DISCLAIMER:

This document does not meet the
current format guidelines of
the Graduate School at
The University of Texas at Austin.

It has been published for
informational use only.

The Dissertation Committee for Sebastian Astroza Certifies that this is the approved version of the following dissertation:

Flexible Multiple Discrete-Continuous Choice Structures and Mixed Modeling

Committee:

Chandra R. Bhat, Supervisor

Jason Abrevaya

Steve Boyles

Randy B. Machemehl

Chandler Stolp

Flexible Multiple Discrete-Continuous Choice Structures and Mixed Modeling

by

Sebastian Astroza

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May, 2018

Flexible Multiple Discrete-Continuous Choice Structures and Mixed Modeling

Sebastian Astroza, Ph.D.

The University of Texas at Austin, 2018

Supervisor: Chandra R. Bhat

In Multiple discrete-continuous (MDC) choice situations, consumers choose one or more alternatives from a set of alternatives jointly with the amount of the chosen alternative to consume. The MDC model that has dominated the recent literature is based on an utility maximization framework. In the utility functional form, each alternative is assumed to have a baseline preference (marginal utility at the point of zero consumption). Stochasticity is usually introduced in these baseline preferences as a kernel stochastic error term to acknowledge the presence of unobserved factors that may impact the utility of each alternative. Researchers have also introduced random structures for the coefficients on the exogenous variables that allow heterogeneity (across individuals) in the sensitivity to exogenous variables. At the same time as there is more emphasis on MDC models today, there is also increasing attention on the analysis of bundle of mixed outcomes. The joint modeling of mixed outcomes is challenging because of the absence of a convenient multivariate distribution to jointly represent the relationship between discrete and continuous outcomes.

The primary objective of this dissertation is to advance the econometric modeling of MDC choice situations, with an emphasis on two aspects of this modeling. The first is to include, in a general way, heterogeneity in the sensitivity to exogenous variables. The second is to extend the joint modeling of mixed outcomes to include MDC outcomes. These two modeling enhancements are undertaken through three specific objectives:(1) formulate and estimate a finite discrete mixture of normals (FDMN) version of the MDCP model (hybrid semi-parametric approach that combines a continuous response surface for the response coefficients with a latent class approach, allowing market segmentation in the MDC context), (2) formulate and estimate a spatial MDC model that considers a multivariate skew-normal (MVSN) distribution for the random coefficients (the MVSN distribution is tractable, parsimonious, and includes the normal distribution as a special interior point case), and (3) propose a new econometric approach for the estimation of joint mixed models that include an MDC outcome. The proposed enhancements are applied to different empirical contexts to analyze several choice processes within the transportation field.

Table of Contents

List of Tables

List of Figures

CHAPTER 1: Introduction

1.1 BACKGROUND AND MOTIVATION

There are several approaches to analyzing the decision process when consumers choose one or more alternatives from a set of alternatives jointly with the amount of the chosen alternative to consume. Classical discrete-continuous choice models assume that alternatives are mutually exclusive and only one alternative can be chosen. Alternatively, multiple discrete-continuous (MDC) models expand the decision by allowing consumers to choose multiple alternatives at the same time, along with the continuous dimension of the amount of consumption. MDC models have been applied in many application contexts, including consumer brand choice and purchase quantity, activity participation and time allocation, household vehicle type and usage, recreational destination choice and number of trips, land-use type and intensity, and stock portfolio selection choice and investment amounts.

The MDC model that has dominated the recent literature is based on a utility maximization framework that assumes a non-linear (but increasing and continuously differentiable) utility function to accommodate the relationship between the decreasing marginal utility (satiation) and the increasing investment in an alternative. The model also assumes that consumers maximize this utility within their budget constraints. In the utility functional form, each alternative is assumed to have a baseline preference, which is the marginal utility of each alternative at the point of zero consumption. Stochasticity is usually introduced in these baseline preferences (for all alternatives) as an additional kernel stochastic error term to acknowledge the presence of unobserved factors that may impact the utility of each alternative. One of the most common distributions used for the kernel stochastic error term (across alternatives) is the multivariate normal distribution, which leads to an MDC probit (MDCP) model structure. Researchers have also introduced random structures for the coefficients on the exogenous variables (or response coefficients) that allow heterogeneity (across individuals) in the sensitivity to exogenous variables in MDC models. The most common assumption in the literature is that the random response coefficients are realizations from a multivariate normal distribution. However, several studies have underscored the potentially serious mis-specification consequences (in terms of theoretical considerations and data fit) of using a multivariate normal response distribution when some other non-normal response distribution is at work.

At the same time as there is more emphasis on MDC models today, there is also increasing attention on the analysis of bundle of mixed outcomes (that is, a mix of continuous and non-continuous variables of different types). The joint modeling of non-commensurate or mixed outcomes is challenging because of the absence of a convenient multivariate distribution to jointly (and directly) represent the relationship between discrete and continuous outcomes. This is particularly the case when one of the dependent outcomes is of a MDC nature. The jointness between outcomes may arise because of the impact (on the multiple choice outcomes) of common underlying exogenous observed variables, or common underlying exogenous unobserved variables, or a combination of the two. For example, we can expect that green lifestyle propensity (an unobserved variable) may jointly impact decisions such as residential location (say represented as a nominal variable in terms of the choice of location in neighborhoods of different population density categories), car ownership (a count outcome), and time-use in travel/activity patterns (an MDC outcome).

This chapter is structured as follows. In Section 1.2., the objectives of the dissertation are defined. In Section 1.3 the MDCP model structure is presented, including the model formulation and estimation. Section 1.4 outlines the dissertation structure.

1.2 DISSERTATION OBJECTIVE

The primary objective of this dissertation is to advance the econometric modeling of MDC choice situations, with an emphasis on two aspects of this modeling. The first is to include, in a general way, heterogeneity in the sensitivity to exogenous variables. The second is to extend the joint modeling of mixed outcomes to include MDC outcomes. These two modeling enhancements are undertaken through three specific objectives:

(1) formulate and estimate a finite discrete mixture of normals (FDMN) version of the MDCP model (that is, using a hybrid semi-parametric approach that combines a continuous response surface for the response coefficients with a latent class approach, allowing the introduction of market segmentation in the MDC context),

(2) formulate and estimate a spatial MDC model that considers a multivariate skew-normal (MVSN) distribution for the random coefficients (the MVSN distribution is tractable, parsimonious in parameters that regulate the distribution and its skewness, and includes the normal distribution as a special interior point case), and

(3) propose a new econometric approach for the estimation of joint mixed models that include an MDC outcome and a nominal discrete outcome, in addition to count, binary/ordinal outcomes, and continuous outcomes.

The proposed enhancements are applied to different empirical contexts to analyze several choice processes within the transportation field. In the following sections, the basics of the MDCP model are explained.

1.3 THE MULTIPLE-DISCRETE CONTINUOUS PROBIT (MDCP) MODEL

1.3.1 Utility functional form

1.3.1.1 Case of only inside goods

Following Bhat (2008), consider a choice scenario where a consumer q (q = 1, 2, …, Q) maximizes his/her utility subject to a binding budget constraint:

$$\max U_q(\boldsymbol{x}_q) = \sum_{k=1}^{K} \frac{\gamma_{qk}}{\alpha_{qk}} \psi_{qk} \left(\left(\frac{x_{qk}}{\gamma_{qk}} + 1 \right)^{\alpha_{qk}} - 1 \right)$$
$$s.t. \quad \sum_{k=1}^{K} p_{qk} x_{qk} = E_q, \tag{1.1}$$

where the utility function $U_q(\boldsymbol{x}_q)$ is quasi-concave, increasing and continuously differentiable, $\boldsymbol{x}_q \geq 0$ is the consumption quantity (vector of dimension K×1 with elements x_{qk}), and γ_{qk}, α_{qk}, and ψ_{qk} are parameters associated with good k and consumer q. The constraint in Equation (1.1) is the linear budget constraint, where E_q is the total expenditure (or income) of consumer q, and p_{qk} is the unit price of good k as experienced by consumer q. The utility function form in Equation (1.1) assumes that there is no essential outside good, so that corner solutions (*i.e.*, zero consumptions) are allowed for all the goods k (the model formulation in presence of outside goods is presented in Section 1.3.1.2). The parameter γ_{qk} in Equation (1.1) allows corner solutions for good k, but also serves the role of a satiation parameter. The role of α_{qk} is to capture satiation effects, with smaller value of α_{qk} implying higher satiation for good k. ψ_{qk} represents the stochastic baseline marginal utility (the marginal utility at the point of zero consumption).

The utility function in Equation (1.1) constitutes a valid utility function if $\gamma_{qk} > 0$, $\alpha_{qk} \leq 1$, and $\psi_{qk} > 0$ for all q and k. In empirical terms, it is not possible to disentangle the two effects of the γ_{qk} and α_{qk} parameters, which leads to empirical identification issues and estimation breakdowns when one attempts to estimate both γ_{qk} and α_{qk} parameters for each good. Researchers have either constrained α_{qk} to zero for all goods (technically, assumed $\alpha_{qk} \to 0$ for all k) and estimated the parameters, or constrained γ_{qk} to 1 for all goods and estimated the α_{qk} parameters. The first case is usually referred as the $\boldsymbol{\gamma}$ **profile**, and the second case is known as the $\boldsymbol{\alpha}$ **profile**, as presented in Equation (1.2).

$$\begin{aligned} &\gamma \text{ profile} \quad U_q(\boldsymbol{x}_q) = \sum_{k=1}^{K} \gamma_{qk} \psi_{qk} \ln\left(\frac{x_{qk}}{\gamma_{qk}} + 1\right) \\ &\alpha \text{ profile} \quad U_q(\boldsymbol{x}_q) = \sum_{k=1}^{K} \frac{\psi_{qk}}{\alpha_{qk}} \left(\left(x_{qk} + 1\right)^{\alpha_{qk}} - 1\right) \end{aligned} \tag{1.2}$$

Bhat (2008) suggests testing both profiles and selecting the model with the best fit. However, in this section, we will retain the general utility form of Equation (1.1) to keep the presentation general.

To complete the model structure, stochasticity is added by parameterizing the baseline utility as follows:

$$\psi_{qk} = \exp(\boldsymbol{\beta}_q' \boldsymbol{z}_{qk} + \xi_{qk}), \tag{1.3}$$

where $\boldsymbol{z}_{qk}$ is a D-dimensional vector of attributes that characterize good k and the consumer q (including a dummy variable for each good except one, to capture intrinsic preferences for each good except one good that forms the base), $\boldsymbol{\beta}_q$ is a consumer-specific vector of coefficients (of dimension $D \times 1$), and ξ_{qk} captures the idiosyncratic (unobserved) characteristics that impact the baseline utility of good k and consumer q. We assume that the error terms ξ_{qk} are multivariate normally distributed across goods k for a given consumer q: $\boldsymbol{\xi}_q = (\xi_{q1}, \xi_{q2}, ..., \xi_{qK})' \sim MVN_K(\boldsymbol{0}_K, \boldsymbol{\Lambda})$, where $MVN_K(\boldsymbol{0}_K, \boldsymbol{\Lambda})$ indicates a K-variate normal distribution with a mean vector of zeros denoted by $\boldsymbol{0}_K$ and a covariance matrix $\boldsymbol{\Lambda}$. Further, to allow taste variation due to unobserved individual attributes, $\boldsymbol{\beta}_q$ is typically considered as a realization from a multivariate normal

distribution: $\boldsymbol{\beta}_q \sim MVN_D(\boldsymbol{b},\boldsymbol{\Omega})$. As we mentioned earlier, one of the main objectives of this dissertation is using more flexible error structures for $\boldsymbol{\beta}_q$ and $\boldsymbol{\xi}_q$ (see chapters 2 and 3). The vectors $\boldsymbol{\beta}_q$ and $\boldsymbol{\xi}_q$ are assumed to be independent of each other. For future reference, we also write $\boldsymbol{\beta}_q = \boldsymbol{b} + \tilde{\boldsymbol{\beta}}_q$, where $\tilde{\boldsymbol{\beta}}_q \sim MVN_D(0_D,\boldsymbol{\Omega})$. Note, however, that the parameters (in the $\boldsymbol{\beta}_q$ vector) on the dummy variables specific to each alternative have to be fixed parameters, since their randomness is already captured in the covariance matrix $\boldsymbol{\Lambda}$.

1.3.1.2 Case of outside and inside goods

In the presence of both outside goods and inside goods, we label the outside goods as the first K_1 goods, and label the inside goods as the following K_2 goods (K_1+K_2=K). Consequently, the utility functional form of Equation (1.1) needs to be modified as follows:

$$U_q(\mathrm{x}_q) = \sum_{j=1}^{K_1} \frac{\psi_{qj}}{\alpha_{qj}} \left(x_{qj} + \gamma_{qj}\right)^{\alpha_{qj}} + \sum_{k=K_1+1}^{K} \frac{\gamma_{qk}}{\alpha_{qk}} \psi_{qk} \left(\left(\frac{x_{qk}}{\gamma_{qk}} + 1 \right)^{\alpha_{qk}} - 1 \right) \tag{1.4}$$

Obviously, we need $\gamma_{qj} > 0$ and $x_{qj} + \gamma_{qj} > 0$ for all j=1,2,…,K_1. The magnitude of γ_{qj} may be interpreted as the required lower bound for consumption of the outside good j. Similarly to the only inside goods case, γ_{qj} and α_{qj} parameters cannot be estimated simultaneously, and analysts should choose between the γ profile and the α profile.

