On Accommodating Spatial Interactions in a Generalized Heterogeneous Data Model (GHDM) of Mixed Types of Dependent Variables

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, USA
Phone: 1-512-471-4535; Fax: 1-512-475-8744
Email: bhat@mail.utexas.edu
and
King Abdulaziz University, Jeddah 21589, Saudi Arabia

Abdul R. Pinjari
University of South Florida
Department of Civil and Environmental Engineering
4202 E Fowler Ave, ENB 118, Tampa, FL 33620, USA
Phone: 1-813-974-9671; Fax: 1-813-974-2957
Email: apinjari@usf.edu

Subodh K. Dubey
University of South Australia
Institute for Choice
140 Arthur St, Level 13, North Sydney, NSW 2060, Australia
Phone: +61-426-580-105
Email: subbits@gmail.com

Amin S. Hamdi
King Abdulaziz University
Department of Civil Engineering
P.O. Box 80204, Jeddah 21589, Saudi Arabia
Phone: +966-2-640-2000 Ext. 72542; Fax: +966-2-695-2179
Email: ahamdi@kau.edu.sa

ABSTRACT
We develop an econometric framework for incorporating spatial dependence in integrated model systems of latent variables and multidimensional mixed data outcomes. The framework combines Bhat's Generalized Heterogeneous Data Model (GHDM) with a spatial (social) formulation to parsimoniously introduce spatial (social) dependencies through latent constructs. The applicability of the spatial GHDM framework is demonstrated through an empirical analysis of spatial dependencies in a multidimensional mixed data bundle comprising a variety of household choices – household commute distance, residential location (density) choice, vehicle ownership, parents' commute mode choice, and children's school mode choice – along with other measurement variables for two latent constructs – parent's *safety concerns about children walking/biking to school* and *active lifestyle propensity*. The GHDM framework identifies an intricate web of causal relationships and endogeneity among the endogenous variables. Furthermore, the spatial (social) version of the GHDM model reveals a high level of spatial (social) dependency in the latent *active lifestyle propensity* of different households and moderate level of spatial dependency in *parents' safety concerns*. Ignoring spatial (social) dependencies in the empirical model results in inferior data fit, potential bias and statistical insignificance of the parameters corresponding to nominal variables, and underestimation of policy impacts.

Keywords: Spatial econometrics; Multidimensional mixed data models; Latent variables; Maximum approximate composite marginal likelihood (MACML) estimation.

1. INTRODUCTION

Multi-dimensional dependent outcome models are of interest in several fields, including land-use and transportation, biology, finance, and econometrics, just to name a few. The primary motivation for modeling dependent outcomes jointly is that there may be common underlying unobserved factors (attitudes, values, and lifestyle factors) of decision-makers that impact multiple dependent outcomes simultaneously. Ignoring the jointness and considering each dimension separately invites the pitfalls of (1) inefficient estimation of covariate effects for each outcome because such an approach fails to borrow information on other outcomes (Teixeira-Pinto and Harezlak, 2013), (2) multiple statistical testing requirements for specification analysis, which even then offer relatively poor statistical power in testing and poor control of type I error rates (De Leon and Zhu, 2008), and (3) inconsistent estimates of the structural effect of one endogenous variable on another (see Bhat and Guo, 2007). The last of these problems is particularly troubling, since it leads to what is typically referred to in the econometric literature as the "sample selection" or the "endogeneity in variables" problem. That is, modeling each outcome independently with a recursive pattern of influence among the outcomes is tantamount to a strictly sequential decision-making process, which is not consistent with the bundled (or package) nature of multiple outcomes. For example, in a land-use and transportation context, households that are environmentally conscious (and/or auto-averse in their lifestyle) may choose to locate in transit and pedestrian friendly neighborhoods that are characterized by high land use density (the word "auto" in this paper will be used to refer to motorized vehicles in the household). Then, a cross-sectional data set may indicate low auto ownership levels in high land use density areas, but at least part of this effect can be attributed to the purely associative effect of auto-averse households choosing to own fewer autos and residing in high density areas (rather than the low auto ownership being a sole causal result of living in a high density neighborhood). Ignoring this issue will, in general, lead to a misleading conclusion about the causal effect of land-use on auto ownership, which can, in turn, lead to misinformed land-use policies. A way out to more accurately capture causal effects is to model the choice dimensions together in a joint equations modeling framework that accounts for correlated unobserved effects as well as possible causal inter-relationships between endogenous outcomes.

To be sure, there has been a substantial amount of work in the econometric literature on the simultaneous modeling of multiple continuous variables. However, there has been relatively little emphasis on multiple non-continuous variables (see De Leon and Chough, 2013). Bhat (2015a) provides a review of the many different approaches for modeling multiple and mixed data outcomes, and proposes a relatively general modeling framework, which he refers to as the General Heterogeneous Data Model (GHDM) system.

Even as there has been increasing emphasis on mixed data outcome modeling, there also has been a growing interest in accommodating spatial (and social) dependency effects among decision-makers. This is because spatial/social interactions can be exploited by decision-makers to achieve desired system end-states.[1] As a simple illustration of this point, consider household auto ownership, and assume that the number of autos owned by a household influences that of the

[1] In the current paper, we will refer to social/spatial interactions in the strict context of some form of dyadic interaction between individuals located in close social or spatial proximity. Also, our model can be used to capture interactions among decision makers due to proximity in space, or due to any other proximity measure based on social dimensions (such as income earnings, presence of children, virtual social networks of friends/family, or other measures). But for labeling conciseness, we will adopt the terminology of "spatial dependence" (rather than "spatial/social dependence), with the understanding that the proposed model is applicable to any form of proximity-based dyadic interaction processes (and not simply spatial proximity).

household's residential neighbors. Then, a limited-funding information campaign to reduce auto dependency (and promote the use of non-motorized modes of transportation) would do well to target individuals from different neighborhoods, rather than targeting individuals from the same neighborhood. Doing so will benefit from the "ripple wave" (or spatial multiplier) effect caused by intra-neighborhood social exchanges, so that the aggregate-level effect of the information campaign on auto ownership can be substantial. Within the context of accommodating spatial dependencies, spatial lag and spatial error-type autoregressive structures developed for continuous dependent variables are being considered for non-continuous dependent outcomes (see reviews of this literature in Elhorst, 2010, Anselin, 2010, Franzese *et al*., 2010, Ferdous and Bhat, 2013, Bhat *et al*., 2014a, Bhat, 2014, and Bhat, 2015b).[2] Unfortunately, in the case of non-continuous outcomes, accommodating spatial dependence, in general, leads to multidimensional integration of the order of the number of decision-makers for count and ordered-response outcomes, and of the order of the number of decision-makers times the number of alternatives minus one for nominal (unordered-response) outcomes. Typical simulation-based methods, including the frequentist recursive importance sampling (RIS) estimator (which is a generalization of the more familiar Geweke-Hajivassiliou-Keane or GHK simulator; see Beron and Vijverberg, 2004) and the Bayesian Markov Chain Monte Carlo (MCMC)-based estimator (see LeSage and Pace, 2009), become impractical if not infeasible with moderate to large estimation sample sizes (see Bhat, 2011 and Smirnov, 2010). But, recently, Bhat and colleagues have suggested a composite marginal likelihood (CML) inference approach for estimating spatial binary/ordered-response probit/count models, and the maximum approximate composite marginal likelihood (MACML) inference approach for estimating spatial unordered-response multinomial probit (MNP) models (see Bhat, 2014 for a review). These methods are easy to implement, require no simulation, and involve only univariate and bivariate cumulative normal distribution function evaluations. However, all earlier spatial model studies, regardless of the estimation technique used, have focused on a single dependent outcome for each decision maker, rather than multiple and mixed dependent outcomes for each decision maker. On the other hand, when a host of dependent outcomes are co-determined because of common underlying unobserved factors or psychological constructs (attitudes, values, lifestyles, *etc*.), it is very likely that spatial dependence will exist not across just one of those outcomes but across all the outcomes.

In the current paper, we use the important insight that the analyst can generate spatial dependence across multiple and mixed outcomes by specifying spatial dependence in the "soft" psychological construct (latent) variables. In doing so, we combine the GHDM formulation with a spatial formulation. Then, since the mixed outcomes are specified to be a function of a much smaller set of the unobserved psychological constructs in measurement equations, it immediately generates spatial dependence across all outcomes. This is a powerful concept that we have not seen invoked in the literature. While a tantalizingly simple concept, we believe that this has the potential to transform the landscape of spatial econometrics in mixed data modeling. As evidence, we would like to point out that no earlier study in the econometric literature that we are aware of has undertaken a spatial dependence analysis in the context of a relatively large mixed multidimensional model system, as we undertake in this paper. Also, to our knowledge, this is the first study to propose such a methodological structure for introducing spatial dependence in

[2] Of course, the spatial lag and spatial error specifications can be combined together in a Kelejian-Prucha specification (see Elhorst, 2010), or the spatial lag could be combined with spatially lagged exogenous variable effects in a Spatial Durbin specification (see Bhat *et al*., 2014a). In all of these cases, the spatial dependence leads also to spatial heteroscedasticity in the random error terms.

multiple mixed outcomes. At the same time, from a conceptual standpoint, we are able to better disentangle true causal effects from spurious self-selection effects (because the same unobserved factors impact multiple endogenous variables) and spatial dependence effects (because of diffusion of unobserved attitudes and lifestyles based on spatial proximity). Therefore, one can use the model to more accurately examine policy impacts that involve a combination of direct causal effects, self-selection effects, and spatial/social diffusion effects.

Section 2 presents the formulation of the spatial GHDM model along with the MACML estimation approach. Section 3 presents an application of the model to a multidimensional choice bundle consisting of residential location (nominal variable), commute distance (continuous variable), vehicle ownership (count variable), parents' commute mode choice (binary variable), and children's school mode choice (nominal variable). Section 4 concludes the paper.

2. THE SPATIAL GHDM MODEL FORMULATION

There are two components to the GHDM model: (1) the latent variable structural equation model (SEM) system, and (2) the latent variable measurement equation model (MEM) system. In the former system, latent psychological constructs (or latent variables) relevant to the endogenous outcomes of interest in the latter system are posited, based on theoretical psychological considerations, earlier qualitative/quantitative studies, and intuition. These latent variables are expressed as linear functions of exogenous observed variables and stochastic error terms. In the latter measurement system, the mixed outcomes of interest ("endogenous" variables), as well as any subjective indicators of the latent variable vector $\boldsymbol{z}^*$, are written as functions of $\boldsymbol{z}^*$, exogenous covariates, and each other. The structural and measurement equation systems are then estimated jointly based on statistical testing using nested predictive likelihood ratio tests and non-nested adjusted predictive likelihood ratio tests.[3]

2.1 Latent Variable SEM System

Let l be the index for latent psychological constructs $l = (1,2,...,L)$ and q be the index for individuals $q = (1,2,...,Q)$. Then the latent construct z^*_{ql} may be written as a linear function of covariates using a spatial auto-correlation or spatial lag structure as follows:

$$z^*_{ql} = \boldsymbol{\alpha}_l' \boldsymbol{s}_q + \eta_{ql} + \delta_l \sum_{q'=1}^{Q} w_{qq'} z^*_{q'l} \tag{1}$$

where $\boldsymbol{s}_q$ is an (F×1) vector of observed covariates (excluding a constant), $\boldsymbol{\alpha}_l$ is the corresponding (F×1) vector of coefficients, η_{ql} is a random error term assumed to be distributed standard normal, $\delta_l (-1 < \delta_l < 1)$ is the spatial autoregressive parameter (this parameter needs to be bounded in magnitude by the value of one, but can take both positive or negative values; however, we expect the parameter to be positive because attitudes/preferences are likely to be reinforcing through social interactions), and $w_{qq'}$ is an element of a (Q×Q) row normalized spatial weight matrix **W**

[3] There is some level of subjectivity in the number and "labels" of the latent variables posited in the structural equation system. An alternative is to use exploratory factor analysis to identify the latent factors (or latent constructs) through analytic variance minimization, as done in psychology. However, unlike studies in the psychological field that typically collect a battery of items (and sometimes hundreds) of indicators, most economic and transportation studies collect few indicators of the latent factors. So, it is the norm in these fields to posit latent constructs based on a combination of intuitiveness, judgment, and earlier studies (see Bhat, 2015a for a detailed discussion).

with $w_{qq} = 0$ and $\sum_{q' \neq q}^{Q} w_{qq'} = 1 \ \forall q$.[4] This spatial weight matrix is the one that engenders dependencies, with the elements typically parametrized as a decreasing function of geographic or social distance.[5] Next, define the following notations to write Equation (1) in matrix form for all Q individuals.

$\boldsymbol{z}_q^* = (z_{q1}^*, z_{q2}^*, ..., z_{qL}^*)' \ [(L \times 1) \text{ vector}]$, $\boldsymbol{z}^* = \left[(\boldsymbol{z}_1^*)', (\boldsymbol{z}_2^*)', ..., (\boldsymbol{z}_Q^*)'\right]' \ [(QL \times 1) \text{ vector}]$,

$\widetilde{\boldsymbol{s}}_q = \mathbf{IDEN}_L \otimes \boldsymbol{s}_q' \ [(L \times LF) \text{ matrix}]$, $\widetilde{\mathbf{s}} = (\widetilde{\mathbf{s}}_1', \widetilde{\mathbf{s}}_2', ..., \widetilde{\mathbf{s}}_Q')' \ [(QL \times LF) \text{ matrix}]$,

$\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1', \boldsymbol{\alpha}_2', ..., \boldsymbol{\alpha}_L')' \ [(LF \times 1) \text{ vector}]$, $\boldsymbol{\eta}_q = (\eta_{q1}, \eta_{q2}, ..., \eta_{qL})' \ [(L \times 1) \text{ vector}]$,

$\boldsymbol{\eta} = (\boldsymbol{\eta}_1', \boldsymbol{\eta}_2', ..., \boldsymbol{\eta}_Q')' \ [(QL \times 1) \text{ vector}]$, $\boldsymbol{\delta} = (\delta_1, \delta_2, ..., \delta_L)' \ [(L \times 1) \text{ vector}]$,

$\widetilde{\boldsymbol{\delta}} = \mathbf{1}_Q \otimes \boldsymbol{\delta} \ [(QL \times 1) \text{ vector}]$, with " $\otimes$ " representing the Kronecker product.

$\mathbf{1}_Q$ in the notation above is a vector of size Q with all its elements equal to 1. To allow correlation among the latent variables of an individual, we assume a standard multivariate normal (MVN) distribution for $\boldsymbol{\eta}_q : \boldsymbol{\eta}_q \sim \text{MVN}_L[\mathbf{0}_L, \boldsymbol{\Gamma}]$, where $\mathbf{0}_L$ is an (L×1) column vector of zeros, and $\boldsymbol{\Gamma}$ is the correlation matrix of size (L×L). We also assume $\boldsymbol{\eta}_q$ to be independent across individuals (*i.e.*, $\text{Cov}(\boldsymbol{\eta}_q, \boldsymbol{\eta}_{q'}) = 0, \forall q \neq q'$). With this, Equation (1) may be written in matrix form for all Q individuals as follows:

$$\boldsymbol{z}^* = \mathbf{S}\widetilde{\boldsymbol{s}}\boldsymbol{\alpha} + \mathbf{S}\boldsymbol{\eta} \tag{2}$$

where $\mathbf{S} = \left[\mathbf{IDEN}_{QL} - \widetilde{\boldsymbol{\delta}}.*(\mathbf{W} \otimes \mathbf{IDEN}_L)\right]^{-1} \ [(QL \times QL) \text{ matrix}]$, " .* " represents the element by element product, $\mathbf{IDEN}_{QL}$ is an identity matrix of size QL, and $\mathbf{W}$ is a (Q×Q) row normalized

[4] For notation ease, we use the same vector $\boldsymbol{s}_q$ in the equations for all latent variables l. However, this in no way constrains the same exogenous variables to appear in all latent variables because the coefficient on this vector is latent construct-specific (note the subscript on $\boldsymbol{\alpha}_l$). Thus, if a specific variable in $\boldsymbol{s}_q$ does not appear as a determinant of a latent construct z_{ql}^*, this is accommodated by having the corresponding element of the $\boldsymbol{\alpha}_l$ vector set to zero.