1.3.2 Model estimation

To find the optimal allocation of goods, the analyst should construct the Lagrangian and derive the Karush-Kuhn-Tucker (KKT) conditions. The Lagrangian function for the problem, after substituting Equation (1.3) in Equation (1.1) is:

$$\mathrm{L}_q = \sum_{k=1}^{K} \frac{\gamma_{qk}}{\alpha_{qk}} \exp(\boldsymbol{b}'\boldsymbol{z}_{qk} + \tilde{\boldsymbol{\beta}}_q'\boldsymbol{z}_{qk} + \xi_{qk}) \left\{ \left(\frac{x_{qk}}{\gamma_{qk}} + 1 \right)^{\alpha_{qk}} - 1 \right\} - \lambda_q \left[\sum_{k=1}^{K} p_{qk} x_{qk} - E_q \right], \tag{1.5}$$

where λ_q is the Lagrangian multiplier for the expenditure constraint, which represents the marginal utility of total expenditure (or income). The KKT first-order conditions for the optimal consumption allocations (the x_{qk}^* values) are:

$$\exp(\boldsymbol{b}'\boldsymbol{z}_{qk} + \tilde{\boldsymbol{\beta}}_q'\boldsymbol{z}_{qk} + \xi_{qk})\left(\frac{x_{qk}^*}{\gamma_k}+1\right)^{\alpha_k-1} - \lambda_q\, p_{qk} = 0, \text{ if } x_{qk}^* > 0,\ k = 1,2,...,K$$
$$\exp(\boldsymbol{b}'\boldsymbol{z}_{qk} + \tilde{\boldsymbol{\beta}}_q'\boldsymbol{z}_{qk} + \xi_{qk})\left(\frac{x_{qk}^*}{\gamma_k}+1\right)^{\alpha_k-1} - \lambda_q\, p_{qk} < 0, \text{ if } x_{qk}^* = 0,\ k = 1,2,...,K. \quad (1.6)$$

The optimal demand satisfies the conditions above plus the budget constraint. The budget constraint implies that only K–1 of the x_{qk}^* values need to be estimated, since the quantity consumed of any one good is automatically determined from the quantities consumed of all the other goods. To accommodate this constraint, let m_q be the consumed good with the lowest value of k for the q^{th} consumer. For instance, if the choice set has seven goods $(K = 7)$ and the consumer q chooses goods 2, 3 and 5, then $m_q = 2$. The order in which the goods are organized does not affect the model formulation or estimation, since the definition of m_q only serves as a reference to compare marginal utilities (note also that the consumer q should choose at least one good given that $E_q > 0$). For the good m_q, the Lagrangian multiplier may then be written as:

$$\lambda_q = \frac{\exp(\boldsymbol{b}'\boldsymbol{z}_{qm_q} + \tilde{\boldsymbol{\beta}}_q'\boldsymbol{z}_{qm_q} + \xi_{qm_q})}{p_{qm_q}}\left(\frac{x_{qm_q}^*}{\gamma_{qm_q}}+1\right)^{\alpha_{qm_q}-1}. \quad (1.7)$$

Substituting for λ_q from above into Equation (1.6) for the other goods k ($k = 1,2,...,K;\ k \neq m_q$), and taking logarithms, we can rewrite the KKT conditions as:

$$V_{qk} + \tilde{\boldsymbol{\beta}}_q'\boldsymbol{z}_{qk} + \xi_{qk} = V_{qm_q} + \tilde{\boldsymbol{\beta}}_q'\boldsymbol{z}_{qm_q} + \xi_{qm_q}, \text{ if } x_{qk}^* > 0,\ k = 1,2,...,K,\ k \neq m_q$$
$$V_{qk} + \tilde{\boldsymbol{\beta}}_q'\boldsymbol{z}_{qk} + \xi_{qk} < V_{qm_q} + \tilde{\boldsymbol{\beta}}_q'\boldsymbol{z}_{qm_q} + \xi_{qm_q}, \text{ if } x_{qk}^* = 0,\ k = 1,2,...,K,\ k \neq m_q, \quad (1.8)$$

where $V_{qk} = \mathrm{b}'\mathrm{z}_{qk} + (\alpha_{qk} - 1)\ln\left(\frac{x_{qk}^*}{\gamma_{qk}}+1\right) - \ln p_{qk}$. Letting $y_{qk} = V_{qk} + \tilde{\boldsymbol{\beta}}_q'\,\boldsymbol{z}_{qk} + \xi_{qk}$, and $y_{qkm_q}^* = y_{qk} - y_{qm_q}$, the KKT conditions in Equation (1.8) are equivalent to:

$$y_{qkm_q}^* = 0, \text{ if } x_{qk}^* > 0,\ k = 1,2,...,K,\ k \neq m_q$$
$$y_{qkm_q}^* < 0, \text{ if } x_{qk}^* = 0,\ k = 1,2,...,K,\ k \neq m_q. \quad (1.9)$$

Three important identification issues need to be noted here because the KKT conditions above are based on differences, as reflected in the $y^*_{qkm_q}$ terms. First, a constant cannot be identified in the $\boldsymbol{b}'\boldsymbol{z}_{qk}$ term for one of the *K* goods. Similarly, consumer-specific variables that do not vary across goods can be introduced for *K*–1 goods, with the remaining good being the base. Second, only the covariance matrix of the error differences is estimable. Taking the difference with respect to the first good, only the elements of the covariance matrix $\boldsymbol{\Lambda}_1$ of $\varepsilon_{qk1} = \xi_{qk} - \xi_{q1}$, $k \neq 1$ are estimable. However, the KKT conditions take the difference against the first consumed good m_q by consumer *q*. Thus, in translating the KKT conditions to the consumption probability for consumer *q*, the covariance matrix $\boldsymbol{\Lambda}_{m_q}$ is desired. Since m_q will vary across consumers *q*, $\boldsymbol{\Lambda}_{m_q}$ will also vary across consumers. But all the $\boldsymbol{\Lambda}_{m_q}$ matrices must originate in the same covariance matrix $\boldsymbol{\Lambda}$ for the original error term vector ξ_q. To achieve this consistency, $\boldsymbol{\Lambda}$ is constructed from $\boldsymbol{\Lambda}_1$ by adding an additional row on top and an additional column to the left. All elements of this additional row and column are filled with values of zeros. $\boldsymbol{\Lambda}_{m_q}$ may then be obtained appropriately for each consumer *q* based on the same $\boldsymbol{\Lambda}$ matrix. Third, an additional scale normalization needs to be imposed on $\boldsymbol{\Lambda}$ if there is no price variation across goods for each consumer *q* (*i.e.*, if $p_{qk} = \tilde{p}_q \; \forall k$ and $\forall q$). For instance, one can normalize the element of $\boldsymbol{\Lambda}$ in the second row and second column to the value of one. But, if there is some price variation across goods for even a subset of consumers, there is no need for this scale normalization and all the *K*(*K*–1)/2 parameters of the full covariance matrix of $\boldsymbol{\Lambda}_1$ are estimable (see Bhat, 2008 for a discussion of this scale normalization issue).

The parameters to be estimated include the α_{qk} parameters (for an α-profile), the γ_{qk} parameters (for a γ-profile), the $\boldsymbol{b}$ vector, and the elements of the covariance matrices $\boldsymbol{\Omega}$ and $\boldsymbol{\Lambda}$. In the rest of this section, we will use the following key notation: $f_G(.;\boldsymbol{\mu},\boldsymbol{\Sigma})$ for the multivariate normal density function of dimension *G* with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, $\boldsymbol{\omega}_{\boldsymbol{\Sigma}}$ for the diagonal matrix of standard deviations of $\boldsymbol{\Sigma}$ (with its r^{th} element being $\omega_{\Sigma,r}$), $\phi_G(.;\boldsymbol{\Sigma}^*)$ for the multivariate standard normal density function of dimension *G* and correlation matrix $\boldsymbol{\Sigma}^*$, such that $\boldsymbol{\Sigma}^* = \boldsymbol{\omega}_{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\boldsymbol{\omega}_{\boldsymbol{\Sigma}}^{-1}$, $F_G(.;\boldsymbol{\mu},\boldsymbol{\Sigma})$ for the multivariate normal cumulative distribution function

of dimension G with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, and $\Phi_G(.;\boldsymbol{\Sigma}^*)$ for the multivariate standard normal cumulative distribution function of dimension G and correlation matrix $\boldsymbol{\Sigma}^*$.

Using the marginal and conditional distribution properties of the multivariate normal distribution, the above likelihood function can be written as:

$$\begin{aligned} L_q &= \det(\mathbf{J}_q) \times f_{L_{q,C}}(\boldsymbol{0}_{L_{q,C}}; \tilde{\boldsymbol{H}}_{q,C}, \tilde{\boldsymbol{\Psi}}_{q,C}) \times F_{L_{q,NC}}(\boldsymbol{0}_{L_{q,NC}}; \breve{\boldsymbol{H}}_{q,NC}, \breve{\boldsymbol{\Psi}}_{q,NC}) \\ &= \det(\mathbf{J}_q) \times \left(\prod_{g=1}^{L_{q,C}} \omega_{\tilde{\Psi}_{q,C},g} \right)^{-1} \left(\phi_{L_{q,C}}(\boldsymbol{\omega}_{\tilde{\Psi}_{q,C}}^{-1}(-\tilde{\boldsymbol{H}}_{q,C}), \tilde{\boldsymbol{\Psi}}^*_{q,C}) \right) \times \Phi_{L_{q,NC}}(\boldsymbol{\omega}_{\breve{\Psi}_{q,NC}}^{-1}(-\breve{\boldsymbol{H}}_{q,NC}), \breve{\boldsymbol{\Psi}}^*_{q,NC}), \end{aligned} \tag{1.13}$$

where $\breve{\boldsymbol{H}}_{q,NC} = \tilde{\boldsymbol{H}}_{q,NC} + \tilde{\boldsymbol{\Psi}}_{q,NC,C}(\tilde{\boldsymbol{\Psi}}_{q,C})^{-1}(-\tilde{\boldsymbol{H}}_{q,C})$, $\breve{\boldsymbol{\Psi}}_{q,NC} = \tilde{\boldsymbol{\Psi}}_{q,NC} - \tilde{\boldsymbol{\Psi}}_{q,NC,C}(\tilde{\boldsymbol{\Psi}}_{q,C})^{-1}\tilde{\boldsymbol{\Psi}}'_{q,NC,C}$, $\tilde{\boldsymbol{\Psi}}^*_{q,C} = \boldsymbol{\omega}_{\tilde{\Psi}_{q,C}}^{-1}\tilde{\boldsymbol{\Psi}}_{q,C}\boldsymbol{\omega}_{\tilde{\Psi}_{q,C}}^{-1}$, and $\breve{\boldsymbol{\Psi}}^*_{q,NC} = \boldsymbol{\omega}_{\breve{\Psi}_{q,NC}}^{-1}\breve{\boldsymbol{\Psi}}_{q,NC}\boldsymbol{\omega}_{\breve{\Psi}_{q,NC}}^{-1}$.

The multivariate normal cumulative distribution (MVNCD) function in Equation (1.13) is of dimension $L_{q,NC}$, which can have a dimensionality of up to (K–1). Typical simulation-based methods to approximate this MVNCD function can get inaccurate and time-consuming as K increases. An alternative is to use the maximum approximate composite marginal likelihood (MACML) approach (Bhat, 2011), in which the multiple integrals are evaluated using a fast analytic approximation method. The MACML estimator is based solely on univariate and bivariate cumulative normal distribution evaluations, regardless of the dimensionality of integration, which considerably reduces computation time compared to other simulation techniques to evaluate multidimensional integrals (see Bhat and Sidharthan, 2011 for an extended simulation analysis of the ability of the MACML method to recover parameters). The MACML approach was proposed to estimate mixed multinomial probit models (MNP), but can be extended to other modeling frameworks that result in MVNCD function evaluations, such as the proposed MDCP modeling framework.

There is one very important issue that still needs to be dealt with. This concerns the positive definiteness of covariance matrices. The positive-definiteness of $\tilde{\boldsymbol{\Psi}}_q$ in the likelihood function can be ensured by using a Cholesky-decomposition of the matrices $\boldsymbol{\Omega}$ and $\boldsymbol{\Lambda}$, and estimating these Cholesky-decomposed parameters. Note that, to obtain the Cholesky factor for $\boldsymbol{\Lambda}$, we first obtain the Cholesky factor for $\boldsymbol{\Lambda}_1$, and then add a column of zeros as the first column and a row of zeros as the first row to the Cholesky factor $\boldsymbol{\Lambda}_1$.

1.4 DISSERTATION OUTLINE

The rest of the dissertation is structured as follows. Chapter 2 presents a FDMN version of the MDCP model. Chapter 3 provides the formulation of a spatial MDC model with an MVSN distribution for the random coefficients and the kernel error term. Chapter 4 describes a new econometric approach for the estimation of joint mixed models that include an MDC outcome. The last chapter concludes the dissertation by summarizing the findings in the previous chapters, discussing limitations of the current work, and suggesting directions for future research.

CHAPTER 2: Allowing a General Form for Unobserved Heterogeneity in the MDCP Model

The material in this chapter is drawn substantially from the following published paper:

Bhat, C. R., Astroza, S., and Bhat, A. C. (2016). On allowing a general form for unobserved heterogeneity in the multiple discrete–continuous probit model: Formulation and application to tourism travel. *Transportation Research Part B* 86, 223-249.

In this chapter, we propose a new econometric formulation and an associated estimation method for a finite discrete mixture of normals (FDMN) version of the MDCP model. In the next section the concept of unobserved heterogeneity is discussed in the context of MDC models. In Section 2.2, a MDCP model is formulated using a general form for the unobserved heterogeneity. In Section 2.3, simulation exercises are undertaken to examine the ability of the estimation method to recover parameters from finite samples. Section 2.4 describes an empirical application of the proposed framework to analyze individual-level decisions regarding recreational destination locations and the number of trips to each destination. Section 2.5 summarizes the main findings of this chapter.

2.1 UNOBSERVED HETEROGENEITY IN MDC MODELS

Researchers have introduced random structures for the coefficients on the exogenous variables (or response coefficients) that allow heterogeneity (across individuals) in the sensitivity to exogenous variables in discrete choice models. There are three possible approaches to introduce randomness in the response coefficients. The first approach uses continuous random structures for the coefficients on the exogenous variables. Within this approach, the most common assumption is that the random response coefficients are realizations from a multivariate normal

distribution.[1] But this can lead to a misspecification if some other non-normal distribution characterizes the taste heterogeneity for one or more coefficients (see Train, 1998; Amador *et al*., 2005; Train and Sonnier, 2005; Hensher *et al*., 2005; Fosgerau, 2005; Greene *et al*., 2006; Balcombe *et al*., 2009; and Torres *et al*., 2011). The second approach uses a discrete distribution for the response coefficients. This approach leads to the familiar latent class model with an endogenous segmentation that allocates individuals probabilistically to segments as a function of exogenous variables (see Bhat, 1997; Greene and Hensher, 2003; Train, 2008; Bastin *et al*., 2010; Cherchi *et al*., 2009; and Sobhani *et al*., 2013). The problem with this approach, however, is that homogeneity in response is assumed within each latent class. The third approach uses a hybrid semi-parametric approach that combines a continuous response surface for the coefficients with a latent class approach (see, for example, Campbell *et al*., 2010; Bujosa *et al*., 2010; Greene and Hensher, 2013; and Xiong and Mannering, 2013). In this approach, the response coefficients are typically assumed to be realizations of a discrete mixture of multivariate normal distributions. That is, the relationship between the propensity variable and exogenous variables is assumed to belong to one of several latent (discrete) classes. Within each of these classes, the coefficients are drawn from a continuous multivariate normal distribution. The resulting finite discrete mixture of normal (FDMN) model generalizes the heterogeneity form because the normally distributed random parameters approach and the latent class approach consist of special cases—the first approach resulting when there is only one latent class and the

[1] To put things in context within the broader literature on accommodating non-normal coefficients, note that the second latent segmentation approach is equivalent to a non-parametric approach in which all random coefficients are assumed to have the same number of nodal points, with the number of nodal points being equal to the number of latent segments. The nodal points correspond to the segment-specific values (for each coefficient) in the latent segmentation set-up, and the probability masses at these nodal points for each individual correspond to the segment membership probabilities for that individual. This latent segmentation set-up is a restrictive version of a more general non-parametric specification in which the number of nodal points is allowed to vary across coefficients and both the nodal points and probability masses are separately estimated for each coefficient (see, for example, Bastin *et al.*, 2010). However, this general non-parametric approach is seldom used because consistency is achieved only in very large samples and parameter estimates generally have high variance (Mittelhammer and Judge, 2011). On the other hand, a continuous distribution offers substantial efficiency in the number of mixing parameters. In this regard, the paper by Bhat and Sidharthan (2012) is of particular note because it enables a non-normal (skew) continuous distribution to be used. However, their approach still enforces a unimodal distribution for the coefficients. The finite discrete mixture of normals (FDMN) is a good hybrid semi-parametric approach that combines the flexibility of the discrete mixture distribution with the efficiency advantage of the continuous distributions. It is by far the most widely used semi-parametric approach in the statistical and econometric literature because of this good balance between flexibility and efficiency (see Geweke and Keane, 1999, Frühwirth-Schnatter, 2011, and Ferdous *et al.*, 2011).

second resulting when the multivariate normal distribution becomes degenerate within each latent class.

Several earlier studies have included heterogeneity in the sensitivity to exogenous variables in the MDC context. Bhat *et al*. (2013a) proposed an estimation approach for the MDCP model that allows taste variation through the inclusion of random parameters. They demonstrated the ability to recover the parameters based on a simulation experiment, using both cross-sectional and panel data, and applied the model to analyze recreational long-distance travel. On the same topic of recreational travel, Kuriyama *et al*. (2010) proposed a latent class KKT model based on the linear expenditure system with translated constant elasticity of substitution utility functions proposed by Hanemman (1978). Sobhani *et al*. (2013) and Wafa *et al*. (2015) use a latent class approach with the MDCEV kernel structure. In Sobhani *et al*. (2013), the authors propose an estimation approach combining the full information maximum likelihood and the expectation maximization approaches. The latent class MDCEV model is applied to study non-workers' daily decisions regarding vehicle type and usage in conjunction with activity type and accompaniment choice decisions. Wafa *et al*. (2015) proposed a latent class MDCEV model to study the spatial transferability of activity-travel models.

In this dissertation, we propose an FDMN version of the MDCP model. To our knowledge, this is the first such formulation and application of an MDCP model in the econometric literature. We also propose the use of Bhat's (2011) maximum approximate composite marginal likelihood (MACML) inference approach for the estimation. This approach is computationally efficient and does not involve quasi-Monte Carlo simulation techniques of the type proposed in Bhat (2000) and Bhat (2001). The advantage of the MACML approach relative to simulation techniques is that it involves only univariate and bivariate cumulative normal distribution function evaluations in the likelihood function, regardless of the number of alternatives or segments in the latent classification. Using a 2012 New Zealand Domestic Travel Survey data set, the model is applied to analyze individual-level decisions regarding recreational destination locations and the number of trips to each destination. The results provide insights into the demographic and other factors that influence individuals' preferences for different recreational destinations, and show that the FDMN MDCP model is able to identify different discrete segments of the sample, each one of them with different stochastic effects of the exogenous variables on destination choice (and the effects varying across the discrete segments).

2.2 METHODOLOGY

2.2.1 Model formulation

Following the formulation in section 1.4, consider a choice scenario where a consumer q (q = 1, 2,…, Q) belonging to a segment g (g = 1, 2,…, G) maximizes his/her utility subject to a binding constraint, as shown in Equation (2.1):

$$\max U_q(\boldsymbol{x}_q) \,|\, \text{segment } g = \sum_{k=1}^{K} \frac{\gamma_{qgk}}{\alpha_{qgk}} \psi_{qgk} \left(\left(\frac{x_{qk}}{\gamma_{qgk}} + 1 \right)^{\alpha_{qgk}} - 1 \right)$$
$$\text{s.t. } \sum_{k=1}^{K} p_{qk} x_{qk} = E_q, \tag{2.1}$$

where the utility function $U_q(\boldsymbol{x}_q)$, given that consumer q belongs to segment g, is quasi-concave, increasing and continuously differentiable; $\boldsymbol{x}_q$ is the consumption quantity (vector of dimension K×1 with elements x_{qk} so that $x_{qk} \geq 0$ for all k; k is an index for good k), and γ_{qgk}, α_{qgk}, and ψ_{qgk} are parameters associated with good k and consumer q, given that consumer q belongs to segment g.[2] In the budget constraint, E_q is the total expenditure (or income) of consumer q, and p_{qk} is the unit price of good k as experienced by consumer q. Assume, for now, that there is no essential outside good, so that corner solutions (zero consumptions) are possible for all goods k (relaxing this assumption is straightforward and simplifies the analysis considerably). The parameter ψ_{qgk} represents the baseline marginal utility for good k, given that the individual q belongs to population segment g (*i.e.*, ψ_{qgk} is the marginal utility of good k at the point of no consumption of good k, given that q belongs to segment g). The parameter γ_{qgk} allows a corner solution for good k and also serves as a translation-based satiation parameter, while α_{qgk} serves as an exponential-based satiation parameter. As discussed in detail in Bhat (2008), only one parameter of the set γ_{qgk} or α_{qgk} will be empirically identified, so the analyst will have to estimate a γ-profile (in which $\alpha_{qgk} \to 0$) or an α-profile (in which the γ_{qgk} terms are normalized to the value of one). Both these profiles can be estimated, and the one that provides a

[2] Though we will refer to the alternatives for consumption as "goods" in this section, it is important to note that the consumption alternatives can also refer to consumption of different types of activities or, as in the empirical analysis of the current chapter, to "consumption" of different destination regions for leisure trips.

better data fit may be selected. Also, for the γ-profile, we will need $\gamma_{qgk} > 0 \ \forall k$, and, for the α-profile, we will need $\alpha_{qgk} \leq 1 \ \forall k$. In the current research, we will retain the general utility form of Equation (2.1) to keep the presentation general.

In Equation (2.2) we introduce observed heterogeneity across individuals within segment *g* and stochasticity through the baseline marginal utility function ψ_{qgk}:

$$\psi_{qgk} = \exp(\boldsymbol{\beta}'_{qg} \boldsymbol{z}_{qk}), \tag{2.2}$$

where $\boldsymbol{z}_{qk}$ is a *D*-dimensional vector of attributes that characterizes good *k* and the consumer *q* (including a constant for each good except one, to capture intrinsic preferences for each good relative to a base good); $\boldsymbol{\beta}_{qg}$ is a consumer-specific vector of coefficients (of dimension *D*×1) that allows unobserved taste variation across all consumers *q* in segment *g* and allows different observed responsiveness across all consumers *q* based on different values of the elements of the vector $\boldsymbol{z}_{qk}$. In the current research, we consider $\boldsymbol{\beta}_{qg}$ to be a realization from a multivariate normal distribution: $\boldsymbol{\beta}_{qg} \sim f_D(\boldsymbol{b}_g, \boldsymbol{\Omega}_g)$. For future reference, we also write $\boldsymbol{\beta}_{qg} = \boldsymbol{b}_g + \tilde{\boldsymbol{\beta}}_{qg}$, where $\tilde{\boldsymbol{\beta}}_{qg} \sim f_D(\boldsymbol{0}_D, \boldsymbol{\Omega}_g)$.

The optimal consumption vector $\boldsymbol{x}_q$ can be solved based on the constrained optimization problem of Equation (2.1) by forming the Lagrangian function and applying the KKT conditions, conditional on the individual belonging to segment *g*. The Lagrangian function for the problem is provided in Equation (2.3):

$$\ell_q \mid q \in \text{segment } g = \sum_{k=1}^{K} \frac{\gamma_{qgk}}{\alpha_{qgk}} \exp(\boldsymbol{b}'_g \boldsymbol{z}_{qk} + \tilde{\boldsymbol{\beta}}'_{qg} \boldsymbol{z}_{qk}) \left\{ \left(\frac{x_{qk}}{\gamma_{qgk}} + 1 \right)^{\alpha_{qgk}} - 1 \right\} - \lambda_{qg} \left[\sum_{k=1}^{K} p_{qk} x_{qk} - E_q \right], \tag{2.3}$$

where λ_{qg} is a segment *g*-specific Lagrangian multiplier associated with the expenditure constraint. The KKT first-order conditions for the optimal consumption x^*_{qk}, given that consumer *q* belongs to segment *g*, are as shown in Equation (2.4):

$$\begin{aligned} &\exp(\boldsymbol{b}'_g \boldsymbol{z}_{qk} + \tilde{\boldsymbol{\beta}}'_{qg} \boldsymbol{z}_{qk}) \left(\frac{x^*_{qk}}{\gamma_{qgk}} + 1 \right)^{\alpha_{qgk}-1} - \lambda_{qg} \ p_{qk} = 0, \text{ if } x^*_{qk} > 0, \ k = 1,2,...,K \\ &\exp(\boldsymbol{b}'_g \boldsymbol{z}_{qk} + \tilde{\boldsymbol{\beta}}'_{qg} \boldsymbol{z}_{qk}) \left(\frac{x^*_{qk}}{\gamma_{qgk}} + 1 \right)^{\alpha_{qgk}-1} - \lambda_{qg} \ p_{qk} < 0, \text{ if } x^*_{qk} = 0, \ k = 1,2,...,K. \end{aligned} \tag{2.4}$$

The optimal demand, conditional on individual q belonging to segment g, satisfies the above conditions and the budget constraint $\sum_{k=1}^{K} p_{qk} x_{qk}^{*} = E_q$. The budget constraint implies that only K–1 of the x_{qk}^{*} values need to be estimated. To accommodate this singularity, let m_q be, without loss of generality, the consumed good with the lowest value of k for the q^{th} consumer (note that the consumer must consume at least one good given $E_q > 0$). For this m_q^{th} good, $x_{qm_q}^{*} > 0$, which implies Equation (2.5) from the first set of KKT conditions in Equation (2.4):

$$\lambda_{qg} = \frac{\exp(\boldsymbol{b}_g' \boldsymbol{z}_{qm_q} + \tilde{\boldsymbol{\beta}}_{qg}' \boldsymbol{z}_{qm_q})}{p_{qm_q}} \left(\frac{x_{qm_q}^{*}}{\gamma_{qgm_q}} + 1 \right)^{\alpha_{qgm_q} - 1}. \tag{2.5}$$

Substituting back in Equation (2.4) for the other goods k ($k = 1,2,...,K$; $k \neq m_q$), and taking logarithms and simplifying, we may write the KKT conditions as Equation (2.6):

$$\begin{aligned} &y_{qgkm_q}^{*} = 0, \text{ if } x_{qk}^{*} > 0, \; k = 1,2,...,K, \; k \neq m_q, \\ &y_{qgkm_q}^{*} < 0, \text{ if } x_{qk}^{*} = 0, \; k = 1,2,...,K, \; k \neq m_q. \end{aligned} \tag{2.6}$$

where $y_{qgkm_q}^{*} = y_{qgk} - y_{qgm_q}$; $y_{qgk}^{*} = V_{qgk} + \tilde{\boldsymbol{\beta}}_{qg}' \boldsymbol{z}_{qk}$; and $V_{qgk} = \boldsymbol{b}_g' \boldsymbol{z}_{qk} + (\alpha_{qgk} - 1) \ln\left(\frac{x_{qk}^{*}}{\gamma_{qgk}} + 1 \right) - \ln p_{qgk}$.

The above conditions are conditional on individual q belonging to segment g. Within this context, two important identification considerations need to be noted (additional identifications considerations due to multiple segments will be noted later). First, a dummy variable (or constant) corresponding to one of the K goods should not be introduced, since only differences in the y_{qgk}^{*} terms matter (this is similar to a standard discrete choice model). Similarly, consumer-specific variables that do not vary across goods can be introduced only for (*K–1*) goods, with the remaining alternative being the base. Let the first alternative be the base for the dummy variable and for consumer-specific variables that do not vary across goods. That is, let $z_{q1constant} = 0$ (and correspondingly, the element in $\boldsymbol{b}_g$ corresponding to this first alternative's constant is fixed at 0 and the variance element contribution in $\boldsymbol{\Omega}_g$ corresponding to this alternative's constant is also fixed at 0; in addition, all covariance elements in $\boldsymbol{\Omega}_g$ corresponding to this first alternative's constant also are set to zero). Also, let $z_{q1l} = 0$ for all consumer-specific variables l that do not

vary across goods (and correspondingly, the elements in $\boldsymbol{b}_g$ for these variables for the first alternative are fixed at zero and so are all variances/covariances in $\boldsymbol{\Omega}_g$ for these variables for the first alternative).

2.2.2 Consumer role in a finite mixture of segments

The derivation thus far is based on the notion that consumer *q* belongs to a single segment *g*. But now consider the case that consumer *q* belongs to a finite mixture of segments—that is, the actual assignment of consumer *q* to a specific segment is not observed, but we are able to attribute different probabilities $\pi_{qg}\,(g=1,2,\ldots,G)$ to consumer *q* belonging to different segments. We require that $0 \le \pi_{qg} \le 1$, and $\sum_{g=1}^{G} \pi_{qg} = 1$. To enforce these restrictions, and following Bhat (1997), we use the logit link function shown in Equation (2.7):

$$\pi_{qg} = \frac{\exp(\boldsymbol{\mu}_g' \boldsymbol{w}_q)}{\sum_{h=1}^{G} \exp(\boldsymbol{\mu}_h' \boldsymbol{w}_h)}, \tag{2.7}$$

where $\boldsymbol{w}_q$ is a vector of individual exogenous variables, and $\boldsymbol{\mu}_1 = \boldsymbol{0}$ serves as a vector identification condition. This probabilistic assignment to segments is tantamount to using a mixture of multivariate normal distributions for $\boldsymbol{\beta}_q$: $\lambda(\boldsymbol{\beta}_q = \mathbf{a}) \,|\, \boldsymbol{b}, \boldsymbol{\Omega} = \sum_{g=1}^{G} \pi_{qg}\, f_D(\mathbf{a}; \boldsymbol{b}_g, \boldsymbol{\Omega}_g)$, where $f_D(.; \boldsymbol{b}_g, \boldsymbol{\Omega}_g)$ is the multivariate normal density function with mean vector $\boldsymbol{b}_g$ and covariance matrix $\boldsymbol{\Omega}_g$. $\boldsymbol{b}$ is a vector obtained by stacking the $\boldsymbol{b}_g$ vectors vertically, and $\boldsymbol{\Omega}$ is the matrix obtained by block-diagonally stacking the $\boldsymbol{\Omega}_g$ matrices. Specifically, one may write $\psi_{qk} = \exp(\boldsymbol{\beta}_q' z_{qk})$, which, with the mixture of MVN distributions as above for $\boldsymbol{\beta}_q$, leads to the segment-specific baseline utility functions of the form of Equation (2.2) with a probabilistic segment assignment π_{qg}. The mixture of normal distributions is a semi-parametric distribution that relaxes the normal distribution for $\boldsymbol{\beta}_q$ commonly used in typical MDC models, while allowing the distribution itself to be a function of individual-level attributes through the π_{qg} terms. The mixture distribution effectively combines the flexibility of the latent class model with the parsimony of the continuous multivariate normal distribution assumption for $\boldsymbol{\beta}_q$. In

particular, if each individual belongs to a single segment that is known *a priori* (that is, if $\pi_{qg} = 1$ for a specific segment g and zero for all other segments, and if this is known a priori for each individual q) and $\gamma_{qgk} = \gamma_{qk} \; \forall \, g$ and $\alpha_{qgk} = \alpha_{qk} \; \forall \, g,$ the model collapses to a random-coefficient MDCP model (or RC-MDCP in the rest of this chapter) as in Bhat *et al*. (2013a). On the other hand, if the multivariate normal distribution within each segment becomes degenerate (*i.e.*, $\mathbf{\Omega}_g = 0$ for all g), then the model collapses to a latent class MDCP (LC-MDCP) model.