[5] Note that the framework is extendable to include general forms of spatial and social dependence. This is because the weight matrix $\mathbf{W}$ can accommodate general forms of dependence. For example, $\mathbf{W}$ itself can be parameterized as a finite mixture of several weight matrices, each weight matrix being related to a specific proximity measure k: $\mathbf{W} = \sum_{k=1}^{K} \varphi_k \mathbf{W}_k$, where φ_k is the weight on the k^{th} proximity variable in determining dependency between individuals ($\sum_{k=1}^{K} \varphi_k = 1$), and $\mathbf{W}_k$ is a matrix with its elements representing a measure of distance between individuals on the k^{th} covariate (for example, see Yang and Allenby, 2003). The important issue though is that the weight matrix should be such that the dyadic interactions between decision-makers fades with spatial or social distance. In the empirical analysis of this paper, we prespecify the elements of $\mathbf{W}$ to be a fixed decreasing function of a single exogenous variable (distance between residences of households). This is standard in much of the transportation literature to acknowledge that the home-end generally tends to be the hub of socialization and interaction. Future studies, however, can use more enhanced (and multi-dimensional) definitions for $\mathbf{W}$, including multiple distance-based separations, using information from social-networking data.

weight matrix. It is now easy to see that $\boldsymbol{z}^*$ is distributed MVN with mean $\boldsymbol{B}$ and correlation matrix $\boldsymbol{\Xi}$. That is, $\boldsymbol{z}^* \sim \text{MVN}_{QL}(\boldsymbol{B},\boldsymbol{\Xi})$, where $\mathbf{B} = \mathbf{S}\tilde{\mathbf{s}}\boldsymbol{\alpha}$ and $\boldsymbol{\Xi} = \mathbf{S}\left[\mathbf{IDEN}_Q \otimes \boldsymbol{\Gamma}\right]\mathbf{S}'$.[6]

The reader will note that Equation (2) is not simply a linear regression equation system with spatial dependence. This is because the left side $\boldsymbol{z}^*$ is unobserved. But when $\boldsymbol{z}^*$ gets included as a determinant of the measurement equation outcomes (see next section), it provides a vehicle to estimate the parameters embedded in Equation (2) through the observations on the MEM outcomes.

2.2 Latent Variable MEM System

We consider a multidimensional mixed outcome system comprising H continuous outcomes, N ordinal outcomes, C count outcomes, and G nominal outcomes, all indicators of the underlying latent consruct vector $\boldsymbol{z}^*$. Let $E = (H + N + C)$. Also, let I_g be the number of alternatives corresponding to the g^{th} (g=1,2,…,G) nominal variable ($I_g \geq 3$), and $\ddot{G} = \sum_{g=1}^{G} I_g$, $\tilde{G} = \sum_{g=1}^{G} (I_g - 1)$. All the E+G outcomes are a function of an ($A\times1$) vector $\boldsymbol{x}_q$ of exogenous variables, which includes a constant, independent variables, as well as possibly the observed values of other endogenous outcomes (introduced as the observed continuous value for continuous endogenous variables, or as observed dummy variables representing each category for nominal variables, or as the observed count value or as corresponding observed dummy variables for each count value). The effects of the endogenous outcomes are "uncorrupted causal" influences after controlling for error correlations across the many dimensions as well as spatial dependencies (engendered by the latent stochastic construct vector $\boldsymbol{z}^*$). These endogenous effects correspond to recursive influences among the dependent variable outcomes.[7]

The observation mechanisms for all the non-continuous outcomes are assumed to be based on underlying latent continuous variables. For each of the ordinal and count outcomes, there is a corresponding underlying latent continuous variable (this is not the same as the latent construct variables in the SEM, but represent underlying variables that are mapped into the actual observed limited-dependent (or non-continous) MEM outcomes (*i.e.*, the observed ordinal. Count, and

---

[6] It is also possible to include the unobserved continuous constructs $z^*_{ql'}$ on the right side of each z^*_{ql} in Equation (1), for $l' \neq l$. However, it may not be easy to justify *a priori* inter-relationships between unobserved variables, and so we prefer a "reduced form" structure as in Equation (1) with a general covariance structure for the latent variables with $\boldsymbol{\eta}_q \sim \text{MVN}_L\left[\mathbf{0}_L, \boldsymbol{\Gamma}\right]$. In cases where it may indeed be appropriate to allow inter-relationships between the latent variables, the econometric identification of the system is possible if a recursive relationship is used so that some latent variables appear as right side variables in the equations for other latent variables in a recursive fashion. Bhat (2015a) presents a detailed discussion of identification conditions for this situation. Let $\pi_{ll'}$ be the effect of the latent variable $z^*_{ql'}$ on z^*_{ql}. Collect these $\pi_{ll'}$ terms (many of which will have to be constrained to zero for recursivity purposes, and $\pi_{ll'}$=0 for l=l') into an $L\times L$ matrix $\boldsymbol{\Pi}$. Then, the only change to our spatial econometric system would be that the matrix $\mathbf{S}$ needs to be re-defined as follows, and the non-zero (and identifiable) elements of $\boldsymbol{\Pi}$ added as parameters to be estimated: $\mathbf{S} = \left[\mathbf{IDEN}_{QL} - \tilde{\boldsymbol{\delta}}.*(\mathbf{W} \otimes \mathbf{IDEN}_L) - \mathbf{IDEN}_Q \otimes \boldsymbol{\Pi}\right]^{-1}$ $\left[(QL \times QL) \text{ matrix}\right]$.

[7] In joint limited-dependent variables systems in which one or more dependent variables are not observed on a continuous scale, such as the joint system considered in this paper that has discrete dependent, count, and ordinal variables, the structural effects of one limited-dependent variable on another can only be in a single direction. See Maddala (1983) and Bhat (2015a) for a more detailed explanation.

nominal outcomes). For each of the nominal outcomes, there are I_g underlying latent continuous variables. The translation from the underlying latent continuous variables to the actual observed outcome for these non-continous outcomes depend on the outcome type, and is discussed in more detail in Appendix A. Based on the many notations introduced there, and collecting the continuous outcomes along with the underlying latent continuous variables across all the non-continuous outcomes into a $(E+\ddot{G})$-dimensional vector $(\boldsymbol{yU})_q$, one can write the following matrix equation for each individual:

$$(\boldsymbol{yU})_q = \ddot{\boldsymbol{b}}\boldsymbol{x}_q + \ddot{\boldsymbol{c}}\,\boldsymbol{z}_q^* + \boldsymbol{\xi}_q, \text{ with } \text{Var}(\boldsymbol{\xi}_q) = \ddot{\boldsymbol{\Sigma}}\ [(E+\ddot{G})\times(E+\ddot{G}) \text{ matrix}] \tag{3}$$

where $\ddot{\boldsymbol{b}}$ is a $[(E+\ddot{\text{G}})\times A \text{ matrix}]$ of the effects of the vector $\boldsymbol{x}_q$ on the underlying latent continuous variables, and $\ddot{\boldsymbol{c}}$ is a $[(E+\ddot{\text{G}})\times L \text{ matrix}]$ of the loadings of the latent constructs on the underlying latent continuous variables.[8] Now, the above Equation (3) for an individual *q* may be used to write a compact expression of endogenous variable equations for all *Q* individuals as:

$$\boldsymbol{yU} = \vec{\boldsymbol{b}}\boldsymbol{x} + \vec{\boldsymbol{c}}\boldsymbol{z}^* + \boldsymbol{\xi}, \tag{4}$$

where $\boldsymbol{yU} = [(\boldsymbol{yU})_1', (\boldsymbol{yU})_2', ..., (\boldsymbol{yU})_Q']'$ $[Q(E+\ddot{\text{G}})\times 1 \text{ vector}]$, $\boldsymbol{\xi} = (\boldsymbol{\xi}_1', \boldsymbol{\xi}_2', ..., \boldsymbol{\xi}_Q')'$ $[Q(E+\ddot{\text{G}})\times 1 \text{ vector}]$, $\boldsymbol{x} = (\boldsymbol{x}_1', \boldsymbol{x}_2', ..., \boldsymbol{x}_Q')'$ $[QA\times 1 \text{ vector}]$, $\vec{\boldsymbol{b}} = \mathbf{IDEN}_Q \otimes \ddot{\boldsymbol{b}}$ $[Q(E+\ddot{\text{G}})\times QA \text{ matrix}]$, and $\vec{\mathbf{c}} = (\mathbf{IDEN}_Q \otimes \ddot{\mathbf{c}})$ $[Q(E+\ddot{\text{G}})\times QL \text{ matrix}]$.

To develop the reduced form model system, substitute the right side of structural Equation (2) in the above equation, as below:

$$\begin{aligned}\boldsymbol{yU} &= \vec{\boldsymbol{b}}\boldsymbol{x} + \vec{\boldsymbol{c}}\left[\mathbf{S}\tilde{\boldsymbol{s}}\boldsymbol{\alpha} + \mathbf{S}\boldsymbol{\eta}\right] + \boldsymbol{\xi} \\ &= \vec{\boldsymbol{b}}\boldsymbol{x} + \vec{\boldsymbol{c}}\left[\boldsymbol{B} + \mathbf{S}\boldsymbol{\eta}\right] + \boldsymbol{\xi} \\ &= (\vec{\boldsymbol{b}}\boldsymbol{x} + \vec{\boldsymbol{c}}\boldsymbol{B}) + (\vec{\boldsymbol{c}}\mathbf{S}\boldsymbol{\eta} + \boldsymbol{\xi})\end{aligned} \tag{5}$$

Then, $\boldsymbol{yU} \sim \text{MVN}_{Q(E+\ddot{G})}\left[(\vec{\boldsymbol{b}}\boldsymbol{x} + \vec{\boldsymbol{c}}\boldsymbol{B}), (\vec{\boldsymbol{c}}\boldsymbol{\Xi}\vec{\boldsymbol{c}}' + \mathbf{IDEN}_Q \otimes \vec{\boldsymbol{\Sigma}})\right]$.

---

[8] Note that even if all the outcomes in the vector $\boldsymbol{yU}_q$ are continuous, estimating each outcome independently would lead to inconsistent estimates (because of endogeneity bias) if there is at least one other endogenous continuous outcome impacting each outcome. To see this, consider the very simple case of two continuous outcomes, a single latent construct, and the first continuous variable also appearing on the right side of the second continuous variable's regression as an element of the vector $\boldsymbol{x}_q = (\boldsymbol{t}_q', y_{q1})'$. Then, at the individual level, the vector Equation system (4) comprises the following two equations:

$$y_{q1} = \boldsymbol{\gamma}_1'\boldsymbol{t}_q + d_1 z_{q1}^* + \varepsilon_{q1} \qquad = \boldsymbol{\gamma}_1'\boldsymbol{t}_q + d_1(\boldsymbol{\alpha}_l'\boldsymbol{s}_q + \eta_{ql}) + \varepsilon_{q1} \qquad = (\boldsymbol{\gamma}_1'\boldsymbol{t}_q + d_1\boldsymbol{\alpha}_l'\boldsymbol{s}_q) + (d_1\eta_{ql} + \varepsilon_{q1})$$

$$y_{q2} = \boldsymbol{\gamma}_2'\boldsymbol{t}_q + \mu y_{q1} + d_2 z_{q1}^* + \varepsilon_{q2} = \boldsymbol{\gamma}_2'\boldsymbol{t}_q + \mu y_{q1} + d_2(\boldsymbol{\alpha}_l'\boldsymbol{s}_q + \eta_{ql}) + \varepsilon_{q2} = (\boldsymbol{\gamma}_2'\boldsymbol{t}_q + \mu y_{q1} + d_2\boldsymbol{\alpha}_l'\boldsymbol{s}_q) + (d_2\eta_{ql} + \varepsilon_{q2})$$

From the expressions above, it is clear that estimating the second equation individually will provide inconsistent estimates because the variable y_{q1} is correlated with the error term $(d_2\eta_{ql} + \varepsilon_{q2})$ in that equation (because of the common error term η_{ql} originating from the latent construct z_{ql}^*. Of course, the situation becomes even more serious (in terms of inconsistency) because not all the $(\boldsymbol{yU})_q$ elements are observed continuous outcomes, but represent latent underlying variables of observed non-continuous outcomes.

Two important points to be noted here. First is that the spatial dependence in the latent construct vector z^* permeates into spatial dependence among individuals for each outcome through the **S** matrix in the first line of Equation (5). For example, in the empirical context of the current paper, a latent construct labeled as "active lifestyle propensity" (ALP) positively impacts the likelihood of a child in the household walking or bicycling to school. By allowing spatial dependence in the ALP across households (based on proximity of household residences), we immediately allow proximate spatial dependence in the walking/bicycling propensity of children to school (that is, households in close residential proximity are likely to interact socially, leading to a diffusion of attitudes regarding active lifestyles and therefore a jointly higher (or lower) likelihood of children of households living close by walking/bicycling to school). Second, given that the number of latent constructs are much fewer than the number of outcomes, spatial dependence is engendered in a very parsimonious fashion within each outcome. In the empirical analysis in this paper, ALP, in addition to influencing children's school mode choice, also positively impacts five other outcomes. Rather than allow spatial dependencies separately within each of the five other outcomes (which would lead to a model that would be proliferate in parameters), our model generates spatial dependency within each of the outcomes based on the spatial dependency in the ALP construct. This is also reasonable from a conceptual standpoint, in that the underlying attitudes are the ones likely to "diffuse" through interactions and then these attitudes impact the outcomes. Thus, just as in the case of children's school mode choice, the implication is that, for example, there is spatial dependence in parents' use of walk/bicycle/public transportation for the commute because of social interactions through which the ALP attitude permeates among individuals living close to one another.

Of course, as should be clear from the last line of Equation (5), jointness is engendered in the outcomes for each individual because of the stochastic nature of the latent constructs (as manifested in the correlation matrix $\mathbf{\Gamma}$ characterizing the $\boldsymbol{\eta}_q$ vector that enters into the $\mathbf{\Xi}$ matrix). Furthermore, because of the spatial dependence in the latent constructs, the net implication is that there is jointness created across all outcomes and across all individuals (note the **S** matrix that also enters into the $\mathbf{\Xi}$ covariance matrix). In summary, our proposed method is a simple, yet powerful and parsimonious way, to incorporate both jointness in outcomes as well as spatial dependencies in outcomes in mixed data modeling.

2.3 Model Estimation

The model in this paper combines a joint mixed outcome system with spatial dependence. In contrast, the previous econometric literature has focused on aspatial joint outcome model systems or on spatial single outcome models.[9] We begin this section by providing an overview of estimators that focus on joint model outcome systems without spatial dependence and single outcome models with spatial dependence.

There have been several estimatiom methods proposed for situations when there are aspatial joint model systems with a few mixed outcome variables (the reader is referred to De Leon and Chough, 2013 for a good review of these methods for mixed outcome systems). The methods include two-stage methods, such as the control function approach or the two stage residual

[9] The reader is referred to Yang and Lee (2015) for estimation methods in the context of spatial dependence in multivariate continuous outcomes. Their approach does not engender spatial dependence through a lower-dimensional latent construct system as we do here, nor does it consider a mix of continuous and non-continuous outcomes as in the current paper.

correction (2SRI) approach (see Terza *et al*., 2008 and Petrin and Train, 2010). But it can be a challenge in two-stage approaches to find good instruments (in fact, we believe the assumptions made in identifying instruments are rather heroic, with all kinds of conceptual justifications provided in the past for instrument selection that we personally find, at best, amusing). The approach also constitutes a limited information approach that can be fraught with econometric efficiency and collinearity problems (Puhani, 2000). Further, even in systems with but two or three mixed outcome variables, the analytic correction or a bootstrapping empirical estimator for obtaining the correct standard errors can be cumbersome (Bhat, 2015a). Further, these two-stage methods do not fare well when there are many mixed outcome variables of interest, where GHDM-type models become appealing because jointness is engendered through a much smaller set of latent constructs. In these models, there are too many constraints that need to be preserved in the measurement equation system, which render control function methods rather ineffective. Additionally, there are also sequential likelihood estimation methods that have been considered in aspatial GHDM-type models. These generally require indicators separate from the outcome variables of interest that provide information on the latent constructs. The methods estimate the structural equation system first using the indicators, and then use the predicted latent constructs as exogenous error-free variables in the measurement equation system. But this approach is deficient because it will, in general, lead to inconsistent and biased estimates (see Hoshino and Bentler, 2013 for a detailed discussion of this issue). Besides, these sequential methods are generally not applicable when there are no indicators separate from the outcomes of interest themselves. On the other hand, the full-information maximum likelihood (FIML) estimator of GHDM-type models is consistent, asymptotically normal, and efficient, subject to the correct parametric assumptions on the stochastic terms and the usual other regularity conditions. But the FIML estimator can be computationally difficult because of multi-dimensional integrals in the optimization function. Typically, a simulated FIML estimator (labeled as the maximum simulated likelihood or MSL estimator) is needed because of the analytic intractability of the integration in the FIML estimator. In such an MSL inference approach, consistency, efficiency, and asymptotic normality of the estimator is critically predicated on the condition that the number of simulation draws rises faster than the square root of the number of individuals in the estimation sample. Unfortunately, for many practical situations, the computational cost associated with the number of simulation draws to ensure good asymptotic estimator properties can be prohibitive and literally infeasible (in the context of the computation resources available and the time available for estimation) as the number of dimensions of integration increases. This is particularly so because the accuracy of simulation techniques is known to degrade rapidly at medium-to-high dimensions, and the simulation noise increases substantially. Increasingly, therefore, a composite marginal likelihood (CML) estimator is used in aspatial mixed model systems where the likelihood function is replaced with a surrogate likelihood function of substantially lower dimensionality. In these CML-based approaches (see Bhat, 2014 for a comprehensive review), the "trick" is to develop a function that is the product of the probability of easily computed marginal events. Bhat and colleagues (see, for example, LaMondia and Bhat, 2011, Sidharthan and Bhat, 2012, and Bhat, 2014, 2015b,c) use a pairwise marginal likelihood in which the probability of pairs of outcomes are first developed, and then these are compounded across all outcomes to develop the CML. For mixed outcome systems with only continuous, binary, ordered, and count outcomes, the CML function contains only bivariate normal cumulative distribution function evaluations. But when nominal outcomes are included, the CML involves a multivariate normal cumulative distribution (MVNCD) function. However, in Bhat's maximum approximate CML (or MACML) procedure, this MVNCD evaluation is

analytically approximated so that, once again, only univariate and bivariate normal cumulative distribution functions need be evaluated. The properties of the CML estimator may be derived using the theory of estimating equations (see Bhat, 2014 for full details). Specifically, under usual regularity assumptions, combined with the normality assumptions on the error terms, the CML estimator is consistent and asymptotically normal distributed (this is because of the unbiasedness of the CML score function, which is a linear combination of proper score functions associated with the marginal event probabilities forming the composite likelihood). A substantial advantage of the CML (or its cousin, the MACML) is that it is very computationally efficient because of its simulation-free nature. Further, while the CML estimator loses some asymptotic efficiency from a theoretical perspective relative to a full likelihood estimator (because information embedded in the higher dimension components of the full information estimator are ignored by the CML estimator), many studies have found that the efficiency loss of the CML estimator (relative to the maximum likelihood (ML) estimator) is negligible to small in applications. Also, CML procedures are typically more robust to mis-specification in the higher dimensions characterizing the overall joint distribution space of all the outcomes, because it relies only on the distribution characterizing the underlying lower dimensional process of pairs of outcomes. That is, the consistency of the CML approach is predicated only on the correctness of the assumed lower dimensional distribution, and not on the correctness of the entire multivariate distribution of all outcomes as in the ML. Additionally, when MSL has to be used, as is the case in most mixed systems because of the intractability of the integrals in the likelihood function, there is once again an efficiency loss in the MSL relative to the ML. Overall, between the CML and the MSL, multiple studies (see Bhat, 2014 for an exhaustive review) have shown that little to no finite sample efficiency loss (and sometimes even efficiency gains) with the CML estimator relative to the MSL estimator.

Of course, when spatial dependencies are considered even in models with a single non-continuous outcome, all the two-stage and limited information approaches have further problems (see Sidharthan and Bhat, 2012 and Arbia, 2014 for reviews of estimation methods for spatial econometric models for univariate non-continuous outcomes; readers may also want to refer to a special issue of *Spatial Economic Analysis* edited by Elhorst *et al*. (2016) for a collection of recent papers on spatial dependence). For example, Klier and McMillen's (2008) linearized version of Pinkse and Slade's (1998) Generalized Method of Moments (GMM) approach is based on a two-step instrumental variable estimation technique after linearizing around zero interdependence, and so tends to work well only for the case of large estimation sample sizes and weak spatial dependence. Also, while it may be more robust relative to full information maximum likelihood to stochastic term functional forms, it loses substantial efficiency because of ignoring dependencies across observations (and identifying spatial parameters using only error term heteroscedasticity). As a result of such limitations of limited-information approaches, it is typical to assume normal distribution errors in the models and use the simulation-based full-information maximum likelihood (FIML) recursive importance sampling (RIS) estimator in the frequentist estimation of spatial models with a non-continuous outcome. Unfortunately, this FIML RIS estimator gets very cumbersome even for moderate to large sample sizes, because the dimensionality of the integrals in the likelihood function to be simulated is of the order of the number of observations in binary/ordered-response outcome models, and of the order of the number of observations times the number of alternative minus one in nominal outcome models. To address this issue, Bhat *et al.* (2010) introduced the composite marginal likelihood (CML) inference approach for the estimation of a spatially dependent binary/ordered-response outcome. Bhat (2011) later proposed the MACML approach for accommodating spatial dependence patterns

in more general outcomes (including nominal outcomes), The CML inference approach, also later used for spatial dependence modeling for a binary outcome under the label of partial maximum likelihood estimation (PMLE) by Wang *et al*. (2013) (that is, Wang *et al*.'s PMLE is exactly the same as Bhat *et al*.'s (2010) CML), replaces the likelihood function with a surrogate likelihood function of substantially lower dimensionality. Example applications of the CML for spatial dependence modeling for a single binary, ordered response, or count outcome include Bhat *et al*. (2010), Castro *et al*., 2012, Ferdous and Bhat (2013), Castro *et al*., 2013, Narayanamoorthy *et al*., (2013), and Bhat *et al*. (2014a), while example applications of the CML for spatial dependence modeling for a nominal outcome include Sener and Bhat (2012), Sidharthan and Bhat (2012), and Paleti *et al*. (2013a).

As indicated earlier, in this paper, for the first time, we combine the modeling of multiple and mixed dependent outcomes with spatial dependence across all outcomes. In such a situation, the problems mentioned above of two-stage methods as well as full-information techniques for the case of mixed outcomes without spatial dependence and single outcomes with spatial dependence compound in terms of limitations and problems. However, the CML approach (and the MACML approach if there are nominal outcomes) still retains its appeal because a pairwise approach across outcomes as well as across observations can still can be relatively easily implemented. In combination with the insight that jointness and spatial dependence can be parsimoniously introduced through the stochastic latent constructs, our approach offers a new methodology for the estimation of spatially dependent mixed outcome model systems. Because the details of this methodology require the notations in Appendix A, as well as because the methodology is very notationally intensive in terms of its overall blueprint, we relegate it to Appendix B.

3. AN EMPIRICAL APPLICATION

In this section, we demonstrate an empirical application of the proposed spatial GHDM by analyzing a multidimensional mixed data bundle of households' long-term and short-term travel-related chocies. Figure 1 depicts the conceptualization of the mixed data bundle. The latent variables (constructs) are represented by the ovals, while the endogenous outcomes (*i.e.*, household choice variables) considered are identified in the rectangular boxes. The two latent variables are parents' *safety concern regarding children walking/bicycling to school* (SCWBS) and household-level *active lifestyle propensity* (ALP). The endogenous outcomes include indicators of the two latent variables. These indicators are identified toward the top of Figure 1, and include three likert scale based ordinal variables to measure SCWBS (see top left corner) – parental concern about violence/crime along the route to school, traffic speed along route, and the amount of traffic along route. The indicators for ALP include three count variables (see top right corner) measuring household-level weekly usage of physically active travel modes – number of episodes in the past week of each of walking, biking, and public transit modes.

The remaining variables in Figure 1 represent endogenous outcomes of actual interest in this study (though they also serve as indicators of the two latent variables, and are conceptually no different from the ordinal/count indicator variables toward the top of the figure). At the bottom of the figure are a continuous variable (household's commute distance) and a count variable (household auto ownership), while the binary and multinomial variables appear just below the latent variables. The binary variable corresponds to an aggregate representation of parents' commute mode choice (=1 if at least one parent in the household uses public transit, walk, or bicycle for commuting, 0 otherwise). The nominal variables are the children's school travel mode (as we will note later, the school mode choice of only one randomly picked child in the household

was recorded in the survey data used in our empirical analysis; however, the method we propose can easily be extended to include the school mode choice of each child if such data were available), and another nominal variable for residential location choice based on neighborhood density of households (households per square mile in the Census block group of the househod's residence, as obtained from the 2010 decennial Census data). This last nominal variable and commute distance (the continuous variable), taken together characterize household residential location in the current empirical analysis. Further details on the construction and descriptive statistics of each of the outcome variables are provided later.

As noted earlier, in addition to the indicator variables toward the top of Figure 1, the actual endogenous outcomes of interest (below the latent variable ovals) also represent manifestations of the latent variables. In the figure, alternative effects of the latent constructs on the actual endogenous outcomes of interest were considered, and the final specification for the effects of the latent construct effects was based on statistical testing (see Bhat, 2015a for a discussion). While we discuss the specification results later in Section 3.2.3, our *a priori* hypothesis (consistent with Figure 1) is that households with a higher SCWBS (relative to other households) will be less likely to let their children walk/bicycle to school. Also, households with a higher ALP are more likely to have their children walk/bicycle to school (due to the potential physical activity benefits of doing so), commute to work by non-auto modes, and reside in dense neighborhoods (see Walker and Li, 2007, Bhat, 2015a, and Bhat, 2015c). All these influences of latent variables on endogenous variables are depicted by dotted lines/arrows in Figure 1.

Finally, in Figure 1, solid arrows from one endogenous outcome to the other endogenous outcome represent causal (recursive) relationships, after accounting for associations among the endogenous outcomes caused by the stochastic latent variables. Note that the figure represents one set of relationships among the endogenous outcomes based on testing a variety of different relationships identifiable in the GHDM framework (see Bhat, 2015a for detailed discussed on identification issues in the GHDM framework). We discuss these endogenous inter-relationships later in Section 3.2.4.

The selection of the choice bundle in Figure 1 is motivated from a couple of reasons. First, most multidimensional choice studies in the literature have focused on modeling only two or three of the dimensions of residential location, auto ownership, commute distance, and parents' (or adults') commute mode choice (Abraham and Hunt, 1997, Bhat and Guo, 2007, Pinjari *et al*., 2011, Paleti *et al*., 2013b, and Bhat *et al*., 2014b). Here we model all of these, as well as children's school travel mode choice as part of a bundle of travel behavior and residential location choice decisions. While numerous studies exist in the literature on modeling children's school mode choice as a function of sociodemographic characteristics, residential location, vehicle ownership and other attruibutes (see, for example, Yarlagadda and Srinivasan, 2008, Sidharthan *et al*., 2011, and McDonald, 2008), none of these earlier studies consider children's school mode choice as part of a bundle of travel behavior and location choice decisions. Therefore, these studies ignore potential endogeneity between children's school mode choice and other choices such as residential location attributes (density) and auto ownership. Ignoring such endogeneity might result in biased estimation of the influence of residential location attributes and potentially distorted policy implications of, for example, neo-urbanist initiatives to densify neighborhoods (see Section 3.3 later). A second reason for the choice bundle used here is that a number of studies have incorporated spatial dependency when analyzing the above identified choice dimensions individually (see for example, Sidharthan *et al*., 2011 in the context of children's school mode

choice), but no earlier study, to our knowledge, has considered spatial dependence in multiple and mixed outcomes simultaneously.

3.1 Empirical Data

The primary data source used for this study is the California add-on sample of the 2009 National Household Travel Survey (NHTS) conducted by the US Department of Transportation. The add-on survey sample includes detailed information about socio-demographic, residence, vehicle, and activity-travel characteristics for a 24-hour survey period from 21,225 households in the state. Of these, only about 13.5% of households (a) had children and (b) were targeted for collection of parents' concern on safety issues related to their children's travel and the actual school mode choice of a single randomly picked school-going child. In our analysis, we focused only on such households with at least one worker. Further, recognizing potential differences between different regions of the state, we narrowed down our analysis to households from the following contiguous (based on shared boundaries) ten counties in southern California: San Luis Obispo, Kern, Santa Barbara, Ventura, Los-Angeles, Orange, San Bernardino, Riverside, San Diego, and Imperial. After some additional minor cleaning, the resulting final estimation sample comprised 1538 households.

To conserve on space, we relegate details of the exogenous variable characteristics of the sample to Section 3 of the online supplement.

3.1.1 Dependent (Endogenous) Outcome (Variable) Characteristics in the Sample

Table 1 provides the descriptive statistics of the endogenous outcomes in the sample, which are briefly described here.

Continuous Outcome

The household-level average commute distance (or, household commute distance), measured as the average of one-way commute distance reported across all commuters in the household, is the only continuous dependent variable in our empirical analysis. The sample average of household commute distance is 15.15 miles. The average reported commute time in southern California is about 26.9 minutes (Lin, 2012) which, given the average commute distance of 15.15 miles, translates to an average speed of 33.7 miles/hour, a reasonable commute travel speed for an urban scenario. For model estimation purposes, we used the natural logarithm of the household commute distance variable.

Ordinal Outcomes

The three ordinal variables considered in this analysis correspond to parents' concerns about crime and traffic along their children's route to school (see second column panel in the top portion of Table 1). All of these ordinal variables, measured on a 5-point Likert scale, are used in the measurement equations to identify the latent construct SCWBS of the household. The descriptive statistics of these variables in the sample suggests that speed and amount of traffic along the children's school travel route are matters of greater concern than violence/crime along the route.

Count Outcomes

There are four count variables: number of bicycling episodes in the past week, number of walking episodes in the past week, number of times public transit used in past week, and auto ownership. The first three count variables were recorded for every individual in the household. We aggregated

the individual-level variables to the household level for use in the measurement equations. The descriptive statistics of these variables suggest a greater amount of walking than bicycling, in terms of the number of trips per week. Further, only 33% of the households in the sample used public transportation at least once in the past week. Finally, in the context of household auto ownership, a vast majority of households in the sample own at least one vehicle, with one half of the households owning two vehicles and about 40% of the households owning 3 or more vehicles.

Binary Outcome

The only binary outcome in the current study is an indicator for the use of public transportation, walk, or bike as the commute mode by at least one commuter in the household on the survey day. This variable is labeled as "parents' commute mode choice". For ease of presentation, we will refer to the walk, bicycle, and public transportation modes collectively as non-auto modes. The descriptive statistics in Table 1 show that only about 8% of the households in the sample used non-auto modes for commuting.

Multinomial Outcomes

The two multinomial outcomes are residential location choice and children's school mode choice. The following four categories were considered for residential housing density variable: (1) less than 1000 hh./sq. mile, (2) 1000-1999 hh./sq. mile, (3) 2000-3999 hh./sq. mile, and (4) 4000 or more hh./sq. mile. As may be observed from the descriptive statistics of this variable, 60% of the households live in very low (less than 1000 hh./sq. mile) to low (1000-1999 hh./sq. mile) density locations. Also, only about 11% of the households live in high (4000 or more hh./sq. mile) density locations. For estimating the parameters of this variable, we consider the residential housing density category of "less than 1000 hh./sq. mile" as the base category.

For the children's school mode choice, the following four categories are considered: (1) car (either driven by parents or others), (2) bus (school bus or public transportation), (3) walk/bicycle and (4) other modes (taxicab, street car or others). The car mode is the predominant (about 70%) mode of children's school travel. But a non-significant proportion of children use the bus (about 10%) and walk/bicycle (18.5%) modes. The car mode is the base category.

3.2 Model Estimation Results

A variety of different empirical model specifications were estimated in this study, including alternative weight matrices for spatial dependency, the influences of exogenous variables on the latent constructs, the impacts of exogenous variables and latent constructs on the endogenous outcomes, and alternative recursive inter-relationships among the endogenous outcomes. The final empirical model specification was determined based on a combination of statistical data fit, parsimony in specification, and ease in interpretation.

3.2.1 Selection of the Weight Matrix

The spatial weight matrix contains information on the nature and decay of spatial dependencies with spatial separation. To construct this matrix, we first developed a matrix of distances between each (and every) pair of households. The distances were measured between the centroids of the census tracts of the household locations. Next, the following six different weight configurations were considered: (1) a same/contiguity tract indicator (*i.e.*, $w_{qq'} = 1$ if households q and q' are in the same tract or in continguous tracts, and 0 otherwise), (2) a shared boundary length measure (computed as the perimeter of the census tract for two households q and q' in the same tract, and

as the shred boundary length if the two households are in continguus tracts), (3) inverse of continuous distance, (4) inverse of the exponential of continuous distance, (5) inverse of the square root of continuous distance, and (6) inverse of the square of continuous distance. The best weight configuration is chosen based on a composite likelihood information criterion (CLIC) statistic. The weight configuration that provides the highest value of the CLIC statistic is the preferred one (see Bhat, 2011; Sidharthan and Bhat, 2012). In our analysis, this came out to be the inverse of the square root of distance, with the best specification resulting when the spatial dependence reduces to zero beyond a threshold distance of one mile. Details of the CLIC statistics for the alternative weight configurations are available from the authors on request.

3.2.2 Parameter Estimates of the Structural Equations for Latent Variables

The parameter estimates of the structural equations for the two latent variables, SCWBS and ALP, are presented in Table 2 and discussed below.

Safety concerns about children walking/biking to school (SCWBS)

The parameter estimates suggest that parents of younger children exhibit a higher SCWBS than parents of older children. This is intuitive as parents are likely to feel more confident in their older children's ability to navigate around motorized traffic on their path to school, and also be less vulnerable to violence/crime on streets. The finding is also consistent with that reported earlier by Alton *et al*. (2007) and Seraj *et al*. (2012). Parents also exhibit a higher level of safety concern for girls than boys. This may be because girls are more likely to be victims of sexual offenses, and also perhaps due to lingering cultural biases that provide boys more independence than girls (see McLean and Anderson, 2009 and Seraj *et al*., 2012).

Table 2 further indicates that households with lower education levels and households with lower income levels exhibit a lower level of safety concern, perhaps because such households may be concerned about basic needs such as food and shelter that precede safety concerns in Maslow's hierarchy of needs (Huitt, 2007). Besides, it has been documented in the literature (Seccombe, 2002) that lower income families tend to have a greater tendency of resiliency, particularly in the context of "adapting to risk in order to maintain competence in adverse conditions" (Orthner *et al*., 2004).