The use of latent classes, as in the current research, requires labeling restrictions for identifiability. In particular, the parameter space includes $G!$ subspaces, each associated with a different way of labeling the mixture components. To prevent the interchange of the mixture components, we impose the labeling restriction that the constants specific to the second alternative are increasing across the segments, *i.e.*: $b_{11}<b_{21}<b_{31}<\ldots b_{G1}$ (b_{11} refers to coefficient on the dummy variable for the second alternative in the first segment, b_{21} refers to the coefficient on the dummy variable for the second alternative in the second segment, and so on until b_{G1} refers to the coefficient on the dummy variable for the second alternative in the G^{th} segment).[3] To implement the labeling restriction, we parameterize the b_{g1} values as follows: $b_{g1} = b_{g-1,1} + \exp(\kappa_g)$ for g=2,…,G. Such a labeling restriction is needed because the same model specification (and likelihood function value) results simply by interchanging the sequence in which the segments are numbered. Technically, therefore, multiple sets of parameters (corresponding to a swap of segment values) result in the same likelihood function, creating an identification problem. This identification problem is resolved through the imposition of the labeling restriction above so that the segments become non-interchangeable.[4] Finally, an additional scale normalization needs to be imposed on $\mathbf{\Omega}_g$ for one of the g segments if there is no price variation across goods for each consumer q (*i.e.*, if $p_{qk} = \tilde{p}_q \;\; \forall k$ and $\forall q$). For instance, one can normalize the variance of the second alternative's constant in the first segment

[3] As clearly indicated earlier, the constant coefficients for the first alternative are set to zero in every segment g (g = 1, 2, …, G) for identification.

[4] Of course, the labeling restriction discussed above, which we use in the simulation experiments in the next section, is only one of several possible restrictions to identify the model, Thus, in our empirical analysis, where we do not use constants in the baseline utilities of the alternatives (for reasons discussed in the empirical section), we implement another version of the labeling restriction by requiring that the constant in the first segment in the segment membership model has a maintained value of zero, and there is a descending order of magnitudes of the constants for the other segments in the segment membership model.

($g = 1$) to the value of one. But, if there is price variation across even a subset of goods for a subset of consumers, there is no need for this additional scale normalization (see Bhat, 2008).

2.2.3 Model estimation

If a γ-profile is used, the parameter γ_{qgk} may be parameterized as $\exp(\tilde{\boldsymbol{\theta}}_g' \boldsymbol{a}_{qk})$, where $\boldsymbol{a}_{qk}$ is a vector of explanatory variables and $\tilde{\boldsymbol{\theta}}_g$ is a corresponding vector of parameters. On the other hand, if an α-profile is used, the parameter α_{qgk} may be parameterized as $1-\exp(\ddot{\boldsymbol{\theta}}_g' \tilde{\boldsymbol{a}}_{qk})$ (to maintain the restriction that $\alpha_{qgk} < 1$) or as $\dfrac{1}{1+\exp(-\ddot{\boldsymbol{\theta}}_g' \tilde{\boldsymbol{a}}_{qk})}$ (to maintain the stronger restrictions that $0 < \alpha_{qgk} < 1$; this stronger restriction often helps create stability in estimation).

Let $\boldsymbol{\theta} = (\boldsymbol{b}_1', \boldsymbol{b}_2', \ldots, \boldsymbol{b}_G'; \bar{\boldsymbol{\Omega}}_1', \bar{\boldsymbol{\Omega}}_2', \ldots, \bar{\boldsymbol{\Omega}}_G'; \boldsymbol{\mu}_1', \boldsymbol{\mu}_2', \ldots, \boldsymbol{\mu}_G'; \tilde{\boldsymbol{\theta}}_1', \tilde{\boldsymbol{\theta}}_2', \ldots, \tilde{\boldsymbol{\theta}}_G')'$ if a γ-profile is estimated and $\boldsymbol{\theta} = (\boldsymbol{b}_1', \boldsymbol{b}_2', \ldots, \boldsymbol{b}_G'; \bar{\boldsymbol{\Omega}}_1', \bar{\boldsymbol{\Omega}}_2', \ldots, \bar{\boldsymbol{\Omega}}_G'; \boldsymbol{\mu}_1', \boldsymbol{\mu}_2', \ldots, \boldsymbol{\mu}_G'; \ddot{\boldsymbol{\theta}}_1', \ddot{\boldsymbol{\theta}}_2', \ldots, \ddot{\boldsymbol{\theta}}_G')'$ if an α-profile is estimated, with $\bar{\boldsymbol{\Omega}}_{\mathbf{g}}$ representing the row vectorization of the upper diagonal elements of $\boldsymbol{\Omega}_{\mathbf{g}}$. To formulate the estimation procedure, we will use the following notation: $f_S(.; \boldsymbol{\eta}, \boldsymbol{\Sigma})$ for the multivariate normal density function of dimension *S* with mean vector $\boldsymbol{\eta}$ and covariance matrix $\boldsymbol{\Sigma}$; $\boldsymbol{\omega}_{\boldsymbol{\Sigma}}$ for the diagonal matrix of standard deviations of $\boldsymbol{\Sigma}$ (with its r^{th} element being $\omega_{\boldsymbol{\Sigma},r}$); $\phi_S(.; \boldsymbol{\Sigma}^*)$ for the multivariate standard normal density function of dimension *S* and correlation matrix $\boldsymbol{\Sigma}^*$—such that $\boldsymbol{\Sigma}^* = \boldsymbol{\omega}_{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Sigma} \boldsymbol{\omega}_{\boldsymbol{\Sigma}}^{-1}$, $F_S(.; \boldsymbol{\eta}, \boldsymbol{\Sigma})$ for the multivariate normal cumulative distribution function of dimension *S* with mean vector $\boldsymbol{\eta}$ and covariance matrix $\boldsymbol{\Sigma}$—and $\Phi_S(.; \boldsymbol{\Sigma}^*)$ for the multivariate standard normal cumulative distribution function of dimension *S* and correlation matrix $\boldsymbol{\Sigma}^*$.

The derivation of the likelihood function is identical to the process presented in Section 1.4.2. The reader is referred to Appendix A for a detailed explanation. The likelihood function can be written as shown in Equation (2.8):

$$L_{qg} = \det(\mathbf{J}_{qg}) \times f_{L_{q,C}}(\mathbf{0}_{L_{q,C}}; \tilde{\mathbf{H}}_{qg,C}, \tilde{\mathbf{\Psi}}_{qg,C}) \times F_{L_{q,NC}}(\mathbf{0}_{L_{q,NC}}; \breve{\mathbf{H}}_{qg,NC}, \breve{\mathbf{\Psi}}_{qg,NC})$$
$$= \det(\mathbf{J}_{qg}) \times \left(\prod_{s=1}^{L_{q,C}} \omega_{\tilde{\mathbf{\Psi}}_{qg,C},s} \right)^{-1} \left(\phi_{L_{q,C}}(\boldsymbol{\omega}^{-1}_{\tilde{\mathbf{\Psi}}_{qg,C}}(-\tilde{\mathbf{H}}_{qg,C}), \tilde{\mathbf{\Psi}}^{*}_{qg,C}) \right) \times \Phi_{L_{q,NC}}(\boldsymbol{\omega}^{-1}_{\breve{\mathbf{\Psi}}_{qg,NC}}(-\breve{\mathbf{H}}_{qg,NC}), \breve{\mathbf{\Psi}}^{*}_{qg,NC}), \quad (2.8)$$

where $\breve{\boldsymbol{H}}_{qg,NC} = \tilde{\boldsymbol{H}}_{qg,NC} + \tilde{\boldsymbol{\Psi}}_{qg,NC,C}(\tilde{\boldsymbol{\Psi}}_{qg,C})^{-1}(-\tilde{\boldsymbol{H}}_{qg,C})$, $\tilde{\boldsymbol{\Psi}}^{*}_{qg,C} = \boldsymbol{\omega}^{-1}_{\tilde{\boldsymbol{\Psi}}_{qg,C}} \tilde{\boldsymbol{\Psi}}_{qg,C} \boldsymbol{\omega}^{-1}_{\tilde{\boldsymbol{\Psi}}_{qg,C}}$,

$\breve{\boldsymbol{\Psi}}_{qg,NC} = \tilde{\boldsymbol{\Psi}}_{qg,NC} - \tilde{\boldsymbol{\Psi}}_{qg,NC,C}(\tilde{\boldsymbol{\Psi}}_{qg,C})^{-1}\tilde{\boldsymbol{\Psi}}'_{qg,NC,C}$, $\breve{\boldsymbol{\Psi}}^{*}_{qg,NC} = \boldsymbol{\omega}^{-1}_{\breve{\boldsymbol{\Psi}}_{qg,NC}} \breve{\boldsymbol{\Psi}}_{qg,NC} \boldsymbol{\omega}^{-1}_{\breve{\boldsymbol{\Psi}}_{qg,NC}}$, and $\boldsymbol{\omega}_{\boldsymbol{\Xi}}$ represents the diagonal matrix of standard errors corresponding to matrix $\boldsymbol{\Xi}$.

Then, the likelihood function for observation *q* is:

$$L_q(\boldsymbol{\theta}) = \sum_{g=1}^{G} \pi_{qg} L_{qg}(\boldsymbol{\theta}), \quad (2.9)$$

and the likelihood function is then given as:

$$L(\boldsymbol{\theta}) = \prod_{q} L_q(\boldsymbol{\theta}). \quad (2.10)$$

The multivariate normal cumulative distribution (MVNCD) function in Equation (2.10) is of dimension $L_{q,NC}$, which can have a dimensionality of up to (*K*–1). Typical simulation-based methods to approximate this MVNCD function can become inaccurate and time-consuming as *K* increases. An alternative is to use the MACML approach (Bhat, 2011), in which the multiple integrals are evaluated using a fast analytic approximation method.[5] The MACML estimator is based solely on univariate and bivariate cumulative normal distribution evaluations, regardless of the dimensionality of integration, which considerably reduces computation time compared to other simulation techniques used to evaluate multidimensional integrals (see Bhat *et al*., 2013a for an extended simulation analysis of the ability of the MACML method to recover parameters in the simple MDCP model).

One very important issue still needs to be dealt with: the positive definiteness of covariance matrices. The positive-definiteness of $\tilde{\mathbf{\Psi}}_{qg}$ in the likelihood function can be ensured

[5] Note that in the current case, we use only the analytic approximation for the cumulative standard multivariate normal distribution embedded in the MACML; however, we will continue to refer to the approach as MACML for ease in presentation and also because the composite marginal likelihood (CML) inference approach subsumes the maximum likelihood (ML) inference approach used here as a special case.

by applying a Cholesky decomposition to the matrices $\boldsymbol{\Omega}_g$ (g = 1, 2,…, G), and estimating these Cholesky-decomposed parameters.[6]

2.3 SIMULATION EVALUATION

The simulation exercises undertaken in this section examine the ability of the MACML estimator to recover parameters from finite samples in an FDMN MDCP model by generating simulated data sets with known underlying model parameters. To examine the robustness of the MACML approach when applied to different numbers of mixtures, we consider both two- and three-mixture models. In addition, we examine the effects of (a) assuming that coefficients are fixed and not stochastic within each segment (that is, using the LC-MDCP model), and (b) assuming normality of the response coefficient when non-normality is present and thus using a single segment when multiple segments are present (that is, using the RC-MDCP model).

2.3.1 Experimental design

In the design, we consider the case with three alternatives. In each of the two- and three-mixture cases, we consider two independent variables in the $\boldsymbol{z}_{qk}$ vector in the baseline utility for each alternative. That is, consider the following for the $\boldsymbol{z}_{qk}$ vectors:

$$\boldsymbol{z}_{q1} = \left[0,0,z_{q1,1},z_{q1,2}\right],\ \boldsymbol{z}_{q2} = \left[1,0,z_{q2,1},z_{q2,2}\right], \text{and } \boldsymbol{z}_{q3} = \left[0,1,z_{q3,1},z_{q3,2}\right], \tag{2.11}$$

---

[6] Previous research in latent segmentation (see, for example, Bhat, 1997 and Sobhani *et al.*, 2013) has highlighted several estimation challenges in terms of stability and convergence. Most of these studies recommend the Expectation Maximization (EM) method to find good initial values to start the full information likelihood function iterations. However, the EM method also leads to long estimation times. In this research, to obtain good start values as well as minimize estimation time, we implemented the following steps to estimate a model with S segments, with the parameters at the end of each step serving as the initial start values for the iterations associated with the subsequent step. In particular, we first estimated an MDCP model with only a single vector of constants in the baseline utilities of the alternatives (that is, we estimated a constants-only model as though there were only one latent segment). Second, we used the constants from the first step as initial values for the baseline utility constants in the first segment, made a random perturbation of these values by increasing or decreasing these values between 5% to 10% for the baseline utility constants in the remaining S-1 segments, used a constants-only specification for the segment membership probabilities with the constant for the first segment constrained to zero, and the other constants perturbed from zero in a way that adheres to the labeling restriction as discussed at the top of Section 2.2.2, and estimated a latent segmentation MDCP model (with only constants and no random parameters). Third, we used the results of the previous step as a starting point and added exogenous variables to the segment membership model to get a good segment membership specification. Fourth, we introduced exogenous variables in the baseline utilities of the MDCP-specific models for each segment to obtain a good latent segmentation model simultaneously with a refined segment membership model specification. Finally, we estimated the FDMN version of the model, allowing randomness in the parameters. The entire process was also aided and speeded up by the fact that we coded our own analytic gradient function.

where the last two variables in each $\boldsymbol{z}_{qk}$ (k=1,2,3) correspond to the two independent variables. The first variable in $\boldsymbol{z}_{q2}$ is the constant specific to alternative 2, while the second variable in $\boldsymbol{z}_{q3}$ is the constant specific to alternative 3. The values of the two independent variables for each alternative (*i.e.*, $\boldsymbol{z}_{q1,1}$ and $\boldsymbol{z}_{q1,2}$ for the first alternative; $\boldsymbol{z}_{q2,1}$ and $\boldsymbol{z}_{q2,2}$ for the second alternative; and $\boldsymbol{z}_{q3,1}$ and $\boldsymbol{z}_{q3,2}$ for the third alternative) are drawn from standard univariate normal distributions. In particular, a synthetic sample of 5000 realizations of the exogenous variables is generated corresponding to Q=5000 consumers. Additionally, we generate budget amounts E_q $(q = 1,2,...,Q)$ from a univariate normal distribution with a mean of 150, and truncated between the values of 100 and 200 (the prices of all goods are fixed at the value of one across all consumers). Once generated, the independent variable values and the total budget are held fixed in the rest of the simulation exercise.