Active lifestyle propensity (ALP)

The parameter estimates suggest that Asians and Hispanics tend to exhibit a lower level of ALP than Caucasians and African-Americans. Other studies in the literature (see, for example, Saffer *et al*., 2011 and Sener and Bhat, 2007) have also found such racial differences in physical activity participation and attribute them to cultural differences. The higher levels of ALP among Caucasians may also be attributed to a higher priority placed on physical appearance, perhaps as a facet of identity, for Caucasians in contemporary Western societies (see, for example, Dworkin and Wachs, 2009 and Engelsrud, 2009). Interestingly, income was not found to be a significant correlate of ALP.

Households with a higher fraction of young adults (19-30 years) and a higher fraction of well-educated adults (bachelor's degree or beyond) exhibit a higher level of ALP than households with a lower fraction of young adults and a lower fraction of well-educated adults, respectively. (see also Bauman *et al*., 2012). While the former may simply be an indication of the physiological health status of younger adults relative to their older peers, the latter is presumably a reflection of higher educated individuals being better aware of the health benefits of an active lifestyle (Cutler

and Lleras-Muney, 2006). Finally, consistent with previous findings (Belcher *et al*., 2010), households with young children (less than 16 years of age) have a higher ALP than households with older children (16-18 years). This result may be a consequence of older children being more involved (alone or with their peers) in sedentary activities such as television watching, internet surfing, videogame play, and talking/texting on the phone. Studies by Copperman and Bhat (2007), Sener *et al*. (2008), and Heitzler *et al*. (2011) provide support for this interpretation.

Correlation between the latent constructs

The correlation between the SCWBS and ALP constructs is very small and positive.

Spatial autoregressive parameters

The spatial autoregressive parameter estimates for SCWBS and ALP are 0.447 and 0.846, respectively, and highly statistically significant, confirming our hypothesis that the two latent variables are spatially dependent. As indicated in Section 2.2, the spatial dependency in these latent variables permeates as spatial dependency in all the endogenous variables influenced by these variables. The spatial dependency coefficient for ALP is particularly large, suggesting substantial spillover effects in active lifestyle propensity among those living in close geographic proximity.

3.2.3 Latent Construct Loadings on Endogenous Variables

Table 3 presents the parameter estimates of the loadings of latent constructs on the various endogenous variables. As expected, the loadings of the latent construct SCWBS on all three Likert scale variables measuring parents' SCWBS are positive and highly statistically significant. Consistent with the descriptive statistics, the loading of SCWBS on the "violence/crime" variable is of the smallest magnitude. The SCWBC latent variable also influences the school mode choice of children, with a high SCWBS leading to a higher reluctance among parents to let their child walk, bicycle, or go by bus to school. The intensity of this reluctance is a function of distance, as we discuss further in Section 3.2.5 (in Table 3, we only provide the loading of SCWBS on school mode choice corresponding to a distance of over two miles).

The loadings of the latent construct ALP on all the three count variables measuring the weekly usage of walking, bicycling, and public transit modes are positive, as expected. In addition, households with greater levels of ALP shy away from living in very low density (< 1000 hh./sq. mile) neighborhoods and exhibit a preference for denser neighborhoods (presumably because denser neighborhoods tend to have better walking and biking facilities, and greater proximity to different recreational activity locations; see, for example, Bhat *et al*., 2016). In addition, households with high ALP are more likely to commute by active (that is, non-auto) travel modes as well as encourage children to travel by the non-auto modes. In this sense, the latent construct ALP contributes to residential self-selection, where households that prefer to travel by active travel modes (both for adults' commuting and children's school travel) reside in higher density neighborhoods that allow them to do so. If such self-selection effects are not accounted for, there is a risk of overestimating the influence of residential density on the choice of active travel modes for both commuting and school travel. While a number of studies in the literature discuss residential self-selection effects in the context of commute mode choice (see, for example, Pinjari *et al*., 2008), not many studies highlight such self-selection effects in the context of children's school mode choice.

Interestingly, we did not find any direct statistically significant effects of the ALP latent construct on both auto ownership and household commute distance (however, note that ALP impacts residential density of location, which, in turn, influences auto ownership, as discussed in the next section).

3.2.4 Relationships Among Endogenous Variables

The parameter estimates of the causal and recursive relationships among the endogenous variables are presented in Table 4. The reader will note, however, that regardless of the presence or absence of recursive effects, the model is a joint model because of the presence of stochastic latent variables that impact the many dependent outcomes. For this reason, we also characterize the Table 4 results as "true" causal effects after associations due to common underlying unobserved effects are accommodated. Figure 1 shows a path diagram of these causal relationships, in the form of solid arrows from one endogenous variable to the other. The recursive structure of relationships has been determined after an extensive testing of alternative recursive structures based on overall model fit.

The chain of causal relationships starts at household commute distance, which influences residential location. This is interesting because most other studies use residential location density (the built environment) as an exogenous variable in commute distance modeling (see, for example, Sultana and Weber, 2014). The implication in these earlier studies is that dense neighborhoods engender shorter commutes, ostensibly because there are more employment opportunities in dense areas (the implicit assumption then is that individuals choose work locations after choosing their residential location). Our study, though different from most earlier studies in that it considers commute distance at a household level (as opposed to the individual level of earlier studies), suggests the reverse – that households deliberately choose to live in dense locations to minimize average household commute distance (note also that, in our analysis, we did not find any effects of the latent constructs on household commute distance, suggesting that household commute distance is truly a decision made before all the other decisions modeled). There is also the suggestion in our result that work locations (and work choices in general) are typically determined prior to household location decisions, as also observed by Rashidi *et al*. (2012). Overall, our results do bring to question the notion that densification of neighborhoods by itself will result in shorter commutes, or that urban sprawl will necessarily lead to longer commutes. Next, both household commute distance and residential location influence households' auto ownership; households with a longer commute distance and those living in low density neighborhoods are likely to own more vehicles. These results are consistent with much of the earlier literature (see, for example, Bhat and Guo, 2007; Bhat *et al*., 2009; Aditjandra *et al*., 2012, Bhat *et al*., 2014b, and Brownstone and Fang, 2014). Household auto ownership, in turn, is used in the form of an auto availability variable to explain other endogenous outcomes.[10] In particular, household auto availability, commute distance, and residential location influence adults' commute mode choice, as one would expect (see, for example, Bhat and Sardesai, 2006 and Pinjari *et al*., 2011). And all these four endogenous variables influence children's school mode choice, as discussed in the next section. Finally, residential location and auto availability variables influence the weekly usage of public transit.

[10] As may be observed from the last but one column of Table 4, the auto availability variable is defined on the basis of whether each adult with a driver's license has access to at least one auto in the household.

3.2.5 Children's School Mode Choice Model Component

To conserve on space, we do not provide the full estimation results for each of the endogenous variables in terms of exogenous variable effects. These are available in Section 4 of the online supplement. Here we focus on children's school mode choice, especially because there has been relatively less work on this endogenous variable compared to other endogenous variables, as well as because this is, to our knowledge, the first paper that jointly considers both residential selection effects as well as social interaction effects on children's school mode choice. Table 5 presents the parameter estimates, including those of exogenous variable effects, loadings of the latent constructs, and the effects of endogenous variables (while the last two sets of effects have been touched upon in earlier sections, we discuss these in more detail here).

Exogenous variable effects

The results suggest that the likelihood of using non-auto (bus or walk/bicycle) modes decreases with the increase in the number of workers with the option to work from home or who have a flexible work schedule. Spatial and temporal flexibility in work activity provides flexibility for working parents to adjust their work timings and chauffer children to school (Yarlagadda and Srinivasan, 2008). In addition to these two exogenous variables, we explored the role of other demographics such as age and gender, but did not find a statistically significant influence on school mode choice after controlling for the indirect effects of these variables on children's school mode choice through the latent construct SCWBS.

Moving on to the home-to-school distance variable, children are increasingly likely to walk/bicycle to school as the distance from home to school decreases. Several other studies have also highlighted the influential role distance has on children's school mode choice (Ewing *et al*., 2004, McDonald, 2008, and Kelly and Fu, 2014). In this context, Broberg and Sarjala (2015) suggest that denser school networks with more neighborhood schools located in close proximity to a high proportion of households with school-going children can help increase the share of walking/bicycling to school. Interestingly, children whose residences are farther than 2 miles away from school are more likely to take the bus mode than other children. This could be due to a combination of the distance effect (that is, the walk/bicycle mode become less feasible for distances longer than 2 miles) as well as the potential unavailability of the school bus option for households within a two-mile distance from school. Many school districts provide school buses only for households that live beyond a 2-mile radius. Thus, the school bus may not be an option for children living within a two-mile radius of their school.[11] Besides, specialized schools such as magnet schools and choice schools that draw children from wider geographic regions (that are far beyond 2 miles from the school) have been shown in the literature to have a greater proportion of children traveling by school bus than those in other schools (see Wilson *et al*., 2010).

Latent construct effects

The effects of latent constructs are intuitive and expected, as discussed in Section 3.2.3. We also interacted the SCWBS latent construct with different ranges of home-to-school distance variable. The results suggest, consistent with the findings of Seraj *et al.* (2012), that the influence of SCWBS is moderated (reduced) as the distance decreases, perhaps because of lower exposure to risks and a greater familiarity (hence greater level of comfort) with the travel route for shorter distances.

[11] The survey did not seek information on the availability of the school bus mode. Further, the "bus" alternative in our model includes not only the school bus, but also public transit buses. Future studies of children's school travel mode need to pay more attention to construction of the availability of the bus mode to school.

Endogenous variable effects

As discussed in Section 3.2.4, a variety of different endogenous variables influence children's school mode choice. Note again that these effects are "true" causal effects after accommodating associations engendered among children's school mode choice, parents' commute mode choice, and residential location by the underlying ALP latent construct. The results indicate that adults' use of non-auto modes for commuting tends to increase children's use of non-auto modes for school travel. This may well be because parental non-use of an auto mode for the commute implies less possibility of a child being dropped by the auto mode. Further, children from households with a longer average commute distance or households with higher vehicle availability are less likely to travel to school by non-auto modes, as also observed by Seraj *et al*. (2012).

In terms of residential location density effects, as expected, children from households living in denser neighborhoods (>2000 households per square mile) are more likely to walk or bicycle to school than those in households residing in other neighborhoods. This is presumably because dense neighborhoods tend to have better pedestrian and bicycling facilities. Further, children from households in very high density (greater than 4000 hh./sq. mile) and very low density (less than 1000 hh./sq. mile) areas show a greater propensity to take the school bus than those in households residing in the mid-range residential density (1000-3000 hh./sq. mile). What is more, the propensity to take the bus is highest among those residing in the lowest density neighborhoods, which is rather surprising. This is likely related to the issue discussed earlier of the school bus perhaps being more available when the distance to school is longer, and reinforces the need to pay more attention in future studies on the construction of the availability of the school bus mode. Of course, this result could also be a result of special schools (that tend to offer specialized bus services) being more prevalent in less dense neighborhoods. In any case, the relationship between the availability/use of the bus mode to travel to school and residential density certainly deserves more careful attention in future studies.

Error covariance and spatial dependence

We allowed a non-IID covariance matrix for the error vector among the random utility components of the children's school mode choice alternatives, but an IID error structure sufficed for the current empirical model. However, this does not imply an IID utility structure, because of the presence of the SCSWB and ALP stochastic latent constructs, which engender higher sensitivities among the non-auto (walk, bicycle, and bus) modes than between these modes and the car mode.

The spatial dependence in both the SCWSB and ALP latent constructs permeate into the school bus mode choice decision. That is, children/adults in households in close proximity are more likely to uniformly attribute a higher (or lower) utility for each of the bus and walk/bicycle modes, a very clear sign of social interaction effects and/or unobserved neighborhood location effects that affect modal valuations. It is possible that parents of households living in close proximity interact with one another and share experiences about school travel of their children, or households may band together to facilitate walking and bicycling in a safe and secure way. The net result is a spatial "spillover" effect, leading to a multiplier effect in terms of the effectiveness of programs to promote the use of non-auto modes for school travel. When the SCWSB and ALP of even just a few parents/households are impacted through targeted campaigns, it has a "spillover" impact on other parents/households in close proximity, leading to a "snowballing" effect on the use of non-auto modes of travel to school for children in all households in the neighborhood.

3.2.6 Comparison of the Empirical GHDM Models With and Without Spatial Dependency

Data fit

The log-CML value of the spatial GHDM model is -2580211.30. For the same empirical specification, the log-CML value for the aspatial GHDM model is -2590836.37. The difference in data fit between the spatial model and the aspatial model may be computed using the ADCLRT statistic. The calculated ADCLRT statistic value is 157.59, which is higher than the critical chi-square value with two degrees of freedom at any reasonable level of significance. This clearly underscores the importance of considering spatial dependency. Further, to assess the importance of considering jointness across the endogenous outcomes, we also estimated an Independent Heterogeneous Data Model (IHDM) that ignores the jointness among the endogenous outcomes engendered by the stochastic latent constructs. In this IHDM model, we introduce the exogenous variables (sociodemographic variables) used to explain the latent constructs directly as exogenous variables. The resulting IHDM may be compared to the GHDM using the composite likelihood information criterion (CLIC) introduced by Varin and Vidoni (2005). The model that provides a higher value of CLIC is preferred. The CLIC statistic values for the aspatial GHDM and IHDM models were estimated to be -2595906.78 and -2619331.83, respectively. These CLIC statistics clearly favor the GHDM over the IHDM.

Differences in variable effects

In addition to the differences in data fit, we observed several differences in the estimates of different exogenous and endogenous variable effects. To focus the discussion, we only qualitatively discuss the differences between the spatial and aspatial GHDM models. First, in the context of the structural equations, the spatial model suggested significant differences between Hispanic and Caucasian races in the latent construct *active lifestyle propensity* (ALP). On the other hand, the aspatial model did not reveal statistically significant differences between Hispanics and Caucasians in ALP.

In the measurement equations for non-nominal variables, we did not notice striking differences between the parameter estimates (and corresponding interpretations) of the spatial and aspatial models. For the measurement equations of the nominal variables, however, we noticed some notable differences, as discussed below. In the residential location (density) choice component of the model, the influence of ALP in the aspatial model was not as pronounced as in the spatial model. In children's school mode choice, the spatial model suggested that flexibility in adults' work timings reduces the likelihood of children walking/biking or taking a bus to school. On the other hand, the aspatial model did not reveal any such effect. Moving on to the inter-relationships among endogenous variables, the aspatial model suggested a weak influence (with a small t-statistic value) of household-level commute distance on auto ownership. The spatial model, on the other hand, revealed a stronger influence of household-level commute distance on vehicle ownership. Overall, all these differences between the two models indicate that ignoring spatial dependency may not only lead to deterioration in overall data fit but may also lead to either a bias or statistical insignificance (or a combination of both) of important exogenous and endogenous variable effects.

3.3 Disentangling Different Effects

We indicated in the introductory section that our model is able to disentangle three distinct effects associated with variable impacts. Here, we consider the effects of a neo-urbanist policy aimed at

densification of neighborhoods. To keep the discussion focused, we again examine these effects only in the context of children's travel mode to school.

As identified in Section 3.2.3, households who are inclined to use non-auto modes self-select to live in dense neighborhoods. If this residential self-selection effect is ignored (as is done by the IHDM model), the effect of moving a random household from a low density neighborhood to a high density neighborhood (or, equivalently, densifying an existing low density neighborhood) would be magnified in terms of the increase in non-auto mode use to travel to school. Thus, the difference between the aspatial GHDM and the IHDM in the effect of residential density provides the "spurious" residential self-selection (RSS) effect. Next, consider the aspatial and spatial GHDMs. The latter accommodates spatial/social dependence, while the former does not. That is, the spatial GHDM recognizes the social interactions among households in close proximity, as discussed in the previous section. That is, a random household moved from a low density neighborhood to a high density neighborhood is likely to be influenced by the higher ALP of other households already residing in the high density neighborhood, resulting in a higher ALP of the household and a higher propensity of children in the household to use non-auto modes to school. This, unlike the residential self-selection effect, is a post location spatial interaction effect. The difference between the aspatial GHDM and spatial GHDM provides the social/spatial dependence (SSD) effect.