2.3.1.1 Two-segment case

For the coefficients on the $\boldsymbol{z}_{qk}$ variables, we assume hybrid coefficients as follows:

$$\lambda(\boldsymbol{\beta}_q = \boldsymbol{a}) = \pi_1 \times f_4(\boldsymbol{a}; \boldsymbol{b}_1, \boldsymbol{\Omega}_1) + \pi_2 \times f_4(\boldsymbol{a}; \boldsymbol{b}_2, \boldsymbol{\Omega}_2), \tag{2.12}$$

where $\boldsymbol{b}_1 = (b_{11}, b_{12}, b_{13}, \tilde{b}\,)' = (1.0, 2.0, 0.6, 0.5)'$ for segment 1, and $\boldsymbol{b}_2 = (b_{21}, b_{22}, b_{23}, \tilde{b}\,)' = (2.0, 1.5, 0.2, 0.5)'$ for segment 2. Note that the dimension of b_1 and b_2 are the same as z_{q1}, z_{q2}, and z_{q3} (all of these are 4×1 vectors). That is, b_{11} is the mean constant coefficient on the second alternative in segment 1, b_{12} is the mean constant coefficient on the third alternative in segment 1, b_{13} is the mean coefficient on the first independent variable in the first segment, and $\tilde{b}$ is the mean coefficient on the second independent variable in the first segment. b_{21} through b_{23} are similar to b_{11} through b_{13} but for the second segment, and we maintain the same coefficient $\tilde{b}$ in both segments for the second independent variable. For the covariance matrices $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ of the coefficients we assume:

$$\boldsymbol{\Omega}_1 = \begin{bmatrix} 1 & 0.5 & 0.7 & 0 \\ 0.5 & 1 & 0.8 & 0 \\ 0.7 & 0.8 & 0.9 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \mathbf{L}_{\boldsymbol{\Omega}_1} \mathbf{L}'_{\boldsymbol{\Omega}_1} = \begin{bmatrix} 1 & 0.000 & 0.000 & 0 \\ 0.5 & 0.866 & 0.000 & 0 \\ 0.7 & 0.519 & 0.374 & 0 \\ 0.0 & 0.000 & 0.000 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0.500 & 0.700 & 0 \\ 0 & 0.866 & 0.519 & 0 \\ 0 & 0.000 & 0.374 & 0 \\ 0 & 0.000 & 0.000 & 0 \end{bmatrix}$$

$$\boldsymbol{\Omega}_2 = \begin{bmatrix} 0.81 & 0.54 & 0.72 & 0 \\ 0.54 & 1.0 & 0.8 & 0 \\ 0.72 & 0.8 & 0.89 & 0 \\ 0.00 & 0.00 & 0.00 & 0 \end{bmatrix} = \mathbf{L}_{\boldsymbol{\Omega}_2} \mathbf{L}'_{\boldsymbol{\Omega}_2} = \begin{bmatrix} 0.9 & 0.0 & 0.0 & 0 \\ 0.6 & 0.8 & 0.0 & 0 \\ 0.8 & 0.4 & 0.3 & 0 \\ 0.0 & 0.0 & 0.0 & 0 \end{bmatrix} \begin{bmatrix} 0.9 & 0.6 & 0.8 & 0 \\ 0.0 & 0.8 & 0.4 & 0 \\ 0.0 & 0.0 & 0.3 & 0 \\ 0.0 & 0.0 & 0.0 & 0 \end{bmatrix}$$

As indicated earlier, the positive definiteness of the $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ matrices is ensured in the estimations by reparameterizing the likelihood function in terms of the lower Cholesky factor matrices $\mathbf{L}_{\boldsymbol{\Omega}_1}$ and $\mathbf{L}_{\boldsymbol{\Omega}_2}$, and estimating the associated Cholesky matrix parameters. As should be obvious from the specification of $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$, we assume that the coefficient on the second independent variable (*i.e.*, $\tilde{b}$) is fixed in the simulations (note the zero entries in the last row and column of $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$).[7] Then, in the two-mixture case, there are 11 Cholesky parameters to be estimated: $l_{\boldsymbol{\Omega}_1,2} = 0.5$, $l_{\boldsymbol{\Omega}_1,3} = 0.866$, $l_{\boldsymbol{\Omega}_1,4} = 0.7$, $l_{\boldsymbol{\Omega}_1,5} = 0.519$, $l_{\boldsymbol{\Omega}_1,6} = 0.374$, $l_{\boldsymbol{\Omega}_2,1} = 0.9$, $l_{\boldsymbol{\Omega}_2,2} = 0.6$, $l_{\boldsymbol{\Omega}_2,3} = 0.8$, $l_{\boldsymbol{\Omega}_2,4} = 0.8$, $l_{\boldsymbol{\Omega}_2,5} = 0.4$, and $l_{\boldsymbol{\Omega}_2,6} = 0.3$.

The weight mixture values π_1 and π_2 are set by specifying the vector $\boldsymbol{w}_q$ to include a constant and an independent variable w_{q1} drawn from a standard univariate normal distribution. That is, $\boldsymbol{w}_q = [1, w_{q1}]'$. Also we specify $\boldsymbol{\mu}_1 = [0,0]'$ for normalization and $\boldsymbol{\mu}_2 = [0.6,0.1]'$ for the second segment. Finally, we use a γ-profile in our estimations, and set the satiations parameters for all three alternatives to 1 in both segments. That is, $\gamma_{11} = \gamma_{12} = \gamma_{13} = 1$ for the first segment, and $\gamma_{21} = \gamma_{22} = \gamma_{23} = 1$ for the second segment.

Overall, the parameters to be estimated in the two-mixture case include the following: b_{11}=1, b_{12}=2, b_{13}=0.6, b_{21}=2, b_{22}=1.5, b_{23}=0.2, $\tilde{b}$ =0.5, $l_{\boldsymbol{\Omega}_1,2} = 0.5$, $l_{\boldsymbol{\Omega}_1,3} = 0.866$, $l_{\boldsymbol{\Omega}_1,4} = 0.7$, $l_{\boldsymbol{\Omega}_1,5} = 0.519$, $l_{\boldsymbol{\Omega}_1,6} = 0.374$, $l_{\boldsymbol{\Omega}_2,1} = 0.9$, $l_{\boldsymbol{\Omega}_2,2} = 0.6$, $l_{\boldsymbol{\Omega}_2,3} = 0.8$, $l_{\boldsymbol{\Omega}_2,4} = 0.8$, $l_{\boldsymbol{\Omega}_2,5} = 0.4$, and $l_{\boldsymbol{\Omega}_2,6} = 0.3$, $\mu_{21} = 0.6$, $\mu_{22} = 0.1$, $\gamma_{11} = 1$, $\gamma_{12} = 1$, $\gamma_{13} = 1$, $\gamma_{21} = 1$, $\gamma_{22} = 1$, and $\gamma_{23} = 1$.

2.3.1.2 Three-segment case

In this case, we assume the hybrid coefficients as follows:

[7] We use the general presentation convention that the lower Cholesky matrix of a covariance matrix with a row/column with all zero values (that is, corresponding to a fixed parameter) is obtained by stripping out that row/column from the covariance matrix, obtaining the lower Cholesky matrix for the remaining sub-matrix, and then adding a row/column of zero values to the resulting lower Cholesky matrix.

$$\lambda(\boldsymbol{\beta}_q = \boldsymbol{a}) = \pi_1^* \times f_4(\boldsymbol{a}; \boldsymbol{b}_1^*, \boldsymbol{\Omega}_1^*) + \pi_2^* \times f_4(\boldsymbol{a}; \boldsymbol{b}_2^*, \boldsymbol{\Omega}_2^*) + \pi_3^* \times f_4(\boldsymbol{a}; \boldsymbol{b}_3^*, \boldsymbol{\Omega}_3^*), \quad (2.13)$$

where $\boldsymbol{b}_1^* = \boldsymbol{b}_1, \boldsymbol{b}_2^* = \boldsymbol{b}_2$, and $\boldsymbol{b}_3^* = (3.0,1.0,2.0,0.5)'$, $\boldsymbol{\Omega}_1^* = \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2^* = \boldsymbol{\Omega}_2$, and

$$\boldsymbol{\Omega}_3^* = \begin{bmatrix} 4.0 & 1.0 & 1.2 & 0 \\ 1.0 & 1.25 & 1.1 & 0 \\ 1.2 & 1.1 & 1.81 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \mathbf{L}_{\boldsymbol{\Omega}_3} \mathbf{L}'_{\boldsymbol{\Omega}_3} = \begin{bmatrix} 2.0 & 0 & 0 & 0 \\ 0.5 & 1.0 & 0 & 0 \\ 0.6 & 0.8 & 0.9 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2.0 & 0.5 & 0.6 & 0 \\ 0 & 1.0 & 0.8 & 0 \\ 0 & 0 & 0.9 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The mixture weights π_1^*, π_2^*, and π_3^* are set by specifying $\boldsymbol{\mu}_1^* = \boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2^* = \boldsymbol{\mu}_2$, and $\boldsymbol{\mu}_3 = [0.4,0]'$. Then, the parameters to be estimated in this three-mixture case include: $b_{11}^* = 1$, $b_{12}^* = 2$, $b_{13}^* = 0.6$, $b_{21}^* = 2$, $b_{22}^* = 1.5$, $b_{23}^* = 0.2$, $b_{31}^* = 2$, $b_{32}^* = 0.5$, $b_{33}^* = 0.1$, $\tilde{b}^* = 0.5$, $l^*_{\boldsymbol{\Omega}_1,2} = 0.5$, $l^*_{\boldsymbol{\Omega}_1,3} = 0.866$, $l^*_{\boldsymbol{\Omega}_1,4} = 0.7$, $l^*_{\boldsymbol{\Omega}_1,5} = 0.519$, $l^*_{\boldsymbol{\Omega}_1,6} = 0.374$, $l^*_{\boldsymbol{\Omega}_2,1} = 0.9$, $l^*_{\boldsymbol{\Omega}_2,2} = 0.6$, $l^*_{\boldsymbol{\Omega}_2,3} = 0.8$, $l^*_{\boldsymbol{\Omega}_2,4} = 0.8$, $l^*_{\boldsymbol{\Omega}_2,5} = 0.4$, $l^*_{\boldsymbol{\Omega}_2,6} = 0.3$, $l^*_{\boldsymbol{\Omega}_3,1} = 2$, $l^*_{\boldsymbol{\Omega}_3,2} = 0.5$, $l^*_{\boldsymbol{\Omega}_3,3} = 1.0$, $l^*_{\boldsymbol{\Omega}_3,4} = 0.6$, $l^*_{\boldsymbol{\Omega}_3,5} = 0.8$, $l^*_{\boldsymbol{\Omega}_3,6} = 0.9$, $\mu_{21}^* = 0.6$, $\mu_{22}^* = 0.1$, $\mu_{31}^* = 0.4$, $\mu_{32}^* = 0.0$ $\gamma_{11}^* = 1$, $\gamma_{12}^* = 1$, $\gamma_{13}^* = 1$, $\gamma_{21}^* = 1$, $\gamma_{22}^* = 1$, $\gamma_{23}^* = 1$, $\gamma_{31}^* = 1$, $\gamma_{32}^* = 1$, $\gamma_{33}^* = 1$.

2.3.1.3 Data generation

Using the design presented in the previous sections, we generate the consumption quantity vector $\boldsymbol{x}_q^*$ for each individual using the forecasting algorithm proposed by Pinjari and Bhat (2011a). The above data generation process is undertaken 100 times with different realizations of the $\boldsymbol{\beta}_q$ vector to generate 100 different data sets each for the two- and three-mixture cases.

We estimate two additional models on each of the 100 generated data sets for each of the two- and three-mixture cases. The first model ignores random coefficients on the independent variables in each mixture (latent segment), allowing random coefficients only on the constants. This corresponds to the Latent Class-MCDP (or LC-MDCP) model. Thus, the only Cholesky parameters estimated for the two-mixture case are $l_{\boldsymbol{\Omega}_1,2} = 0.5$, $l_{\boldsymbol{\Omega}_1,3} = 0.866$, $l_{\boldsymbol{\Omega}_2,1} = 0.9$, $l_{\boldsymbol{\Omega}_2,2} = 0.6$, and $l_{\boldsymbol{\Omega}_2,3} = 0.8$. All other Cholesky parameters are effectively held to the value of zero. In Table 2b, for the LC-MDCP model, the only Cholesky parameters estimated are

$l^*_{\Omega_1,2} = 0.5$, $l^*_{\Omega_1,3} = 0.866$, $l^*_{\Omega_2,1} = 0.9$, $l^*_{\Omega_2,2} = 0.6$, $l^*_{\Omega_2,3} = 0.8$, $l^*_{\Omega_3,1} = 2$, $l^*_{\Omega_3,2} = 0.5$, and $l^*_{\Omega_3,3} = 1.0$.

The second model assumes away non-normality by using a single segment for the entire sample (that is, assumes that μ_{21} in the two-mixture case, and μ^*_{21} and μ^*_{31} in the three-segments case, all go to the value of $-\infty$). This is the traditional normally-distributed random-coefficients MDCP (or RC-MDCP) model. Also, in this case $l_{\Omega_2,1} = 0.9$ in the two-segment case, and $l^*_{\Omega_2,1} = 0.9$ and $l^*_{\Omega_3,1} = 2$ in the three-segment case, are not estimable and fixed at 1.0. Additionally, in the two segment case, the following constraints are imposed: $b_{11} = b_{21}$, $b_{12} = b_{22}$, $b_{13} = b_{23}$, $l_{\Omega_1,2} = l_{\Omega_2,2}$, $l_{\Omega_1,3} = l_{\Omega_2,3}$, $l_{\Omega_1,4} = l_{\Omega_2,4}$, $l_{\Omega_1,5} = l_{\Omega_2,5}$, $l_{\Omega_1,6} = l_{\Omega_2,6}$, $\gamma_{11} = \gamma_{21,}$, $\gamma_{12} = \gamma_{22}$, and $\gamma_{13} = \gamma_{23}$. In the three-segment case, the following constraints are imposed: $b^*_{11} = b^*_{21} = b^*_{31}$, $b^*_{12} = b^*_{22} = b^*_{32}$, $b^*_{13} = b^*_{23} = b^*_{33}$, $l^*_{\Omega_1,2} = l^*_{\Omega_2,2} = l^*_{\Omega_3,2}$, $l^*_{\Omega_1,3} = l^*_{\Omega_2,3} = l^*_{\Omega_3,3}$, $l^*_{\Omega_1,4} = l^*_{\Omega_2,4} = l^*_{\Omega_3,4}$, $l^*_{\Omega_1,5} = l^*_{\Omega_2,5} = l^*_{\Omega_3,5}$, $l^*_{\Omega_1,6} = l^*_{\Omega_2,6} = l^*_{\Omega_3,6}$, $\gamma^*_{11} = \gamma^*_{21} = \gamma^*_{31}$, $\gamma^*_{12} = \gamma^*_{22} = \gamma^*_{32}$, and $\gamma^*_{13} = \gamma^*_{23} = \gamma^*_{33}$.