To quantify (and disentangle) the magnitude of the RSS, the "true" causal effect of densification, and the SSD effect, we compute average treatment effects (ATEs) from the IHDM, the aspatial GHDM, and the spatial GHDM models. The ATE measure for a variable provides the expected difference in that variable for a random household if it were located in a specific density configuration i as opposed to another density configuration $i' \neq i$. Here we compute the ATE corresponding to a hypothetical scenario when a household is transplanted from the lowest density (less than 1000 hh./sq. mile) location to the highest density (greater than 4000 hh./sq. mile) location. To calculate the ATE, for each of the models, a realization of the vector $\boldsymbol{yU}$ is constructed (see Equation 5) by appropriately drawing from the distribution of all the relevant parameters ($\vec{\boldsymbol{b}}, \vec{\boldsymbol{c}}, \boldsymbol{B}, \boldsymbol{\Xi}$, and $\vec{\boldsymbol{\Sigma}}$). Then the value of different dependent variables is calculated appropriately by following the chain of causal effects among the endogenous variables. Since residential location density is a nominal variable, the procedure to calculate the ATE is as follows: First, set the value of the residential density variables to zero for all the density categories for all households in the sample and, using the procedure just described above, compute the expected share of each mode for the children's school choice. In doing so, the expected share is computed assuming that all households in the sample live in the lowest density location. Second, set the value of the residential density variables to zero for all the categories except for the highest density category variable (for which a value of one is applied for all the households in the sample), and compute the expected school mode share for each alternative. Finally, compute the ATE for each alternative as the difference between the expected shares obtained between the second and first steps. The above described procedure is repeated 500 times. The mean across the 500 runs was computed as the final ATE effect and the standard deviation across the 500 sets was computed as the standard error estimate.

Table 6 presents the estimated ATE values (and standard errors) for children's school mode choice for the IHDM, the aspatial GHDM, and the spatial GHDM models. The first row under the "IHDM model" heading indicates that a random household that is shifted from a low density location to a high density location is, on an average, likely to reduce auto use probability for children's school travel by 0.082 (standard error of 0.026). Equivalently, if 100 random households

are moved from a low density to a high density location, the auto use mode share among these 100 households will reduce by 8.2%. On the other hand, the aspatial GHDM model estimate predicts a probability reduction of only 0.029 (standard error of 0.015). That is, according to the aspatial GHDM, if 100 random households are moved from a low density to a high density location, the auto use mode share among these 100 households will reduce by only 2.9%. The exaggeration in the reduction in auto use based on the IHDM model is because unobserved residential self-selection effects are not being controlled for. The p-value value for the hypothesis of equality in the ATE estimates is 0.039, clearly rejecting the equality hypothesis even at the 96% level of confidence. Other values in the IHDM and aspatial GHDM columns in the table can be similarly interpreted. The difference in the ATE estimates are statistically significant at the 92% level of confidence for the bus mode and the 99% level of confidence for the walk/bicycle mode. We do not pay much attention to the "other modes" alternative because of the very low percentage of this alternative in the estimation sample. The ATE estimates are also different between the aspatial and spatial GHDM models, even if not that statistically significantly different as between the IHDM and aspatial models, with the spatial model indicating a higher ATE for all modes. This is expected, because the aspatial model ignores the multiplier effect.

The magnitude of the RSS, the true "causal" effect of densification, and the SSD effect are computed as follows. First, the "true causal" effect contribution is considered to be represented by the ATE of the aspatial model. The SSD effect contribution is computed as the difference between the ATE effects from the spatial and aspatial GHDM. Finally, the IHDM ATE that combines (and convolutes) all of the effects is used to subtract the "true" and RSS effects from to obtain the RSS effect. For instance, for the car mode, the "true causal" effect contribution is -0.029, the SSD effect is -0.015 (-0.044-(-0.029)) and the RSS effect is -0.038 (= -0.082-(0.029+0.015)). The percentage contributions of the three effects are then computed and shown in the last column panel of Table 6. As can be observed, for all the three modes with a tangible proportion of users, the RSS effect is of the order of 43-47%, while the "true causal" effect is about 35-41% and the SSD effect is of the order of 16-18%. These results show a substantial residential self-selection effect as well as the presence of SSD effects. The latter result confirms the results in Section 3.2.5 that targeted campaigns where the mode choices of children in just a few parents/households in a neighborhood are impacted can have a "spillover" impact on other parents/households in close proximity. The results also show that tangible "true causal" travel effects of the built environment do exist in the land use-travel behavior association, even after accommodating for the RSS effect.

4. SUMMARY AND CONCLUSIONS

This paper develops a framework for incorporating spatial dependencies in integrated model systems of latent variables and multidimensional mixed data outcomes. The framework combines Bhat's Generalized Heterogeneous Data Model (GHDM) with a spatial formulation and introduces spatial dependencies through latent constructs. The resulting spatial GHDM is flexible yet very parsimonious due to the use of latent constructs (of attitudes and lifestyle preferences) as a vehicle for introducing spatial dependencies among the multitude of endogenous variables in multidimensional mixed data model systems. Since the spatial dependencies introduced in latent constructs permeate into all the endogenous outcomes influenced by the latent constructs, the approach obviates the need for incorporating spatial dependencies separately for each and every endogenous variable.

For estimating the parameters of the proposed spatial GHDM framework, the paper employs the maximum approximate composite marginal likelihood (MACML) approach which

reduces the dimensionality of integration required for estimation into a series of univariate and bivariate normal integrals, regardless of the number of latent constructs and the number of dependent variables in the multidimensional mixed data bundle.

The paper presents an empirical application of the proposed spatial GHDM framework by analyzing a multidimensional mixed data bundle of households' long-term and short-term travel-related chocies in a household travel survey sample from the South California region. The endogenous variables (*i.e.*, household choice variables) considered in the mixed data bundle are: (1) a nominal variable for residential location choice, (2) a count variable for vehicle ownership, (3) a continuous variable for household-level average commute distance, (4) a binary variable for parent's commute mode choice, and (5) a nominal variable for children's school travel mode. Along with these variables, three count variables and three ordinal variables were used to identify two latent constructs – *parents' safety concerns for chidlren walking/biking to school* (SCWBS) and *active lifestyle propensity* (ALP) – to build the GHDM moel. The empirical model reveals interesting insights on the influence of different exogenous variales and the two latent variables on the above-mentioned household choice variables. As importatnly, the GHDM framework helped in identifying an intricate web of causal relationships among the multitude of endogenous variables, as well as disentangling three different effects of variables: the residential self-selection (RSS) effect, the social/spatial dependency (SSD) effect, and the "true" causal effect. In an examination of the effect of neighborhood densification (as part of a neo-urbanist policy) on children's school mode choice, our results showed that the residential self-selection and true causal effects of a densification-based neo-urbanist policy on school mode choice of children are about 45% and 38%, respectively, and also that there is a tangible spatial/social interaction effect at about 17%. Ignoring the residential self-selection effect would substantially overestimate densification effects on school mode choice and other travel choices, while ignoring the spatial/social interaction effects would underestimate densification effects.

In summary, methodologically speaking, the proposed spatial GHDM framework can be a valuable tool for modeling spatial dependencies in multidimensional mixed data outcomes that are becoming of increasing interest in several fields. Empirically speaking, the proposed framework allows for the better disentangling of true causal land use effects from spurious self-selection effects and spatial dependence effects, enabling more accurate policy impact assessment of land use-based policy instruments. We hope that the simple, parsimonious, and elegant way of introducing social/spatial dependence in multi-dimensional mixed models will contribute to empirical research in a variety of disciplines.

ACKNOWLEDGMENTS

This research was partially supported by the U.S. Department of Transportation through the Data-Supported Transportation Operations and Planning (D-STOP) Tier 1 University Transportation Center. The first author would like to acknowledge support from a Humboldt Research Award from the Alexander von Humboldt Foundation, Germany. The authors are grateful to Lisa Macias for her help in formatting this document. Two anonymous referees provided useful comments on an earlier version of this paper.

REFERENCES

Abraham, J., and Hunt, J. (1997). Specification and estimation of nested logit model of home, workplaces, and commuter mode choices by multiple-worker households. *Transportation Research Record: Journal of the Transportation Research Board*, 1606, 17-24.

Aditjandra, P.T., Cao, X.J., and Mulley, C. (2012). Understanding neighbourhood design impact on travel behaviour: An application of structural equations model to a British metropolitan data. *Transportation Research Part A*, 46(1), 22-32.

Alton, D., Adab, P., Roberts, L., and Barrett, T. (2007). Relationship between walking levels and perceptions of the local neighbourhood environment. *Archives of Disease in Childhood*, 92(1), 29-33.

Anselin, L. (2010). Thirty years of spatial econometrics. *Papers in Regional Science*, 89(1), 3-25.

Arbia, G. (2014). *A Primer for Spatial Econometrics*. Palgrave MacMillan, Basingstoke.

Bauman, A.E., Reis, R.S., Sallis, J.F., Wells, J.C., Loos, R.J., Martin, B.W., and Lancet Physical Activity Series Working Group. (2012). Correlates of physical activity: why are some people physically active and others not? *The Lancet*, 380(9838), 258-271.

Belcher, B.R., Berrigan, D., Dodd, K.W., Emken, B.A., Chou, C.P., and Spuijt-Metz, D. (2010). Physical activity in US youth: impact of race/ethnicity, age, gender, and weight status. *Medicine and Science in Sports and Exercise*, 42(12), 2211.

Beron, K.J., and Vijverberg, W.P.M., (2004). Probit in a spatial context: a Monte Carlo analysis. In: Anselin L, Florax RJGM, Rey SJ (eds) *Advances in Spatial Econometrics: Methodology, Tools and Applications*, 169-196, Springer-Verlag, Berlin.

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. (2014). The composite marginal likelihood (CML) inference approach with applications to discrete and mixed dependent variable models. *Foundations and Trends in Econometrics*, 7(1), 1-117.

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 new spatial (social) interaction discrete choice model accommodating for unobserved effects due to endogenous network formation. *Transportation*, 42(5), 879-914.

Bhat, C.R. (2015c). 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., 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 Sardesai, R. (2006). The impact of stop-making and travel time reliability on commute mode choice. *Transportation Research Part B*, 40(9), 709-730.

Bhat, C.R., Sen, S., and Eluru, N. (2009). The impact of demographics, built environment attributes, vehicle characteristics, and gasoline prices on household vehicle holdings and use. *Transportation Research Part B*, 43(1), 1-18.

Bhat, C.R., Sener, I.N., and Eluru, N. (2010). A flexible spatially dependent discrete choice model: formulation and application to teenagers' weekday recreational activity participation, *Transportation Research Part B*, 44(8-9), 903-921.

Bhat, C.R., Paleti, R., and Singh, P. (2014a). A spatial multivariate count model for firm location decisions. *Journal of Regional Science*, 54(3), 462-502.

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

Bhat, C.R., Astroza, S., Bhat, A.C., and Nagel, K. (2016). 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.

Broberg, A., and Sarjala, S. (2015). School travel mode choice and the characteristics of the urban built environment: The case of Helsinki, Finland. *Transport Policy*, 37, 1-10.

Brownstone, D., and Fang, H. (2014). A vehicle ownership and utilization choice model with endogenous residential density. *Journal of Transport and Land Use*, 7(2), 135-151.

Castro, M., Paleti, R., and Bhat, C.R. (2012). A latent variable representation of count data models to accommodate spatial and temporal dependence: application to predicting crash frequency at intersections. *Transportation Research Part B*, 46(1), 253-272.

Castro, M., Paleti, R., and Bhat, C.R. (2013). A spatial generalized ordered response model to examine highway crash injury severity, *Accident Analysis and Prevention*, 52, 188-203.

Copperman, R., and Bhat, C.R. (2007). An analysis of the determinants of children's weekend physical activity participation, *Transportation*, 34(1), 67-87.

Cutler, D.M., and Lleras-Muney, A. (2006). *Education and health: evaluating theories and evidence* (No. w12352). National Bureau of Economic Research.

De Leon, A.R., and Chough, K.C. (2013). *Analysis of Mixed Data: Methods & Applications*, CRC Press, Taylor & Francis Group, Boca Raton, FL.

De Leon, A.R., and Zhu, Y. (2008). ANOVA extensions for mixed discrete and continuous data. *Computational Statistics and Data Analysis*, 52(4), 2218-2227.

Dworkin, S.L., and Wachs, F.L. (2009). *Body Panic: Gender, health, and the selling of fitness*. New York University Press, New York.

Elhorst, J.P. (2010). Applied spatial econometrics: raising the bar. *Spatial Economic Analysis*, 5(1), 9-28.

Elhorst, J.P., Abreu, M., Amaral, P., Bhattacharjee, A., Corrado, L., Fingleton, B., Fuerst, F., Garretsen, H., Igliori, D., Le Gallo, J., McCann, P., Monastiriotis, V., Pryce, G., and Yu, J. (2016). Raising the bar (1), *Spatial Economic Analysis* 11(1), 1-6.

Engelsrud, G. (2009). Aerobic exercise and health-A tenuous connection? In *Normality/Normativity*, Folmarson, K.L. (Ed.), 155-186, Center for Gender Research, University of Uppsala.

Ewing, R., Schroeer, W., and Greene, W. (2004). School location and student travel analysis of factors affecting mode choice. *Transportation Research Record: Journal of the Transportation Research Board*, 1895, 55-63.

Ferdous, N., and Bhat, C.R. (2013). A spatial panel ordered-response model with application to the analysis of urban land-use development intensity patterns. *Journal of Geographical Systems*, 15(1), 1-29.

Franzese, R.J., Hays, J.C., and Schaffer, L. (2010). Spatial, temporal, and spatiotemporal autoregressive probit models of binary outcomes: estimation, interpretation, and presentation. American Political Science Association 2010 Annual Meeting papers, August.

Heitzler, C., Lytle, L., Erickson, D., Sirard, J., Barr-Anderson, D., and Story, M. (2011). Physical activity and sedentary activity patterns among children and adolescents: a latent class analysis approach. *Journal of Physical Activity and Health*, 8(4), 457.

Hoshino, T., and Bentler, P.M. (2013). Bias in factor score regression and a simple solution. In: De Leon, A.R., and Chough, K.C. (eds.), *Analysis of Mixed Data: Methods & Applications*, CRC Press, Taylor & Francis Group, Boca Raton, FL, 43-61.

Huitt, W. (2007). Maslow's hierarchy of needs. *Educational Psychology Interactive*. Valdosta State University, Valdosta, GA.

Kelly, J.A., and Fu, M. (2014). Sustainable school commuting–understanding choices and identifying opportunities: A case study in Dublin, Ireland. *Journal of Transport Geography*, 34, 221-230.

Klier, T. and McMillen, D.P. (2008). Clustering of auto supplier plants in the U.S.: GMM spatial logit for large samples. *ASA Journal of Business & Economic Statistics* 26(4), 460-471.

LaMondia, J.J., and Bhat, C.R. (2011). A study of visitors' leisure travel behavior in the northwest territories of Canada, *Transportation Letters: The International Journal of Transportation Research*, 3(1), 1-19.

LeSage, J.P., and Pace, R.K. (2009). *Introduction to Spatial Econometrics*. Chapman & Hall/CRC, Taylor & Francis Group, Boca Raton.

Lin, J. (2012). Calif. commute times rank 10th longest in US. *California Watch*, http://californiawatch.org/dailyreport/calif-commute-times-rank-10th-longest-us-18614. Accessed on August 17, 2015.

Maddala, G.S. (1983). *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge University Press, Cambridge, UK.

McDonald, N.C. (2008). Children's mode choice for the school trip: the role of distance and school location in walking to school. *Transportation*, 35(1), 23-35.

McLean, C.P., and Anderson, E.R. (2009). Brave men and timid women? A review of the gender differences in fear and anxiety. *Clinical Psychology Review*, 29(6), 496-505.

Narayanamoorthy, S., Paleti, R., and Bhat, C.R. (2013). On accommodating spatial dependence in bicycle and pedestrian injury counts by severity level, *Transportation Research Part B*, 55, 245-264

Orthner, D.K., Jones-Sanpei, H., and Williamson, S. (2004). The resilience and strengths of low-income families. *Family Relations*, 53(2), 159-167.

Paleti, R., Bhat, C.R., Pendyala, R.M., and Goulias, K.G. (2013a). Modeling of household vehicle type choice accommodating spatial dependence effects, *Transportation Research Record: Journal of the Transportation Research Board*, 2343, 86-94

Paleti, R., Bhat, C.R., and Pendyala, R.M. (2013b). Integrated model of residential location, work location, vehicle ownership, and commute tour characteristics. *Transportation Research Record: Journal of the Transportation Research Board*, 2382, 162-172.

Petrin, A. and Train, K. (2010). A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research*, 47(1), 3-13.

Pinjari, A.R., Pendyala, R.M., Bhat, C.R., and Waddell, P.A. (2011). Modeling the choice continuum: an integrated model of residential location, auto ownership, bicycle ownership, and commute tour mode choice decisions. *Transportation*, 38(6), 933-958.

Pinjari, A.R., Eluru, N., Bhat, C.R., Pendyala, R.M., and Spissu, E. (2008). Joint model of choice of residential neighborhood and bicycle ownership: accounting for self-selection and unobserved heterogeneity. *Transportation Research Record: Journal of the Transportation Research Board*, 2082, 17-26.

Pinkse, J. and Slade, M.E. (1998). Contracting in space: an application of spatial statistics to discrete-choice models. *Journal of Econometrics* 85(1), 125-154.

Puhani, P. (2000). The Heckman correction for sample selection and its critique. *Journal of Economic Surveys*, 14(1), 53-68.

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

Saffer, H., Dave, D.M., and Grossman, M. (2011). *Racial, ethnic and gender differences in physical activity*. NBER Working Paper No. 17413, National Bureau of Economic Research, Cambridge, MA.