We make the comparison between the proposed FDMN-MDCP model and the two restrictive formulations above (that is, the LC-MDCP and the RC-MDCP based on the ability to accurately recover model parameters as well as usual nested likelihood ratio tests).

The analytic approximation embedded in the MACML estimator is applied to two of the datasets 10 times with different permutations to obtain the approximation error. The approximation error is negligible, so only one set of permutations for computing the approximation will be considered in each of the 100 datasets. The performance of the MACML inference approach in estimating the parameters of the MDCP model and their standard errors is evaluated as follows:

(1) Estimate the parameters using the analytic approximation in the MACML for each data set *s*. Estimate the standard errors using the Godambe (sandwich) estimator.

(2) Compute the mean estimate for each model parameter across the data sets to obtain **a mean estimate**. Compute the **absolute percentage (finite sample) bias** (APB) of the estimator as:

$$APB(\%) = \left| \frac{\text{mean estimate - true value}}{\text{true value}} \right| \times 100.$$ [8]

(3) Compute the standard deviation for each model parameter across the data sets, and label this as the **finite sample standard error or FSSE** (essentially, this is the empirical standard error).

(4) Compute the median standard error for each model parameter across the data sets and label this as **the asymptotic standard error or ASE** (essentially, this is the standard error of the distribution of the estimator as the sample size increases).

(5) Next, to evaluate the accuracy of the asymptotic standard error formula as computed using the MACML inference approach for the finite sample size used, compute the APB associated with the ASE of the estimator as:

$$APBASE(\%) = \left| \frac{\text{ASE - FSSE}}{\text{FSSE}} \right| \times 100$$

2.3.2 Simulation results

2.3.2.1 Recoverability of parameters in the MDCP with the mixture model

Tables 2.1a and 2.1b present the results for the simulation. Table 2.1a corresponds to the two-segment case, while Table 2.1b corresponds to the three-segment case. The second column presents the true values used in generating the data samples. The third column labeled "Parameter Estimates" provides the mean value (across the data sets) of each parameter as well as the corresponding APB measure, while the fourth broad column labeled "Standard Error Estimates" provides the ASE, FSSE, and the APBASE values for the parameter standard errors.

[8] In case a true parameter value is zero, the APB is computed by taking the difference of the mean estimate from the true value (= 0), dividing this difference by the value of 1 in the denominator, and multiplying by 100.

Table 2.1a. Evaluation of the ability to recover true parameters for the two-segment case

Parameter	True Value	MACML Method				
		Parameter Estimates		Standard Error Estimates		
		Mean Estimate	Absolute Percentage Bias (APB)	Asymptotic Standard Error (ASE)	Finite Sample Standard Error (FSSE)	Absolute Percentage Bias of Asymptotic Standard Error (APBASE)
b_{11}	1.000	1.063	6.3%	0.157	0.136	14.8%
b_{12}	2.000	1.997	0.2%	0.387	0.438	11.7%
b_{13}	0.600	0.586	2.4%	0.063	0.061	3.2%
b_{21}	2.000	1.901	4.9%	0.419	0.407	2.9%
b_{22}	1.500	1.503	0.2%	0.141	0.136	3.5%
b_{23}	0.200	0.196	2.2%	0.032	0.035	8.2%
$\tilde{b}$	0.500	0.500	0.0%	0.013	0.012	12.4%
$l_{\Omega_1,2}$	0.500	0.476	4.7%	0.055	0.057	2.5%
$l_{\Omega_1,3}$	0.866	0.865	0.2%	0.040	0.044	9.6%
$l_{\Omega_1,4}$	0.700	0.666	4.9%	0.049	0.036	35.6%
$l_{\Omega_1,5}$	0.519	0.529	2.0%	0.051	0.049	3.0%
$l_{\Omega_1,6}$	0.374	0.378	1.0%	0.026	0.028	5.9%
$l_{\Omega_2,1}$	0.900	0.898	0.2%	0.023	0.021	10.7%
$l_{\Omega_2,2}$	0.600	0.598	0.4%	0.031	0.032	3.5%
$l_{\Omega_2,3}$	0.800	0.796	0.5%	0.021	0.020	4.0%
$l_{\Omega_2,4}$	0.800	0.795	0.6%	0.025	0.025	0.8%
$l_{\Omega_2,5}$	0.400	0.392	2.1%	0.021	0.018	17.2%
$l_{\Omega_2,6}$	0.300	0.299	0.3%	0.015	0.014	4.5%
μ_{21}	0.600	0.505	15.8%	0.206	0.183	12.9%
μ_{22}	0.100	0.112	11.9%	0.046	0.050	8.2%
γ_1	1.000	1.038	3.8%	0.125	0.115	8.4%
γ_2	1.000	1.008	0.8%	0.146	0.136	7.0%
γ_3	1.000	1.103	10.3%	0.396	0.395	0.2%
Overall Mean Value Across Parameters		**3.2% (for APB)**		**0.115**	**0.113**	**7.8%**
Mean Time (mins)		18.3				
Std. dev of Time		7.5				
% of Runs Converged		100%				

Table 2.1b. Evaluation of the ability to recover true parameters for the three-segment case

Parameter	True Value	MACML Method				
		Parameter Estimates		Standard Error Estimates		
		Mean Estimate	Absolute Percentage Bias (APB)	Asymptotic Standard Error (ASE)	Finite Sample Standard Error (FSSE)	Absolute Percentage Bias of Asymptotic Standard Error (APBASE)
b^*_{11}	1.000	1.003	0.3%	0.296	0.285	4.0%
b^*_{12}	2.000	1.885	5.8%	0.433	0.421	2.7%
b^*_{13}	0.600	0.547	8.8%	0.215	0.192	11.7%
b^*_{21}	2.000	1.844	7.8%	0.599	0.566	5.9%
b^*_{22}	1.500	1.432	4.5%	0.295	0.282	4.5%
b^*_{23}	0.200	0.206	3.2%	0.093	0.092	1.3%
b^*_{31}	3.000	3.378	12.6%	0.082	0.080	3.0%
b^*_{32}	1.300	1.235	5.0%	0.032	0.038	14.1%
b^*_{33}	0.300	0.346	15.3%	0.175	0.155	12.7%
$\tilde{b}^*$	0.500	0.499	0.3%	0.023	0.024	3.5%
$l^*_{\Omega_1,2}$	0.500	0.506	1.3%	0.137	0.124	10.1%
$l^*_{\Omega_1,3}$	0.866	0.863	0.4%	0.109	0.104	4.5%
$l^*_{\Omega_1,4}$	0.700	0.675	3.5%	0.082	0.072	14.7%
$l^*_{\Omega_1,5}$	0.519	0.500	3.7%	0.081	0.090	9.6%
$l^*_{\Omega_1,6}$	0.374	0.383	2.4%	0.056	0.052	7.6%
$l^*_{\Omega_2,1}$	0.900	0.920	2.3%	0.098	0.093	5.4%
$l^*_{\Omega_2,2}$	0.600	0.582	3.1%	0.075	0.074	0.4%
$l^*_{\Omega_2,3}$	0.800	0.790	1.3%	0.047	0.054	13.1%
$l^*_{\Omega_2,4}$	0.800	0.784	2.0%	0.105	0.084	24.5%
$l^*_{\Omega_2,5}$	0.400	0.401	0.3%	0.087	0.086	1.4%
$l^*_{\Omega_2,6}$	0.300	0.304	1.3%	0.058	0.057	0.9%
$l^*_{\Omega_3,1}$	2.000	2.030	1.5%	0.057	0.061	6.1%
$l^*_{\Omega_3,2}$	0.500	0.575	14.9%	0.067	0.053	25.8%
$l^*_{\Omega_3,3}$	1.000	0.986	1.4%	0.070	0.072	3.7%
$l^*_{\Omega_3,4}$	0.600	0.580	3.3%	0.050	0.067	25.5%
$l^*_{\Omega_3,5}$	0.800	0.886	10.7%	0.341	0.400	14.7%
$l^*_{\Omega_3,6}$	0.900	1.060	17.8%	0.060	0.055	8.5%
μ^*_{21}	0.600	0.687	14.6%	0.541	0.503	7.6%
μ^*_{22}	0.100	0.112	12.3%	0.148	0.149	1.0%
μ^*_{31}	0.400	0.342	14.4%	0.225	0.201	11.9%
μ^*_{32}	0.000	0.010	10.0%	0.008	0.007	14.2%
γ_1	1.000	1.041	4.1%	0.267	0.255	4.7%
γ_2	1.000	1.145	14.5%	0.329	0.308	6.8%
γ_3	1.000	1.156	15.6%	0.520	0.473	9.8%
Overall Mean Value		**6.4% (for APB)**		**0.172**	**0.165**	**8.7%**
Mean Time (mins)		72.4				
Std. dev of Time		19.6				

The APB values for the parameter estimates (third column) show that the MACML method does very well in recovering the parameters. The overall mean APB value across all parameters is 3.2% in the two-segment case (see the last row of the column labeled "APB" in Table 2.1a). The APB values are in general higher for the three-segment case (Table 2.1b), with an overall mean value of 6.4% across all parameters, probably due to the many additional parameters that have to be estimated relative to the two-segment model. In general, across the parameters, the APB values are relatively high for the $\boldsymbol{\gamma}$ satiation parameters in both the two- and three-segment cases. The satiation parameters are an important source of non-linearity in the overall utility function (see Equation 2.1), and make the likelihood surface more difficult to track computationally. The APB values of the $\boldsymbol{\mu}$ parameters are also relatively high in both cases (two and three segments) relative to the APB values of the rest of the parameters. These $\boldsymbol{\mu}$ parameters appear in the likelihood function through the mixture (π) probabilities, and it is well established in the literature (see, for example, Sobhani *et al*., 2013) that these mixture probabilities are difficult to pin down because the likelihood surface can be relatively flat for a number of different combinations of the mixture probabilities near the likelihood optimal point.

The finite sample standard errors and the asymptotic standard errors (in the fourth broad column of Tables 2.1a and 2.1b) are close; the average absolute difference is 0.007 and 0.013 for the two- and three-segment cases, respectively. The mean APBASE value across all parameters is 7.8% for the two-segment case and 8.7% for the three-segment case. In both the two- and three-segment cases, the finite sample standard error estimates are generally higher (as a percentage of the mean estimates) for the $\boldsymbol{\gamma}$ and $\boldsymbol{\mu}$ parameters relative to other sets of parameters, reinforcing the finding earlier that the $\boldsymbol{\gamma}$ and $\boldsymbol{\mu}$ parameters are more difficult to recover than other parameters. Some elements of the Cholesky matrix also are difficult to pin down, again because the Cholesky elements enter the likelihood function in a very non-linear fashion as part of the evaluation of the cumulative multivariate normal density and distribution functions.

Overall, the MACML inference approach does well in accurately and precisely recovering parameters in both the two-segment and three-segment FDMN-MDCP model. The reported model estimation times are based on scaling to a desktop computer with an Intel(R) Pentium(R) D CPU@3.20GHz processor and 4GB of RAM. The statistical software GAUSS was used for all the estimations reported in this chapter.

2.3.2.2 Comparison between the proposed model and more restrictive MDCP models

Tables 2.2a and 2.2b present the results for the simulation exercise focusing on the comparison between the proposed FDMN MDCP model and two other, more restrictive versions of the model: the LC-MDCP and the RC-MDCP models. Table 2.2a corresponds to the two-segment case, while Table 2.2b corresponds to the three-segment case. The APB values of the parameters are in general higher in both cases (two and three segments) and in both alternative models relative to the APB values of the parameters in the original model (Tables 2.1a and 2.1b). In the two-segment model, the overall mean APB values across parameters are 28.5% and 26.0% for the LC-MDCP and RC-MDCP models, respectively—significantly higher in comparison with the mean APB value of 3.2% in the proposed model. The difference is even higher in the three-segment model with the overall mean APB values across parameters being 30.8% and 82.9% for the LC-MDCP and RC-MDCP models, respectively, relative to the overall mean APB value of 6.4% in the original model. The superior performance of the FDMN-MDCP model is also evidenced in the higher log-likelihood value, on average, for the FDMN-MDCP model across the 100 estimations (on the 100 data sets). In addition, for each of the 100 data sets, a likelihood ratio test comparing the FDMN-MDCP model with the two other models clearly rejects the other two model in favor of the FDMN-MDCP model (see last row of Tables 2.2a and 2.2b).

Table 2.2a. Effects of ignoring continuous heterogeneity and non-normality in the two-segment model

Parameter	True Value	Latent Class MDCP (LC-MDCP) Model		Random Coeffs. MDCP (RC-MDCP) Model	
		Mean Estimate	Absolute Percentage Bias (APB)	Mean Estimate	Absolute Percentage Bias (APB)
b_{11}	1.000	1.203	20.3%	1.349	34.9%
b_{12}	2.000	1.543	22.9%	1.690	15.5%
b_{13}	0.600	0.890	48.3%	0.337	43.8%
b_{21}	2.000	1.293	35.4%	1.349	32.6%
b_{22}	1.500	1.402	6.5%	1.690	12.7%
b_{23}	0.200	0.289	44.5%	0.337	68.7%
$\tilde{b}$	0.500	0.654	30.8%	0.427	0.4%
$l_{\Omega_1,2}$	0.500	0.592	18.4%	0.376	24.8%
$l_{\Omega_1,3}$	0.866	0.965	11.4%	0.572	34.0%
$l_{\Omega_1,4}$	0.700	--[a]	--	0.942	34.6%
$l_{\Omega_1,5}$	0.519	--[a]	--	0.626	20.6%
$l_{\Omega_1,6}$	0.374	--[a]	--	0.407	8.8%
$l_{\Omega_2,1}$	0.900	0.782	13.1%	--[b]	--
$l_{\Omega_2,2}$	0.600	0.329	45.2%	0.376	37.3%
$l_{\Omega_2,3}$	0.800	0.764	4.5%	0.942	17.8%
$l_{\Omega_2,4}$	0.800	--[a]	--	0.572	28.5%
$l_{\Omega_2,5}$	0.400	--[a]	--	0.626	56.4%
$l_{\Omega_2,6}$	0.300	--[a]	--	0.407	35.6%
μ_{21}	0.600	0.431	28.2%	--[c]	--
μ_{22}	0.100	0.140	40.0%	--[c]	--
γ_1	1.000	1.209	20.9%	1.030	3.0%
γ_2	1.000	1.823	82.3%	1.405	40.5%
γ_3	1.000	1.117	11.7%	1.017	1.7%
Overall Mean Value Across Parameters		**28.5% (for APB)**		**26.0% (for APB)**	
Mean (across 100 data sets) log-likelihood value at convergence		**-39,517.923**		**-39,561.115**	
Number of times the likelihood ratio test statistic favors the FDMN-MDCP model[d]		All one hundred times when compared with $\chi^2_{6,0.95} = 12.59$		All one hundred times when compared with $\chi^2_{13,0.95} = 22.36$	

[a] These parameters are not estimated and are fixed at 0.0.
[b] This parameter is fixed to 1.0 for identification.
[c] These parameters are implicitly fixed to the value of minus infinity.
[d] The mean (across data sets) log-likelihood value at convergence for the FDMN-MDCP model with a two-segment mixture is -38,927.438.

Table 2.2b. Effects of ignoring continuous heterogeneity and non-normality in the three-segment model

Parameter	True Value	Latent Class MDCP Model		Random Coeff. MDCP (RC-MDCP) Model	
		Mean Estimate	Absolute Percentage Bias (APB)	Mean Estimate	Absolute Percentage Bias (APB)
b^*_{11}	1.000	1.543	54.3%	1.738	73.8%
b^*_{12}	2.000	1.276	36.2%	1.277	36.1%
b^*_{13}	0.600	0.320	46.7%	0.246	59.0%
b^*_{21}	2.000	1.652	17.4%	1.738	13.1%
b^*_{22}	1.500	1.724	14.9%	1.277	14.9%
b^*_{23}	0.200	0.102	49.0%	0.246	23.1%
b^*_{31}	3.000	1.592	46.9%	1.738	42.1%
b^*_{32}	1.300	1.035	20.4%	1.277	1.8%
b^*_{33}	0.300	0.472	57.3%	0.246	17.9%
$\tilde{b}^*$	0.500	0.366	26.8%	0.418	16.4%
$l^*_{\Omega_1,2}$	0.500	0.411	17.8%	0.063	87.4%
$l^*_{\Omega_1,3}$	0.866	0.599	30.8%	0.926	6.9%
$l^*_{\Omega_1,4}$	0.700	--[a]	--	0.322	54.0%
$l^*_{\Omega_1,5}$	0.519	--[a]	--	0.618	19.1%
$l^*_{\Omega_1,6}$	0.374	--[a]	--	0.546	46.1%
$l^*_{\Omega_2,1}$	0.900	0.724	19.6%	--[b]	--
$l^*_{\Omega_2,2}$	0.600	0.326	45.7%	0.063	89.5%
$l^*_{\Omega_2,3}$	0.800	0.598	25.3%	0.322	59.7%
$l^*_{\Omega_2,4}$	0.800	--[a]	--	0.926	15.7%
$l^*_{\Omega_2,5}$	0.400	--[a]	--	0.618	54.5%
$l^*_{\Omega_2,6}$	0.300	--[a]	--	0.546	82.1%
$l^*_{\Omega_3,1}$	2.000	1.396	30.2%	--[b]	--
$l^*_{\Omega_3,2}$	0.500	0.398	20.4%	0.063	87.4%
$l^*_{\Omega_3,3}$	1.000	0.733	26.7%	0.322	67.8%
$l^*_{\Omega_3,4}$	0.600	--[a]	--	0.926	54.3%
$l^*_{\Omega_3,5}$	0.800	--[a]	--	0.618	22.7%
$l^*_{\Omega_3,6}$	0.900	--[a]	--	0.546	39.3%
μ^*_{21}	0.600	0.467	22.2%	--[c]	--
μ^*_{22}	0.100	0.156	56.0%	--[c]	--
μ^*_{31}	0.400	0.298	25.5%	--[c]	--
μ^*_{32}	0.000	0.017	17.0%	--[c]	--
γ_1	1.000	1.327	32.7%	2.723	172.3%
γ_2	1.000	1.201	20.1%	2.959	195.9%
γ_3	1.000	1.102	10.2%	2.871	187.1%

Parameter	True Value	Latent Class MDCP Model		Random Coeff. MDCP (RC-MDCP) Model	
		Mean Estimate	Absolute Percentage Bias (APB)	Mean Estimate	Absolute Percentage Bias (APB)
Overall Mean Value Across Parameters		**30.8% (for APB)**		**82.9% (for APB)**	
Mean (across 100 data sets) log-likelihood value at convergence		**-39,599.201**		**-39,797.634**	
Number of times the likelihood ratio test favors the FDMN-MDCP model[d]		All one hundred times when compared with $\chi^2_{9,0.95} = 16.92$		All one hundred times when compared with $\chi^2_{27,0.95} = 40.11$	

[a] These parameters are not estimated and are fixed at 0.0
[b] This parameter is fixed to 1.0 for identification.
[c] These parameters are implicitly fixed to the value of minus infinity.
[d] The mean (across data sets) log-likelihood value at convergence for the FDMN-MDCP model with a three-segment mixture is -39,001.232.

2.4 APPLICATION TO TOURISM TRAVEL

In this section, we demonstrate an application of the proposed model to analyze individual-level decisions regarding recreational destination locations and the number of trips to each destination, using data drawn from the 2012 New Zealand Domestic Travel Survey (DTS).

2.4.1 Empirical context

Tourism has been an important contributor to New Zealand's economy, thanks to the natural and beautiful landscape of the compact island country that also offers an extensive coastline for trekking, swimming, fishing, other water-based activities, and sports. In addition, New Zealand also boasts of some excellent vineyards, offers volcanic/geothermal excursion opportunities, and its forests and pristine landscape have made it a much sought-after location for mainstream Hollywood movies (for example, the Fiordland and Southern Lakes in the southern part of New Zealand were the locations for the mythical Middle Earth in the "Lord of the Rings" trilogy). Overall, tourism contributes 9% of New Zealand's gross domestic product and is also an important source of employment; 10% of New Zealanders work in the tourism industry (see New Zealand Tourism Strategy 2015).

Although the international popularity of New Zealand has increased enormously in the past few years, domestic tourism continues to remain a significant source of income for the tourism industry. According to the New Zealand Tourism Industry Association (TIA, 2012), domestic travelers (New Zealand residents traveling within New Zealand) accounted for about 57% of New Zealand's total tourism industry spend of $23 billion in 2012 (see Statistics New Zealand, 2013). The substantial amount of domestic tourism may be attributed to increased

marketing efforts of leisure activity opportunities within the island nation and more control of the leisure vacation experience through on-line sites. However, it is also a result of a general trend across all countries around the globe of an increasingly compact geographic footprint of leisure travel, spurred by a shift from the traditional long period vacations undertaken during holidays or over the summer to short period leisure travel built around the work weeks (see, for example, White, 2011 and LaMondia and Bhat, 2012). This shift itself may be traced to easier schedule coordination opportunities for short duration leisure pursuits around work weeks, especially for the increasing number of families with multiple working individuals with school-going children.

The growing amount of short distance leisure trips, mostly undertaken using the personal auto mode, has led to increased attention on this leisure travel market among urban transportation planners because of the increased weekend day traffic on city streets and between cities in close proximity, and the concomitant effects on traffic congestion and air quality. Understanding these travel flow patterns can help planning and policy efforts to reduce the negative externalities of such travel. At the same time, unraveling the "push and pull' factors associated with individual and household leisure activity decisions helps cities and regions position themselves as unique and even exotic destinations, with an eye on generating jobs and revenue. This confluence of interest on leisure travel from the transportation and tourism domains has led to many studies in this space in the past decade, with a particular emphasis on destination choice for leisure pursuits. While the early literature in the area considered leisure destination choices as repeated isolated (and independent) decision events for each leisure trip, the more recent literature has moved toward the more realistic representation of destination choices as inter-related decisions for multiple leisure trips over a longer-term period of a month or even a year. Examples of the latter string of multiple discrete-continuous (MDC) studies (with the discrete component being the choice of destination region, and the continuous component being the number of trips to each chosen destination region) include Kuriyama *et al*. (2010), (2011), Van Nostrand *et al*. (2013), von Haefen (2007), Whitehead *et al*. (2010), LaMondia *et al*. (2010), and Bhat *et al*. (2013a). These studies explicitly accommodate variety-seeking and loyalty behavior by considering satiation effects based on Iso-Ahola's (1983) theory of vacation participation in which the individual/family balances needs for familiarity and novelty, within long period budget constraints, to provide an "optimally arousing experience" (see LaMondia *et al*., 2008 for a detailed discussion). In this dissertation, we contribute to leisure destination

choice modeling using the proposed FDMN MDCP model. To our knowledge, this is the first such application in the leisure travel literature.[9]

2.4.2 Data description

The data for this study is derived from three sources. The primary source, as mentioned earlier, is the 2012 New Zealand DTS, which asked survey respondents (New Zealand's residents) to provide information on all one-way trips 40 kilometers or longer from home, overnight trips from home, and flight or ferry trips from home made up to four weeks prior to the survey date (see Ministry of Business, Innovation and Employment, 2013). The survey was targeted at individuals and not households in that only one randomly selected individual (over the age of 15 years) from each sampled household was interviewed. Telephone interviewing was used for the DTS and household telephone numbers were randomly selected from the white pages. Interviews were carried out according to pre-specified quotas for age, sex and region of origin. The process of data collection took place continuously throughout the year.

The survey obtained information on the resident city of the respondent, the city of destination for each trip, the primary reason of each trip, and the primary mode of transportation used to reach the destination. Additionally, the survey also obtained individual and household socio-demographic information. A second data source is a network level of service file that provided information on land travel distance and highway travel time between each city pair within New Zealand (see additional details in the next paragraph). The third data source is a disaggregate spatial land-cover characteristics data obtained from the 2012 Land Cover Database

[9]As with all the earlier MDC leisure studies, this study too focuses on the count of the number of times each leisure destination is visited. Thus, the "continuous" quantity used is actually a count variable, as opposed to a truly continuous measure as required by the theoretical model. But, as demonstrated by von Haefen and Phaneuf (2003), treating the integer count of trips as a continuous variable (within an MDC framework) does not lead to substantial bias in the results or the behavioral implications. This forms the basis for the use of the MDC framework in earlier studies, as well as in the current study, of leisure destination choices over a period of time. Similarly, as in earlier studies, the budget in the MDC formulation is the total number of leisure trips made over a given time period. This budget is "allocated" to the different possible discrete leisure destination locations. While a more reasonable approach would be to allocate a money budget, the operationalization of this alternative approach is extremely difficult because of the many assumptions that need to be made regarding monetary costs of participation per trip. Besides, a more practical problem is that expenditure information is rarely obtained in travel surveys. On the other hand, the number of trips to each destination is readily available from a sample of individuals in a survey, and the total trips (or "budget") is readily obtained by aggregating across the possible destination locations. A related issue in the use of total trips as the "budget" is that the MDC models of leisure destination choice focus on the count of trips to each destination, given the total number of leisure trips during a specified period. In forecasting mode, the latter "budget quantity" is itself predicted in an earlier "trip generation" step, including the choice of making no leisure trips at all.

(LCDB) of the Land Resource Information System (LRIS) of New Zealand. The LCDB provides land-cover information at a 30 meters by 30 meter resolution. From this data, using a geographic information system based procedure, we developed total land area and acreage information for each 30x30 meter^2 grid and by six broadly defined land-cover categories: urban area (including central business districts, commercial and industrial areas, urban parklands, urban dumps, and housing and transportation-related land cover), water area (including rivers, land/ponds, freshwater, and estuarine open water), wetland area (context-dependent combinations of areas such as herbaceous freshwater vegetation, flaxland, and saline vegetation), agricultural area (including vineyards and orchards, perennial crops, short rotation cropland, and grasslands), bare-land area, and forest area (pine forests, mangroves, deciduous hardwoods and other exotic/indigenous forest areas).

The sample formation comprised several steps. First, we selected only leisure trips to primary destinations within New Zealand undertaken by a personal auto (personal auto trips comprise around 90% of all leisure domestic trips within New Zealand; see Ministry of Business, Innovation and Employment, 2008). Second, the leisure destination cities in New Zealand were mapped into one of 16 aggregate destination regions in the current analysis, as identified in Figure 2.1. Nine regions are in the North Island, while seven are in the South Island. This regional classification scheme is the same as that used by the Department of Tourism of New Zealand for its marketing campaigns, and is also the commonly used geo-political partitioning of the country. Third, the total number of trips made by each individual to each region was obtained by appropriate aggregation across trips to cities within each region, and the individual-level trip budget is obtained as the total number of trips of the individual across all regions during the four week period. Fourth, we identified a centroidal city for each of the 16 destination regions, based on the city that attracted the most travelers within each region, and converted the city-to-city land-based travel distance and land-based travel time data to corresponding residence city-to-destination region skims. But travel from one region in one island to another region in another island by auto is possible only through the use of a ferry service (that transports vehicles too) across the Cook Strait between Wellington in the North (located in the Wellington region) and Picton in the South (located in the Marlborough region). On the other hand, the land-based travel time between two regions in different islands from earlier includes only the travel time from the origin point to one of the two ferry terminals plus the travel time from the other ferry terminal to

the destination region. Thus, the total travel time between two regions in different islands should include the 3 hour 15 minute cruise (including ferry terminal times) between the north and south islands. At the end of this step, we obtain the land-based travel distance and the total travel time for each residence city-destination region pairing. Fifth, the travel cost skims were computed as a function of the respondent's reported household income, the estimated cost of vehicle fuel on land, the ferry cost if a ferry crossing is involved, and the land-based distance and total travel time skims (obtained in the previous step) between the respondent's residence city and the centroidal city of each destination region. To calculate the travel cost, we followed the standard approach of valuing travel time at a fixed proportion of one-half of the wage rate (see Hanemann *et al*., 2004 for a detailed discussion). Specifically, the travel cost was computed as:

> Cost (in NZ$) = 2 * (one-way land travel distance in miles * fuel cost per mile + one-way total travel time in hours * (0.5 * hourly wage)) + round-trip ferry cost (as applicable).

The fuel cost per mile is computed at NZ$0.149 per mile based on a fuel cost of NZ$1.75 per liter and a rather high vehicle efficiency factor of 5.3 liters for 100 km (5.3 liters for 62.1 miles or about 44 miles per gallon), given the long distance nature of trips under consideration. The round-trip ferry cost is NZ$145. Sixth, the grid-based land-cover data were translated to a destination region-based land-cover data by suitable aggregation over cells within each destination region. Seventh, individual and household socio-demographic, as well as land cover data by region, were appended to the long distance travel records.