Seccombe, K. (2002). "Beating the odds" versus "changing the odds": Poverty, resilience, and family policy. *Journal of Marriage and Family*, 64(2), 384-394.

Sener, I.N., and Bhat, C.R. (2007). An analysis of the social context of children's weekend discretionary activity participation. *Transportation*, 34(6), 697-721.

Sener, I.N., and Bhat, C.R. (2012). Modeling the spatial and temporal dimensions of recreational activity participation with a focus on physical activities. *Transportation*, 39(3), 627-656.

Sener, I.N., Copperman, R.B., Pendyala, R.M., and Bhat, C.R. (2008). An analysis of children's leisure activity engagement: examining the day of week, location, physical activity level, and fixity dimensions. *Transportation*, 35(5), 673-696.

Seraj, S., Sidharthan, R., Bhat, C.R., Pendyala, R.M., and Goulias, K.G. (2012). Parental attitudes toward children walking and bicycling to school. *Transportation Research Record: Journal of the Transportation Research Board*, 2323, 46-55.

Sidharthan, R., and Bhat, C.R. (2012). Incorporating spatial dynamics and temporal dependency in land use change models. *Geographical Analysis*, 44(4), 321-349.

Sidharthan, R., Bhat, C.R., Pendyala, R.M., and Goulias, K.G. (2011). Model for children's school travel mode choice: accounting for effects of spatial and social interaction. *Transportation Research Record: Journal of the Transportation Research Board*, 2213, 78-86.

Smirnov, O.A. (2010). Modeling spatial discrete choice. *Regional Science and Urban Economics*, 40(5), 292-298.

Sultana, S., and Weber, J. (2014). The nature of urban growth and the commuting transition: endless sprawl or a growth wave? *Urban Studies*, 51(3), 544-576.

Teixeira-Pinto, A., and Harezlak, J. (2013). Factorization and latent variable models for joint analysis of binary and continuous outcomes. In: De Leon, A.R., and Chough, K.C. (eds.), *Analysis of Mixed Data: Methods & Applications*, CRC Press, Taylor & Francis Group, Boca Raton, FL, 81-91.

Terza, J., Basu, A., and Rathouz, P. (2008). Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling. *Journal of Health Economics*, 27(3), 531-543.

Varin, C., and Vidoni, P. (2005). A note on composite likelihood inference and model selection. *Biometrika*, 92(3), 519-528.

Walker, J.L., and Li, J. (2007). Latent lifestyle preferences and household location decisions. *Journal of Geographical Systems*, 9(1), 77-101.

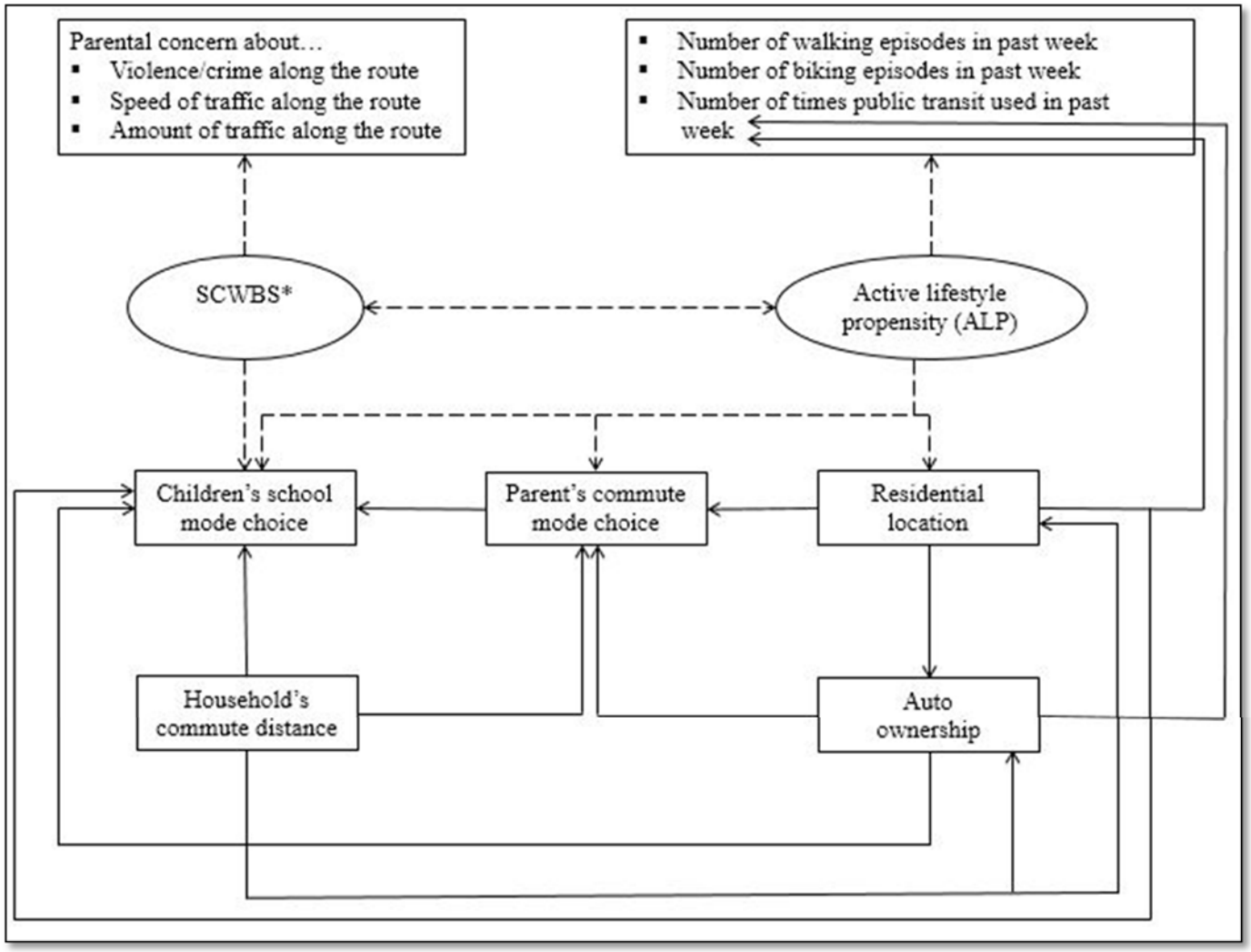
Wang, H., Iglesias, E.M., and Wooldridge, J.M. (2013). Partial maximum likelihood estimation of spatial probit models. *Journal of Econometrics*, 172(1), 77-89.

Wilson, E.J., Marshall, J., Wilson, R., and Krizek, K.J. (2010). By foot, bus or car: children's school travel and school choice policy. *Environment and Planning A*, 42(9), 2168-2185.

Yang, K., and Lee, L-F. (2015). Identification and QML estimation of multivariate and simultaneous spatial autoregressive models. Technical Report, http://kaiyang.weebly.com/uploads/5/0/7/4/5074896/revised_paper.pdf. Accessed August 7, 2016.

Yang, S., and Allenby, G.M. (2003). Modeling interdependent customer preferences. *Journal of Marketing Research*, XL, 282-294.

Yarlagadda, A.K., and Srinivasan, S. (2008). Modeling children's school travel mode and parental escort decisions. *Transportation*, 35(2), 201-218.



* Safety concern regarding children walking/bicycling to school

Figure 1: Conceptual diagram of structural relationships in the empirical model

Table 1: Descriptive statistics of dependent variables

Continuous Outcome		Ordinal Outcomes (Likert scale variables of parents' concern for children walking/bicycling to school or SCWBS)				Count Outcomes				
Average commute distance (miles)			Violence/ crime along the route to school	Speed of traffic along the route to school	Amount of traffic along the route to school		# of bicycling episodes in past week	# of walking episodes in past week	# of times public transit used in past week	Auto ownership
Statistics	Value	Category	Proportion	Proportion	Proportion	Value	Proportion	Proportion	Proportion	Proportion
Avg.	15.15	Not an issue	47.10%	16.82%	14.24%	0	50.88%	6.50%	66.73%	0.74%
Min.	0.11	A little bit of an issue	16.36%	11.15%	12.95%	1	8.71%	2.35%	6.87%	10.37%
Max.	65.00	Somewhat of an issue	13.46%	20.32%	19.36%	2	8.43%	3.18%	6.27%	49.45%
Std.	11.45	Very much an issue	7.83%	17.83%	18.94%	3	7.01%	3.05%	2.35%	24.89%
		A serious issue	15.25%	33.88%	34.51%	4	4.20%	3.96%	1.94%	9.49%
						>=5	20.77%	80.96%	15.84%	5.06%
						Max.	68.00	165.00	60.00	10.00

Unordered Outcomes				
Parent's commute mode	Non-auto mode used		Auto mode used	
	7.88%		92.12%	
Residential location (housing units / sq. mile)	Less than 1000 hhs/sq. mile	1000-1999 hhs/sq. mile	2000-3999 hhs/sq. mile	4000 or more hhs/sq. mile
	32.58%	27.23%	29.54%	10.65%
Children school mode	Car	Bus	Walk/bike	Other modes
	70.37%	9.73%	18.43%	1.47%

Table 2: Parameter estimates of structural equations for the latent constructs

Latent Constructs	Coefficient	T-stat
Safety consciousness with respect to children walking/bicycling to school (SCWBS)		
Age of the school going children (base: 5-10 years old)		
11-15 years old	-0.125	-2.27
16-18 years old	-0.110	-1.90
Gender of the school going children (base: boy)		
Girl	0.077	2.20
Education Status (base: fraction of adults (25 years or more) with some college degree or bachelor's degree in the household)		
Fraction of adults with high school degree or less in the household	-0.218	-2.10
Fraction of adults with graduate degree in the household	0.142	2.11
Household monthly income (base: 75K or more)		
Less than 25K	-0.369	-2.46
25K to 74,999	-0.192	-2.63
Active lifestyle propensity (ALP)		
Race (base: Caucasian or African-American or others)		
Asian	-0.393	-3.93
Hispanic	-0.201	-2.01
Age (base: fraction of adults in the age group 31 or above in the household)		
Fraction of adults in the age group 19-30 years in the household	0.107	2.89
Education Status (base: fraction of adults (25 years or more) with some college degree or less in the household)		
Fraction of adults with high bachelor's degree or less in the household	0.132	3.21
Fraction of adults with graduate degree in the household	0.184	2.14
Number of children in different age groups in the household		
Less than 10 years old	0.143	4.67
11-15 years old	0.154	2.54
16-18 years old	-0.063	-1.63
Correlation between the two latent constructs	0.078	1.67
Spatial autoregressive parameter for the latent construct SCWBS	0.447	2.98
Spatial autoregressive parameter for the latent construct ALP	0.846	5.60

Table 3: Parameter estimates of latent construct loadings on endogenous variables

Dependent variables	Latent Constructs			
	SCWBS		*ALP*	
	Coeff	T-stat	Coeff	T-stat
Walk/bike issue				
Violence/crime along the route	0.131	15.10	----	----
Speed of traffic along the route	2.772	15.57	----	----
Amount of traffic along the route	2.646	16.54	----	----
# bicycling episodes in past week	----	----	0.857	11.74
# walking episodes in past week	----	----	2.065	17.21
# of times public transit used in past week	----	----	0.181	5.86
Residential location (base: less than 1000 hh./sq. mile)				
1000-1999 hh./sq. mile	----	----	0.103	2.34
2000-3999 hh./sq. mile	----	----	0.060	1.90
4000 or more hh./sq. mile	----	----	0.066	2.00
At least one commuter uses public transit/walk/bicycle for commuting	----	----	0.102	4.64
Children's school mode choice (base: Car)				
Bus	-0.308*	-2.26	0.207	3.23
Walk/bicycle	-0.195*	-2.57	0.122	2.07
Other	----	----	----	----
Household average commute distance (miles)	----	----	----	----
Auto ownership	----	----	----	----

***** These coefficients are only for households with distance to school greater than 2 miles. See Table 5 for the loadings of the latent construct SCWBS for different bands of home-to-school distance. All these loadings are negative in sign.

Table 4: Parameters estimates of inter-relationships among endogenous variables

Dependent variables	Natural logarithm of household average commute distance (miles)		Residential location								Each adult with driver license has access to at least one auto*		At least one commuter uses non-auto modes for commuting	
			Less than 1000 hh./sq. mile		1000-1999 hh./sq. mile		2000-3999 hh./sq. mile		4000 or more hh./sq. mile					
	Coeff	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**
Residential location (base: less than 1000 hh./ sq. mile)														
1000-1999 hh./sq. mile	-0.065	-3.25	----	----	----	----	----	----	----	----	----	----	----	----
2000-3999 hh./sq. mile	-0.065	-3.25	----	----	----	----	----	----	----	----	----	----	----	----
4000 or more hh./sq. mile	-0.065	-3.25	----	----	----	----	----	----	----	----	----	----	----	----
Auto ownership	0.045	1.92	----	----	-0.066	-2.10	-0.066	-2.10	-0.166	-2.32	----	----	----	----
At least one commuter uses non-auto modes for commuting	-0.214	-2.10	----	----	----	----	0.131	8.73	0.574	8.85	-1.013	-3.58	----	----
Children school mode (base: car)														
Bus	-0.080	-4.38	0.164	6.07	----	----	----	----	0.098	2.44	-0.319	-2.53	0.291	4.48
Walk/bike	-0.080	-4.38	----	----	----	----	0.176	7.04	0.176	7.04	-0.134	-2.31	0.224	2.73
Others	----	----	----	----	----	----	----	----	----	----	----	----	----	----
Number of times public transit used in past week	----	----	----	----	0.135	2.49	0.135	2.49	1.348	2.83	-1.718	-5.88	----	----

* The auto ownership variable was translated to auto availability per licensed driver in the household.

Table 5: Parameter estimates of the school mode choice component

Explanatory Variables	Children's school mode (base: Car)					
	Bus		Walk/bicycle		Other modes	
	Coeff	T-stat	Coeff	T-stat	Coeff	T-stat
Exogenous variables						
Constants	-1.456	-9.22	-1.167	-3.64	-1.965	-4.94
# workers with work from home option	-0.091	-4.33	----	----	----	----
# workers with flexible work timings	-0.022	-1.57	-0.052	-5.78	----	----
Distance to school						
Less than ¼ mile	----	----	0.498	9.40	----	----
¼ mile to ½ mile	----	----	0.403	7.90	----	----
½ mile to 1 mile	----	----	0.403	7.90	----	----
1 mile to 2 miles	----	----	0.185	4.40	----	----
More than 2 miles	0.286	5.20	----	----	----	----
Latent constructs						
SCWBS (base distance = greater than 2 miles)	-0.308	-2.26	-0.195	-2.57	----	----
SCWBS * Distance to school less than ¼ mile	----	----	0.075	5.37	----	----
SCWBS * Distance to school between ¼ mile and 1 mile	----	----	0.037	2.84	----	----
SCWBS * Distance to school between 1 mile and 2 miles	----	----	0.029	1.93	----	----
Active lifestyle propensity (ALP)	0.207	3.23	0.122	2.07	----	----
Endogenous variables						
At least one commuter uses public transit/walk/bicycle for commuting	0.291	4.48	0.224	2.73	----	----
Natural logarithm of household average commute distance	-0.080	-4.38	-0.080	-4.38	----	----
Each adult with driver license has access to at least one vehicle	-0.319	-2.53	-0.134	-2.31	----	----
Residential location						
Density less than 1000 hh./sq. mile	0.164	6.07	----	----	----	----
1000-1999 hh./sq. mile	----	----	----	----	----	----
2000-3999 hh./sq. mile	----	----	0.176	7.04	----	----
4000 or more hh./sq. mile	0.098	2.44	0.176	7.04	----	----

Table 6: Average treatment effect (ATE) on children's school mode choice of transplanting a random household from the lowest density (less than 1000 hh./sq. mile) residential location to the highest density (4000 or more hh./sq. mile) location

Variable	IHDM model	Aspatial-GHDM model	Spatial-GHDM model	RSS	True causal effect	SSD
Car	-0.082 (0.026)	-0.029 (0.015)	-0.044 (0.014)	46	35	18
Bus	-0.051 (0.018)	-0.021 (0.011)	-0.029 (0.016)	43	41	16
Walk/bike	0.142 (0.036)	0.053 (0.012)	0.076 (0.019)	47	37	16
Other modes	-0.009 (0.008)	-0.003 (0.002)	-0.003 (0.001)	67	33	0

Note: Standard errors are reported in parentheses.

Appendix A: Model Formulation

Let *h* be the index for continuous outcomes $(h = 1, 2, ..., H)$. Then the continuous variable y_{qh} can be written in the usual linear regression fashion as follows:

$$y_{qh} = \boldsymbol{\gamma}_h' \boldsymbol{x}_q + \boldsymbol{d}_h' \boldsymbol{z}_q^* + \varepsilon_{qh} \quad \text{(A.1)}$$

where $\boldsymbol{x}_q$ is an $(A \times 1)$ vector of exogenous variables (including a constant) as well as possibly the observed values of other endogenous variables (continuous, ordinal, count variable, and nominal variables (introduced as dummy variables)), $\boldsymbol{\gamma}_h$ is the corresponding vector of coefficients, $\boldsymbol{d}_h$ is an $(L \times 1)$ vector of latent variable loadings on the h^{th} continuous outcome, and ε_{qh} is a normally distributed random error term. Next, define the following notations to write Equation (A.1) in a compact, matrix form for individual *q*.

$\mathbf{y}_q = (y_{q1}, y_{q2}, ..., y_{qH})'$ $[(H \times 1)$ vector$]$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1', \boldsymbol{\gamma}_2', ..., \boldsymbol{\gamma}_H')'$ $[(H \times A)$ matrix$]$,
$\boldsymbol{d} = (\boldsymbol{d}_1', \boldsymbol{d}_2', ..., \boldsymbol{d}_H')'$ $[(H \times L)$ matrix$]$, and $\boldsymbol{\varepsilon}_q = (\varepsilon_{q1}, \varepsilon_{q2}, ..., \varepsilon_{qH})'[(H \times 1)$ vector$]$.

Now, Equation (A.1) may be written in matrix form for individual *q* as follows:

$$\boldsymbol{y}_q = \boldsymbol{\gamma} \boldsymbol{x}_q + \boldsymbol{d} \boldsymbol{z}_q^* + \boldsymbol{\varepsilon}_q . \quad \text{(A.2)}$$

We assume a diagonal MVN distribution for $\boldsymbol{\varepsilon}_q$: $\boldsymbol{\varepsilon}_q \sim \text{MVN}_H(\mathbf{0}_H, \boldsymbol{\Sigma})$. The non-diagonal elements of $\boldsymbol{\varepsilon}_q$ are assumed to be zero for identification purposes. Also, the $\boldsymbol{\varepsilon}_q$ terms across different individuals are assumed independent of each other.

Next, consider *N* ordinal outcomes (indicators) and let *n* be an index for ordinal outcomes $(n = 1, 2, ..., N)$. Also, let J_n be the number of categories for the n^{th} ordinal outcome ($J_n \geq 2$) and let the corresponding index be j_n $(j_n = 1, 2, ..., J_n)$. Let $\widetilde{y}_{qn}^*$ be the latent underlying variable whose horizontal partitioning leads to the observed outcome a_{qn} for the q^{th} individual's n^{th} ordinal variable. Then, in the usual ordered response formulation, for the individual *q*, we may write:

$$\widetilde{y}_{qn}^* = \widetilde{\boldsymbol{\gamma}}_n' \boldsymbol{x}_q + \widetilde{\boldsymbol{d}}_n' \boldsymbol{z}_q^* + \widetilde{\varepsilon}_{qn}, \ \widetilde{\psi}_{q,n,a_{qn}-1} < \widetilde{y}_{qn}^* < \widetilde{\psi}_{q,n,a_{qn}} \quad \text{(A.3)}$$

where $\boldsymbol{x}_q$ is as defined earlier, $\widetilde{y}_{qn}$ is the ordinal variable outcome category, $\widetilde{\boldsymbol{\gamma}}_n$ is the corresponding vector of coefficients, $\widetilde{\boldsymbol{d}}_n$ is an $(L \times 1)$ vector of latent variable loadings on the n^{th} ordinal outcome, and $\widetilde{\varepsilon}_{qn}$ is a normally distributed random error term. For each ordinal outcome, $\widetilde{\psi}_{q,n,0} < \widetilde{\psi}_{q,n,1} < \widetilde{\psi}_{q,n,2} ... < \widetilde{\psi}_{q,n,J_n-1} < \widetilde{\psi}_{q,n,J_n}$; $\widetilde{\psi}_{q,n,0} = -\infty$, $\widetilde{\psi}_{q,n,1} = 0$, and $\widetilde{\psi}_{q,n,J_n} = +\infty$. Next, define the following notation to write Equation (A.3) in a compact matrix form for individual *q*.

$\widetilde{\boldsymbol{y}}_q^* = (\widetilde{y}_{q1}^*, \widetilde{y}_{q2}^*, ..., \widetilde{y}_{qN}^*)'$ $\left[(N \times 1) \text{ vector}\right]$, $\widetilde{\boldsymbol{\gamma}} = (\widetilde{\boldsymbol{\gamma}}_1', \widetilde{\boldsymbol{\gamma}}_2', ..., \widetilde{\boldsymbol{\gamma}}_N')'$ $\left[(N \times A) \text{ matrix}\right]$,
$\widetilde{\boldsymbol{d}} = (\widetilde{\boldsymbol{d}}_1', \widetilde{\boldsymbol{d}}_2', ..., \widetilde{\boldsymbol{d}}_N')'$ $\left[(N \times L) \text{ matrix}\right]$, $\widetilde{\boldsymbol{\varepsilon}}_q = (\widetilde{\varepsilon}_{q1}, \widetilde{\varepsilon}_{q2}, ..., \widetilde{\varepsilon}_{qN})'$ $\left[(N \times 1) \text{ vector}\right]$.

Also, stack the lower thresholds for the observed outcomes a_{qn} of individual *q* $\widetilde{\psi}_{q,n,a_{qn}-1} (n = 1, 2, ..., N)$ into an $(N \times 1)$ vector $\widetilde{\boldsymbol{\psi}}_{q,low}$ and the corresponding upper thresholds $\widetilde{\psi}_{q,n,a_{qn}} (n = 1, 2, ..., N)$ into another vector $\widetilde{\boldsymbol{\psi}}_{q,up}$.

Now, Equation (A.3) may be written in matrix form for individual q as follows:

$$\widetilde{\boldsymbol{y}}_q^* = \widetilde{\boldsymbol{\gamma}}\boldsymbol{x}_q + \widetilde{\boldsymbol{d}}\boldsymbol{z}_q^* + \widetilde{\boldsymbol{\varepsilon}}_q,\ \ \widetilde{\boldsymbol{\psi}}_{q,low} < \widetilde{\boldsymbol{y}}_q^* < \widetilde{\boldsymbol{\psi}}_{q,up}\,. \tag{A.4}$$

For identification, we assume a diagonal multivariate normal distribution for $\widetilde{\boldsymbol{\varepsilon}}_q$ with all the diagonal elements equal to unity: $\widetilde{\boldsymbol{\varepsilon}}_q \sim \text{MVN}_N(\boldsymbol{0}_N, \textbf{IDEN}_N)$. In addition, the $\widetilde{\boldsymbol{\varepsilon}}_q$ terms are assumed to be independent across individuals.

Let there be C count variables and let c be an index for count outcomes $(c = 1, 2, ..., C)$. Let k_c be the index for count value $(k_c = 0, 1, 2, ..., \infty)$ and let r_{qc} be the actual observed count value. Then, following the recasting of a count model in a generalized ordered-response probit formulation (see Bhat, 2015a), a generalized version of the negative binomial count model may be written as:

$$\breve{y}_{qc}^* = \breve{\boldsymbol{d}}_c'\boldsymbol{z}_q^* + \breve{\varepsilon}_{qc},\ \breve{\psi}_{q,c,r_{qc}-1} < \breve{y}_{qc}^* < \breve{\psi}_{q,c,r_{qc}}\,, \tag{A.5}$$

$$\breve{\psi}_{q,c,r_c} = \Phi^{-1}\left[\frac{\left(1-\upsilon_{qc}\right)^{\theta_c}}{\Gamma(\theta_c)}\sum_{t=0}^{r_c}\left(\frac{\Gamma(\theta_c+t)}{t!}(\upsilon_{qc})^t\right)\right] + \varphi_{c,r_c},\ \upsilon_{qc} = \frac{\lambda_{qc}}{\lambda_{qc}+\theta_{qc}},\ \text{and}\ \lambda_{qc} = e^{\breve{\boldsymbol{\gamma}}_c'\boldsymbol{x}_q}\,. \tag{A.6}$$

In the above equation, $\breve{y}_{qc}^*$ is a latent continuous stochastic propensity variable associated with the count variable c that maps into the observed count r_{qc} through the $\breve{\boldsymbol{\psi}}_{q,c}$ vector (which is a vertically stacked column vector of thresholds $(\breve{\psi}_{q,c,-1}, \breve{\psi}_{q,c,0}, \breve{\psi}_{q,c,1}, \breve{\psi}_{q,c,2}, ...)'$). $\breve{\boldsymbol{d}}_c$ is a $(L\times 1)$ vector of latent variable loadings on the c^{th} count outcome, and $\breve{\varepsilon}_{qc}$ is a standard normal random error term. $\breve{\boldsymbol{\gamma}}_c$ is a column vector of coefficients corresponding to the vector $\boldsymbol{x}_q$. θ_c is a parameter that provides flexibility to the count formulation, and is related to the dispersion parameter in a traditional negative binomial model $(\theta_c > 0 \,\forall\, c)$. $\Gamma(\theta_c)$ is the traditional gamma function; $\Gamma(\theta_c) = \int_{\tilde{t}=0}^{\infty} \tilde{t}^{\,\theta_c - 1} e^{-\tilde{t}}\, d\tilde{t}$. The threshold terms in the $\breve{\boldsymbol{\psi}}_{q,c}$ vector satisfy the ordering condition (*i.e.*, $\breve{\psi}_{q,c,-1} < \breve{\psi}_{q,c,0} < \breve{\psi}_{q,c,1} < \breve{\psi}_{q,c,2} .... < \infty \ \forall\ c$) as long as $\varphi_{c,-1} < \varphi_{c,0} < \varphi_{c,1} < \varphi_{c,2} .... < \infty$. The φ_c terms in the thresholds provide flexibility to accommodate high or low probability masses for specific count outcomes. For identification, we set $\varphi_{c,-1} = -\infty$ and $\varphi_{c,0} = 0$ for all count variables c. In addition, based on empirical testing, we identify a count value e_c^* $(e_c^* \in \{0,1,2,.....\})$ above which φ_{c,k_c} $(k_c \in \{1,2,.....\})$ is held fixed at φ_{c,e_c^*}. Doing so allows the count model to predict beyond the range available in the estimation sample. For later use, let $\boldsymbol{\varphi}_c = (\varphi_{c,1}, \varphi_{c,2}, \ldots \varphi_{c,e_c^*})'$ $(e_c^* \times 1 \text{ vector})$ (assuming $e_c^* > 0$), $\boldsymbol{\varphi} = (\boldsymbol{\varphi}_1', \boldsymbol{\varphi}_2', \ldots \boldsymbol{\varphi}_C')'$ $\left[\left(\sum_c e_c^*\right) \times 1 \text{ vector}\right]$, and $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_C)'$ [C $\times$ 1 vector]. Next define the following notation: $\breve{\boldsymbol{y}}_q^* = (\breve{y}_{q1}^*, \breve{y}_{q2}^*, ..., \breve{y}_{qC}^*)'$ $[(C\times 1) \text{ vector}]$, $\breve{\boldsymbol{d}} = (\breve{\boldsymbol{d}}_1', \breve{\boldsymbol{d}}_2', ..., \breve{\boldsymbol{d}}_C')'$ $[(C \times L) \text{ matrix}]$, $\breve{\boldsymbol{\gamma}} = (\breve{\boldsymbol{\gamma}}_1', \breve{\boldsymbol{\gamma}}_2', ..., \breve{\boldsymbol{\gamma}}_C')'$ $[(C \times A) \text{ matrix}]$, $\breve{\boldsymbol{\varepsilon}}_q = (\breve{\varepsilon}_{q1}, \breve{\varepsilon}_{q2}, ..., \breve{\varepsilon}_{qC})'$ $[(C\times 1) \text{ vector}]$. Also, stack the lower

thresholds of observed counts for the indiviudal q $\breve{\psi}_{q,c,r_{qc}-1}(c=1,2,...,C)$ into a (C×1) vector $\breve{\boldsymbol{\psi}}_{q,low}$ and the upper thresholds $\breve{\psi}_{q,c,r_{qc}}(c=1,2,...,C)$ into another vector $\breve{\boldsymbol{\psi}}_{q,up}$. Now, the latent propensity underlying the count outcomes in Equation (A.5) may be written in matrix form as:

$$\breve{\boldsymbol{y}}_q^* = \breve{\boldsymbol{d}}\boldsymbol{z}_q^* + \breve{\boldsymbol{\varepsilon}}_q, \quad \breve{\boldsymbol{\psi}}_{q,low} < \breve{\boldsymbol{y}}_q^* < \breve{\boldsymbol{\psi}}_{q,up} \tag{A.7}$$

Similar to ordinal variables we assume that the $\breve{\boldsymbol{\varepsilon}}_q$ terms are distributed as follows: $\breve{\boldsymbol{\varepsilon}}_q \sim \text{MVN}_C(\mathbf{0}_C, \mathbf{IDEN}_C)$, with independency across individuals.

Finally, let there be G nominal (unordered-response) variables, and let g be the index for the nominal variables (g = 1, 2, 3,…, G). Also, let I_g be the number of alternatives corresponding to the g^{th} nominal variable ($I_g \geq 3$) and let i_g be the corresponding index (i_g= 1, 2, 3,…, I_g). Consider the g^{th} nominal variable and assume that the individual q chooses the alternative $m_{q,g}$. Also, assume the usual random utility structure for each alternative i_g.

$$U_{qgi_g} = \boldsymbol{b}'_{gi_g}\boldsymbol{x}_q + \boldsymbol{\vartheta}'_{gi_g}(\boldsymbol{\beta}_{gi_g}\boldsymbol{z}_q^*) + \varsigma_{qgi_g}, \tag{A.8}$$

where $\boldsymbol{x}_q$ is as defined earlier, $\boldsymbol{b}_{gi_g}$ is a (A×1) column vector of corresponding coefficients, and ς_{qgi_g} is a normal error term. $\boldsymbol{\beta}_{gi_g}$ is a $(N_{gi_g} \times L)$-matrix of variables interacting with latent variables to influence the utility of alternative i_g, and $\boldsymbol{\vartheta}_{gi_g}$ is a $(N_{gi_g} \times 1)$-column vector of coefficients capturing the effects of latent variables and its interaction effects with other exogenous variables. Let $\boldsymbol{\varsigma}_{\boldsymbol{qg}} = (\varsigma_{qg1}, \varsigma_{qg2}, ...\varsigma_{qgI_g})'$ (I_g×1 vector), with $\boldsymbol{\varsigma}_{\boldsymbol{qg}} \sim MVN_{I_g}(\mathbf{0}, \boldsymbol{\Lambda}_{\mathbf{g}})$ and independent across individuals. Taking the difference with respect to the first alternative, the only estimable elements correspond to the covariance matrix $\breve{\boldsymbol{\Lambda}}_g$ of these error differences, $\breve{\boldsymbol{\varsigma}}_{\boldsymbol{qg}} = (\breve{\varsigma}_{qg2}, \breve{\varsigma}_{qg3}, ..., \breve{\varsigma}_{qgI_g})$ (where $\breve{\varsigma}_{qgi} = \breve{\varsigma}_{qgi} - \breve{\varsigma}_{qg1}, \forall i \neq 1$). Further, the variance term at the top left diagonal of $\breve{\boldsymbol{\Lambda}}_g$ (g=1,2,…,G) is set to 1 to account for scale invariance. $\boldsymbol{\Lambda}_{\mathbf{g}}$ is constructed from $\breve{\boldsymbol{\Lambda}}_{\mathbf{g}}$ by adding a row of zeros on top and a column of zeros to the left. To proceed, define $\boldsymbol{U}_{qg} = (U_{qg1}, U_{qg2}, ..., U_{qgI_g})'$ (I_g×1 vector), $\boldsymbol{b}_g = (\boldsymbol{b}_{g1}, \boldsymbol{b}_{g2}, \boldsymbol{b}_{g3}, ..., \boldsymbol{b}_{gI_g})'$ (I_g×A matrix), and $\boldsymbol{\beta}_g = (\boldsymbol{\beta}'_{g1}, \boldsymbol{\beta}'_{g2}, ..., \boldsymbol{\beta}'_{gI_g})'$ $\left(\sum_{i_g=1}^{I_g} N_{gi_g} \times L\right)$ matrix. Also, define the $\left(I_g \times \sum_{i_g=1}^{I_g} N_{gi_g}\right)$ matrix $\boldsymbol{\vartheta}_{\boldsymbol{g}}$, 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×L matrix), $\vec{G} = \sum_{g=1}^{G} I_g$, $\widetilde{G} = \sum_{g=1}^{G} (I_g - 1)$, $\boldsymbol{U}_{\boldsymbol{q}} = \left(\boldsymbol{U}'_{q1}, \boldsymbol{U}'_{q2}, ..., \boldsymbol{U}'_{qG}\right)'$ ($\vec{G}$×1 vector), $\boldsymbol{\varsigma}_q = (\varsigma_{q1}, \varsigma_{q2}, ..., \varsigma_{qG})'$ ($\vec{G}$×1 vector), $\boldsymbol{b} = (\boldsymbol{b}'_1, \boldsymbol{b}'_2, ..., \boldsymbol{b}'_G)'$ ($\vec{G} \times A$ matrix), $\boldsymbol{\varpi} = (\boldsymbol{\varpi}'_1, \boldsymbol{\varpi}'_2, ..., \boldsymbol{\varpi}'_G)'$ ($\vec{G} \times L$ matrix), and $\boldsymbol{\vartheta} = \text{Vech}(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2, ..., \boldsymbol{\vartheta}_G)$ (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 (A.8) for individual q as:

$$\boldsymbol{U}_q = \boldsymbol{b}\boldsymbol{x}_q + \boldsymbol{\varpi}\, \boldsymbol{z}_q^* + \boldsymbol{\varsigma}_q, \tag{A.9}$$

where $\boldsymbol{\varsigma}_q \sim \text{MVN}_{\vec{G}}(\mathbf{0}_{\vec{G}}, \boldsymbol{\Lambda})$. As earlier, to ensure identification, we specify $\boldsymbol{\Lambda}$ as follows:

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_2 & \mathbf{0} & \mathbf{0} \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Lambda}_3 & \mathbf{0} \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \ \ldots \ \vdots & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \ \ldots & \boldsymbol{\Lambda}_G \end{bmatrix} (\vec{G} \times \vec{G} \text{ matrix}) \tag{A.10}$$

Reduced Form Model System

Let $E = (H + N + C)$ and $\widetilde{E} = (N + C + \widetilde{G})$. Define $\vec{\boldsymbol{y}}_q = \left(\boldsymbol{y}_q', \left[\widetilde{\boldsymbol{y}}_q^{\,*}\right], \left[\breve{\boldsymbol{y}}_q^{\,*}\right] \right)'$ [$E \times 1$ vector], $\vec{\boldsymbol{\gamma}} = (\boldsymbol{\gamma}', \widetilde{\boldsymbol{\gamma}}', \mathbf{0}_{AC})'$ [$E \times A$ matrix], $\vec{\boldsymbol{d}} = (\boldsymbol{d}', \widetilde{\boldsymbol{d}}', \breve{\boldsymbol{d}}')'$ [$E \times L$ matrix], and $\vec{\boldsymbol{\varepsilon}}_q = (\boldsymbol{\varepsilon}_q', \widetilde{\boldsymbol{\varepsilon}}_q', \breve{\boldsymbol{\varepsilon}}_q')'$ [$E \times 1$ vector], where $\mathbf{0}_{AC}$ is a matrix of zeros of dimension $A \times C$. Then, the equations for continuous, ordinal, and count endogenous variables (*i.e.*, Equations A.2, A.4, and A.7) of individual q may be brought together as follows:

$$\vec{\boldsymbol{y}}_q = \vec{\boldsymbol{\gamma}}\boldsymbol{x}_q + \vec{\boldsymbol{d}}\boldsymbol{z}_q^* + \vec{\boldsymbol{\varepsilon}}_q, \text{ with } \text{Var}(\vec{\boldsymbol{\varepsilon}}_q) = \vec{\boldsymbol{\Sigma}} = \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{IDEN}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{IDEN}_C \end{bmatrix} (E \times E \text{ matrix}) \tag{A.11}$$

To combine the above equation with Equation (A.9) for nominal endogenous variables ($\boldsymbol{U}_q$), define $(\boldsymbol{yU})_q = \left[\vec{\boldsymbol{y}}_q{}', \boldsymbol{U}_q' \right]'$ [$(E + \vec{\text{G}}) \times 1$vector], $\vec{\boldsymbol{b}} = (\vec{\boldsymbol{\gamma}}', \boldsymbol{b}')'$ [$(E + \vec{\text{G}}) \times A$ matrix], $\vec{\boldsymbol{c}} = (\vec{\boldsymbol{d}}', \boldsymbol{\varpi}')'$ [$(E + \vec{\text{G}}) \times L$ matrix], and $\boldsymbol{\xi}_q = (\vec{\boldsymbol{\varepsilon}}_q', \boldsymbol{\varsigma}_q')'$ [$(E + \vec{\text{G}}) \times 1$ vector]. Then, the equations for all endogenous variables in the overall model system for individual q may be written compactly as:

$$(\boldsymbol{yU})_q = \vec{\boldsymbol{b}}\boldsymbol{x}_q + \vec{\boldsymbol{c}}\, \boldsymbol{z}_q^* + \boldsymbol{\xi}_q, \text{with } \text{Var}(\boldsymbol{\xi}_q) = \vec{\boldsymbol{\Sigma}} = \begin{bmatrix} \vec{\boldsymbol{\Sigma}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda} \end{bmatrix} [(E + \vec{G}) \times (E + \vec{G}) \text{ matrix}] \tag{A.12}$$

This appears as Equation (3) in the main body of the paper.

Appendix B: Estimation Methodology

Let $\boldsymbol{\lambda}$ be the collection of parameters to be estimated: $\boldsymbol{\lambda} = [\text{Vech}(\boldsymbol{\alpha}), \text{Vech}(\vec{\boldsymbol{\Sigma}}), \text{Vech}(\vec{\boldsymbol{b}}), \text{Vech}(\vec{\boldsymbol{c}}),\ \boldsymbol{\varphi},\ \boldsymbol{\theta},\ \boldsymbol{\delta}]$, where the operator "Vech(.)" vectorizes all the elements of the matrix/vector on which it operates. The identification issues pertaining to the estimability of these parameters in the current spatial-GHDM are the same as those discussed in Bhat (2015a) for the aspatial-GHDM, with the addition of the requirement that all elements of the vector $\boldsymbol{\delta}$ should be bounded in magnitude by the value of 1 (see Sidharthan and Bhat, 2012).

To estimate the model, we work with the latent utility differentials $u_{qgi_g m_{qg}} = \left(U_{qgi_g} - U_{qgm_{qg}}\right)$ of all non-chosen alternatives ($i_g \neq m_{qg}$) with respect to the chosen alternative (m_{qg}) for each nominal variable g and each individual q. Stack the utility differentials into a vector $\boldsymbol{u}_{qg} = \left[\left(u_{qg1m_{qg}}, u_{qg2m_{qg}}, ..., u_{qgI_g m_{qg}}\right)'; i_g \neq m_g\right]$ and then into $\boldsymbol{u}_q = \left(\left[\boldsymbol{u}_{q1}\right]', \left[\boldsymbol{u}_{q2}\right]', ..., \left[\boldsymbol{u}_{qG}\right]'\right)'$. Also, define $(\boldsymbol{yu})_q = \left[\vec{\boldsymbol{y}}_q', \boldsymbol{u}_q'\right]'$ $[(E+\widetilde{G})\times 1 \text{ vector}]$ and $\boldsymbol{yu} = \left[(\boldsymbol{yu})_1', (\boldsymbol{yu})_2', ..., (\boldsymbol{yu})_Q'\right]'$ $[Q(E+\widetilde{G})\times 1 \text{ vector}]$. The distribution of the vector $\boldsymbol{yu}$ may be developed from that of $\boldsymbol{yU}$ using a matrix $\mathbf{M}$ of size $[Q(E+\widetilde{G})\times Q(E+\vec{G})]$, constructed as discussed in Section 1 of the online supplement to this paper (see the online supplement at: http://www.caee.utexas.edu/prof/bhat/ABSTRACTS/Spatial_GHDM/online_supplement.pdf). Then the resulting distribution is $\boldsymbol{yu} \sim \text{MVN}_{Q(E+\widetilde{G})}\left[\widetilde{\boldsymbol{B}}, \widetilde{\boldsymbol{\Omega}}\right]$, where $\widetilde{\boldsymbol{B}} = \mathbf{M}(\vec{\boldsymbol{b}}\boldsymbol{x} + \vec{\boldsymbol{c}}\boldsymbol{B})$ and $\widetilde{\boldsymbol{\Omega}} = \mathbf{M}(\vec{\boldsymbol{c}}\boldsymbol{\Xi}\vec{\boldsymbol{c}}' + \mathbf{IDEN}_Q \otimes \vec{\boldsymbol{\Sigma}})\mathbf{M}'$.

Next, partition $\boldsymbol{yu}$ into two components – one that corresponds to all the continuous variables ($\boldsymbol{y}$) and the other that corresponds to all the ordinal, count, and nominal variables ($\widetilde{\boldsymbol{y}}^*, \breve{\boldsymbol{y}}^*, \boldsymbol{u}$ (utility differences)). That is, $\boldsymbol{yu} = (\boldsymbol{y}', \widetilde{\boldsymbol{u}}')'$, where $\widetilde{\boldsymbol{u}} = \left(\widetilde{\boldsymbol{y}}^{*\prime}, \breve{\boldsymbol{y}}^{*\prime}, \mathbf{u}'\right)'$. Accordingly, the mean vector $\widetilde{\boldsymbol{B}}$ and the variance matrix $\widetilde{\boldsymbol{\Omega}}$ of $\boldsymbol{yu}$ can also be appropriately partitioned as: $\widetilde{\boldsymbol{B}} = \begin{bmatrix} \widetilde{\boldsymbol{B}}_y \\ \widetilde{\boldsymbol{B}}_{\widetilde{u}} \end{bmatrix}$ and $\widetilde{\boldsymbol{\Omega}} = \begin{bmatrix} \widetilde{\boldsymbol{\Omega}}_y & \widetilde{\boldsymbol{\Omega}}_{y\widetilde{u}} \\ \widetilde{\boldsymbol{\Omega}}_{y\widetilde{u}}' & \widetilde{\boldsymbol{\Omega}}_{\widetilde{u}} \end{bmatrix}$ (see Section 1.2 of the online supplement).

One may develop the likelihood function by decomposing the joint distribution of $\boldsymbol{yu} = (\boldsymbol{y}', \widetilde{\boldsymbol{u}}')'$ into a product of marginal and conditional distributions. Specifically, the conditional distribution of $\widetilde{\boldsymbol{u}}$, given $\boldsymbol{y}$, is MVN with mean $\vec{\boldsymbol{B}}_{\widetilde{u}} = \widetilde{\boldsymbol{B}}_{\widetilde{u}} + \widetilde{\boldsymbol{\Omega}}_{y\widetilde{u}}'\widetilde{\boldsymbol{\Omega}}_y^{-1}(\boldsymbol{y} - \widetilde{\mathbf{B}}_y)$ and variance $\vec{\boldsymbol{\Omega}}_{\widetilde{u}} = \widetilde{\boldsymbol{\Omega}}_{\widetilde{u}} - \widetilde{\boldsymbol{\Omega}}_{y\widetilde{u}}'\widetilde{\boldsymbol{\Omega}}_y^{-1}\widetilde{\boldsymbol{\Omega}}_{y\widetilde{u}}$. Furthermore, define the threshold vectors as: $\vec{\boldsymbol{\psi}}_{low} = \left[\widetilde{\boldsymbol{\psi}}_{low}', \breve{\boldsymbol{\psi}}_{low}', \left(-\boldsymbol{\infty}_{Q\widetilde{G}}\right)'\right]'$ $(Q\widetilde{E}\times 1 \text{ vector})$ and $\vec{\boldsymbol{\psi}}_{up} = \left[\widetilde{\boldsymbol{\psi}}_{up}', \breve{\boldsymbol{\psi}}_{up}', \left(\mathbf{0}_{Q\widetilde{G}}\right)'\right]'$ $(Q\widetilde{E}\times 1 \text{ vector})$, where $-\boldsymbol{\infty}_{Q\widetilde{G}}$ is a $Q\widetilde{G}\times 1$-column vector of negative infinities, $\mathbf{0}_{Q\widetilde{G}}$ is another $Q\widetilde{G}\times 1$-column vector of zeros, and $\widetilde{\boldsymbol{\psi}}_{low} = (\widetilde{\boldsymbol{\psi}}_{1,low}', \widetilde{\boldsymbol{\psi}}_{2,low}', ..., \widetilde{\boldsymbol{\psi}}_{Q,low}')'$ $(QN\times 1 \text{ vector})$,

$\tilde{\boldsymbol{\psi}}_{up} = (\tilde{\boldsymbol{\psi}}'_{1,up}, \tilde{\boldsymbol{\psi}}'_{2,up}, ..., \tilde{\boldsymbol{\psi}}'_{Q,up})'$ ($QN \times 1$ vector), $\breve{\boldsymbol{\psi}}_{low} = (\breve{\boldsymbol{\psi}}'_{1,low}, \breve{\boldsymbol{\psi}}'_{2,low}, ..., \breve{\boldsymbol{\psi}}'_{Q,low})'$ ($QC \times 1$ vector), and $\breve{\boldsymbol{\psi}}_{up} = (\breve{\boldsymbol{\psi}}'_{1,up}, \breve{\boldsymbol{\psi}}'_{2,up}, ..., \breve{\boldsymbol{\psi}}'_{Q,up})'$ ($QC \times 1$ vector). Then the likelihood function may be written as:

$$L(\boldsymbol{\lambda}) = f_{QH}(\boldsymbol{y} \mid \tilde{\boldsymbol{B}}_y, \tilde{\boldsymbol{\Omega}}_y) \times \Pr\left[\breve{\boldsymbol{\psi}}_{low} \le \tilde{\boldsymbol{u}} \le \breve{\boldsymbol{\psi}}_{up}\right], \tag{B.1}$$

$$= f_{QH}(\boldsymbol{y} \mid \tilde{\boldsymbol{B}}_y, \tilde{\boldsymbol{\Omega}}_y) \times \int_{D_r} f_{Q\tilde{E}}(\boldsymbol{r} \mid \ddot{\boldsymbol{B}}_{\tilde{u}}, \ddot{\boldsymbol{\Omega}}_{\tilde{u}})\, dr.$$

In the above expression, $f_{QH}(\boldsymbol{y} \mid \tilde{\boldsymbol{B}}_y, \tilde{\boldsymbol{\Omega}}_y)$ is a multivariate density function of dimension QH with mean $\tilde{\boldsymbol{B}}_y$ and covariance $\tilde{\boldsymbol{\Omega}}_y$, evaluated at $\boldsymbol{y}$ (this is the marginal likelihood of the H continuous variable outcomes for all Q individuals). $\Pr\left[\breve{\boldsymbol{\psi}}_{low} \le \tilde{\boldsymbol{u}} \le \breve{\boldsymbol{\psi}}_{up}\right]$ is a $Q\tilde{E}$-dimensional rectangular integral to evaluate the conditional (on $\boldsymbol{y}$) likelihood of all ordinal, count, and nominal variable outcomes for all Q individuals. $D_r = \{\boldsymbol{r} : \breve{\boldsymbol{\psi}}_{low} \le \boldsymbol{r} \le \breve{\boldsymbol{\psi}}_{up}\}$ is the integration domain spanning the multivariate region of the $\tilde{\boldsymbol{u}}$ vector (conditional on $\boldsymbol{y}$), with conditional mean $\ddot{\boldsymbol{B}}_{\tilde{u}}$ and conditional variance $\ddot{\boldsymbol{\Omega}}_{\tilde{u}}$ determined by the observed ordinal and count outcomes, and the range $\left(-\infty_{Q\tilde{G}}, \mathbf{0}_{Q\tilde{G}}\right)$ for the utility differences (taken with respect to the observed choice alternative) for the nominal outcomes. Evaluation of such high dimensional integrals is infeasible with techniques currently available in the literature, as discussed earlier in Section 1 of the main paper. A possible solution to this problem is to use the composite marginal likelihood (CML) approach. In the CML approach, the maximizing function is developed as the product of low dimensional marginal densities (see Bhat, 2014 for a detailed description of the CML approach). For the spatial-GHDM model, the CML function may be written as a product of pairwise marginal densities, across all pairs of individuals, as follows:

$$L_{CML}(\boldsymbol{\lambda}) = \prod_{q=1}^{Q-1} \prod_{q'=q+1}^{Q} f_{2*H}(\boldsymbol{y}_{qq'} \mid \tilde{\boldsymbol{B}}_{qq',\boldsymbol{y}}, \tilde{\boldsymbol{\Omega}}_{qq',\boldsymbol{y}}) \times \Pr\left[\breve{\boldsymbol{\psi}}_{qq',low} \le \tilde{\boldsymbol{u}}_{qq'} \le \breve{\boldsymbol{\psi}}_{qq',up}\right] \tag{B.2}$$

In the above expression, $f_{2*H}(\boldsymbol{y}_{qq'} \mid \tilde{\boldsymbol{B}}_{qq',y}, \tilde{\boldsymbol{\Omega}}_{qq',y})$ is an MVN density function of dimension $2H$ and $\Pr\left[\breve{\boldsymbol{\psi}}_{qq',low} \le \tilde{\boldsymbol{u}}_{qq'} \le \breve{\boldsymbol{\psi}}_{qq',up}\right]$ is a $2\tilde{E}$-dimensional MVN integral. The latter expression can further be simplified into a (pairwise) CML function by taking the product of all pairwise joint probabilities of observed outcomes of both the individuals. Such pairings are enumerated across all pairs of observed outcomes within each individual as well as across the two individuals (please see Section 2 of the online supplement for the notationally intensive details: http://www.caee.utexas.edu/prof/bhat/ABSTRACTS/Spatial_GHDM/online_supplement.pdf).