The final data sample used in the estimation included 3508 individuals. Table 2.3 provides the distribution of these individuals by the number of leisure trips made during the four week period before they were surveyed and by the number of distinct leisure destination regions visited. Although a sizeable fraction (72.3%) of the individuals in the sample make only one trip, a non-insignificant percentage of individuals (27.7%) make more than one trip. Most of the individuals who undertake more than one trip during the survey period prefer to travel to multiple destinations (see the second row and beyond in Table 2.3). For example, 53.3% of individuals making two trips during the survey period visit more than one distinct destination region, while 65% of individuals making three trips visit more than one distinct region. The corresponding numbers are 70.2% and 78.6% for individuals who make four and five or more trips, respectively, during the survey period. Clearly, this is a case of multiple discreteness for individuals who make more than one trip.

Source: www.stats.govt.nz

Figure 2.1. Boundaries of New Zealand regions

Table 2.4 provides descriptive statistics for each of the 16 destination regions. The third broad column presents the mean and standard deviations for the travel impedance skim measures of total travel time, travel distance, and travel cost for each destination region (computed from the residence city-destination region skims developed as discussed earlier in this section). Not

surprisingly, the travel impedance measures are the highest for the Northland region in the North Island (the northernmost region) and the Southland region in the South Island (the southernmost region). As expected, the impedance measures decrease as one gets closer to the center of the country. Interestingly, the impedance measures are lower for the North Island regions compared to the South Island regions. This is because of two-interrelated factors. First, the North Island is

Table 2.3. Recreational travel number of trips

Number of trips	Number of individuals	Number (%) of individuals visiting[a]				
		1 region	2 regions	3 regions	4 regions	5 regions
1	2,535 (72.3%)	2,535 (100%)	0	0	0	0
2	732 (20.9%)	342 (46.7%)	390 (53.3%)	0	0	0
3	180 (5.0%)	63 (35%)	87 (48.3%)	30 (16.7%)	0	0
4	47 (1.3%)	14 (29.8%)	23 (48.9%)	7 (14.9%)	3 (6.4%)	0
5	7 (0.2%)	1 (14.3%)	2 (28.6%)	3 (42.8%)	0	1 (14.3%)
6	3 (0.1%)	2 (66.7%)	1 (33.3%)	0	0	0
7	2 (0.1%)	0	1 (50%)	1 (50%)	0	0
10	2 (0.1%)	0	0	2 (100%)	0	0

[a] Percentages add up to 100% in each row.

Table 2.4. Destination region characteristics

Island	Region	Travel Impedance Measures (Std. Dev.)			Land Cover Percentage					
		Travel Time (hours)	Travel Distance (miles)	Cost (NZ$)	Urban	Water	Wetland	Agricultural	Bare-land	Forest
NORTH ISLAND	Northland	8.31 (6.53)	397.6 (303.9)	314.1 (334.2)	0.75	2.44	0.92	47.92	1.18	46.79
	Auckland	6.53 (6.64)	306.4 (295.6)	265.8 (316.4)	10.68	2.85	0.62	49.04	0.92	35.90
	Waikato	5.98 (5.97)	273.9 (257.8)	241.0 (281.3)	1.14	3.57	0.88	53.10	0.70	40.61
	Bay of Plenty	6.74 (5.60)	313.3 (237.2)	276.9 (272.4)	1.32	2.39	0.27	23.17	0.28	72.57
	Gisborne	7.82 (4.95)	366.2 (205.5)	322.1 (260.1)	0.35	0.36	0.41	46.44	1.55	50.89
	Taranaki	6.41 (4.32)	294.5 (169.9)	261.9 (219.1)	0.98	0.39	0.08	53.81	0.43	44.31
	Manawatu-Wanganui	6.07 (3.72)	279.7 (152.5)	249.2 (198.3)	0.67	0.48	0.32	60.22	0.82	37.50
	Hawke´s Bay	6.30 (4.40)	290.5 (175.1)	258.7 (222.5)	0.59	0.92	0.22	53.93	0.64	43.70
	Wellington	6.46 (3.19)	301.5 (151.4)	266.7 (195.5)	2.53	1.34	0.23	47.82	0.72	47.36
SOUTH ISLAND	Tasman	9.70 (3.74)	392.2 (162.5)	386.1 (266.7)	0.34	1.34	1.29	19.58	3.37	74.08
	Nelson	9.58 (3.61)	388.0 (161.0)	381.6 (261.8)	6.48	3.12	0.25	13.72	0.93	75.49
	Marlborough	8.34 (3.43)	337.4 (156.3)	332.1 (237.3)	0.28	0.56	0.20	43.34	9.97	45.65
	West Coast	10.86 (4.83)	474.7 (210.6)	447.2 (322.4)	0.14	1.43	1.35	15.79	9.41	71.88
	Canterbury	10.48 (5.52)	443.6 (241.6)	425.3 (338.6)	0.71	2.09	0.36	65.67	12.12	19.05
	Otago	14.07 (6.56)	629.5 (291.1)	580.0 (425.9)	0.45	2.76	1.50	73.49	4.73	17.07
	Southland	16.40 (6.86)	749.7 (307.3)	680.9 (471.0)	0.24	2.99	0.98	43.97	4.50	47.32

more populated relative to the South Island (the North Island's population is about 3.2 million, while that of the South Island is about 1 million), which should result in more leisure trips generated from the North Island due to a sheer population size effect. Second, because of the compact nature of the North Island, there are more leisure trips generated per capita in the North than in the South, and most of these trips are destined to within the compact North Island. The net result is that, if one were to draw a horizontal "residential center of gravity" (RCG) line of tourists, it would go through the boundary of the Waikato and Manawaku-Wanganul (MW) regions in the North Island (see Figure 2.1). This is also evidenced in Table 2.4 in that the impedance measures are the smallest for the Waikato and MW regions, and increase as one goes farther away from the horizontal RCG line. Additionally, we should also note that, of the 3508 individuals in the sample, 2588 (73.7%) percent reside in the North Island, and 662 (18.9%) reside in the Waikato-MW regions. The fourth broad column in Table 2.4 provides the percentage of land in each region in each of the six land cover categories (the sum across all columns for each row add up to 100%). Of all the regions, Auckland has the highest percentage of urban land-cover, with Nelson and Wellington being the regions with the second and third highest urban land cover percentages. As we will see later, the high urban land cover is correlated with the intensity of tourist draw. In terms of wetland cover percentages, the highest are for Tasman, West Coast, Otago, and Southland. Nelson is the region with the highest forest land cover.

Table 2.5 provides additional descriptive statistics of the area of each region and destination region characteristics. The third column of the table presents the area of each region. As can be observed from this column and also from Figure 2.1, Canterbury in the south island is the largest region by size across all regions, while Waikato and MW are the largest regions in the North Island. The fourth column shows the number (and the corresponding percentage) of individuals who visited each region at least once. The Waikato region is clearly the one patronized by the most number of individuals, but Auckland, Bay of Plenty, and Canterbury also draw quite a few individuals. However, to get a better picture of attractiveness, the fifth column normalizes the number of people visiting by the area of each region (to accommodate for the fact that there are likely to be size effects here; that is, the larger a region, the more likely it is to be a destination). This column shows that on a per unit area basis, Auckland is by far the most popular destination, followed by Wellington and Nelson. Interestingly, as indicated earlier, these

are the three destinations with the highest percentages of urban land cover, and Nelson is the region with the highest forest cover. The Auckland region includes the famous urban tourist attraction of the City of Auckland as well as such attractions as the Tiritiri Matangi Islands, a haven for nature hikers who want to experience the rich flora and fauna of the region up close (especially of a host of endangered species of birds, each with a unique bird call pattern). The Wellington region, with Wellington City that serves as the capital of the North Island, is well known for Mt. Victoria (that provides a nice walk trail and panoramic views of the city and the Wellington harbor), massage and waxing boutiques in the Lower Hutt area also overlooking the Wellington harbor, and an interactive national museum of New Zealand culture and heritage. Finally, the Nelson region in the north of the South Island, the smallest of all the regions but also the sunniest in all of New Zealand, includes the city of Nelson. Nelson is renowned for its Maori (indigenous Polynesian tribe of New Zealand) arts and craftsmanship, water sports and activities (the Nelson region has the second largest amount of land percentage covered by water, and is liberally sprinkled with freshwater springs, especially near Takaka), and hiking/biking trails in the Abel Tasman National Park and other pristine forest land. Also interesting to note is that Tasman, West Coast, Otago, and Southland are some of the regions with the lowest number of visiting individuals per unit area, and these regions all have a relatively high wetland cover percentage as identified earlier, suggesting an inverse relationship between wetland cover percentage and tourist draw (perhaps because there is little to do within wetlands). The sixth broad column presents statistics on the number of visits to a destination region among those who visited the destination region at least once. The mean and maximum values from this column suggest that Auckland, Waikato, Bay of Plenty, Wellington, Canterbury, and Otago have the most loyal following.

Table 2.5. Recreational travel destination choice and number of trips

Island	Destination Region	Area (miles^2)	Total number (%) of individuals visiting each region*	Number of visiting individuals per unit area (per miles^2)	Number of trips among those who visit each destination			
					Mean	Min.	Max.	Std. Dev.
NORTH ISLAND	Northland	5,383	290 (8.3%)	0.0539	1.16	1	4	0.44
	Auckland	2,162	575 (16.4%)	0.2660	1.17	1	6	0.49
	Waikato	9,883	788 (22.5%)	0.0798	1.19	1	7	0.53
	Bay of Plenty	4,806	454 (12.9%)	0.0945	1.20	1	8	0.61
	Gisborne	3,224	42 (1.2%)	0.0129	1.17	1	4	0.53
	Taranaki	2,808	104 (3.0%)	0.0370	1.12	1	3	0.35
	Manawatu-Wanganui	8,577	288 (8.2%)	0.0337	1.13	1	4	0.38
	Hawke´s Bay	5,469	185 (5.3%)	0.0339	1.09	1	3	0.31
	Wellington	3,137	328 (9.4%)	0.1046	1.18	1	4	0.47
SOUTH ISLAND	Tasman	3,778	70 (2.0%)	0.0186	1.16	1	3	0.50
	Nelson	172	31 (0.9%)	0.1805	1.06	1	2	0.25
	Marlborough	4,820	74 (2.1%)	0.0153	1.07	1	2	0.25
	West Coast	9,010	77 (2.2%)	0.0085	1.13	1	4	0.47
	Canterbury	17,508	465 (13.3%)	0.0267	1.21	1	6	0.52
	Otago	12,351	260 (7.4%)	0.0210	1.22	1	6	0.56
	Southland	13,261	80 (2.3%)	0.0060	1.06	1	2	0.24

* Total percentage across all rows in this column add up to more than 100% because some travelers visit more than one destination region.

2.4.3 Variable specification and model formulation

The number of destination region alternatives in the MDCP model is 16. Thus, rather than including 15 alternative-specific constants in the baseline preference and 16 region-specific satiation parameters (in addition to other explanatory variables) in each latent segment, we adopted an "unlabeled" MDCP specification in which the baseline preferences and satiations are captured through attributes of the individual regions. For identification in this unlabeled alternatives context, the constant for the first segment is constrained to zero, and the constants for other segments are constrained to be descending from the second segment forward.

2.4.3.1 Baseline preference specification

The first independent variable we used in the baseline preference (that is, as part of the $\boldsymbol{z}_{qk}$ vector in Equation (2.2)) is the logarithm of the area of each region, to proxy for the number of elemental destination opportunities within each aggregate region (see Bhat *et al*., 1998). The expectation is that large regions are more likely to be chosen as a recreation destination based on a sheer "volume of opportunities" effect. The coefficient on this size variable may be viewed as an inclusive value characterizing the presence of common unobserved destination region attributes affecting the utility of elemental alternatives within each region. As in traditional discrete choice models, we expect this coefficient to be positive and less than one. If less than one, the implication is that there are common unobserved region attributes that lead to higher sensitivity across elemental alternatives within a region than across different regions. The net effect is that there is an inelastic influence of increasing region size on the region's baseline utility. That is, compared to the case when the coefficient is one, the rise in the baseline utility of a region due to an increase in the region's size is much less when the coefficient is estimated to be less than one in magnitude (because of more redistribution of leisure trips across elemental destinations within the same region rather than across different regions).

The next set of variables we considered are land-cover effects, captured by interacting the land-cover percentage by category in each destination region with the travel time from each individual's residence city to the centroidal city of each destination region. We computed a land-cover accessibility measure of the Hansen-type (Fotheringham, 1983) for individual q and land-cover type i as presented by destination region k as $AC_{qki}=LC_{ki}/[f(TT_{qk})]$, where LC_{ki} is the percentage area in land-cover category i (i = urban, water, wetland, agricultural, bare-land, and

forest) in destination region *k*, TT_{qk} is the travel time (in hours) from individual *q*'s residence city to the centroid of destination region *k*, and *f*(.) is a function.[10] The accessibility measures proxy the intensity of opportunities for recreational participation specific to each land-use category in a destination region normalized by a measure of impedance (function of travel time) for individual *q* to reach those opportunities. In the empirical analysis, a host of functional forms can be tested for the travel time measure. In our specifications, we considered both a linear form, $f(TT_{qk}) = TT_{qk}$, as well as a logarithmic form, $f(TT_{qk}) = \ln(TT_{qk})$. The logarithmic form penalizes destination regions less for being far away from the residential location of the individual. In both cases, a positive coefficient on an accessibility measure implies that individuals are attracted toward proximal destination regions with a substantial percentage of area in the corresponding land use. Our expectation, based on the descriptive statistics, is a positive coefficient on the urban land cover accessibility variable, though things are less clear from the descriptive analysis regarding the nature of effects of other accessibility variables. Based on our specification tests, the linear form is the preferred functional form for $f(TT_{qk})$.

The land cover-based accessibility effects (which are specific to each land cover category) capture any preferences individuals have for specific types of activities that may be featured in each destination region (as manifested in the land-cover category percentages). However, these effects do not capture an overall diversity index for each destination region. That is, it is possible that some individuals may be drawn to destination regions that have a good diversity of activity participation opportunities as well as are relatively close by. We proxy this effect by constructing a diversity index of land-cover types for each destination region, based on generalizing a similar index proposed originally by Bhat and Gossen (2004). This land cover diversity index is computed as a fraction between 0 and 1 for each destination region. Regions with a value closer to one have a richer land-cover mix than regions with a value closer to zero. The actual form of the land-cover diversity index for destination region *k* is:

[10] We do not introduce the land-cover percentages themselves directly in the baseline preference because these percentages do not vary across individuals in the sample. Thus, destination region land-cover percentages by themselves do not provide adequate variation to estimate parameters (because there are only 16 destination regions). But, by interacting these land cover percentages with individual-specific travel times to each region, we obtain rich variation across individuals in the resulting accessibility measures.

Land-cover diversity index $D_k = 1 - \left\{ \frac{\sum_{i=1}^{I} \left| LC_{ki} - \frac{100}{I} \right|}{\left(\frac{200(I-1)}{I} \right)} \right\}$, (2.14)

where LC_{ki} is the percentage area in land-cover category i in destination region k (as earlier) and I=6 (that is, we have six land cover categories) in our empirical context. The functional form would assign the value of zero if a region's land-cover is only in one category, and would assign a value of 1 if a region's land-cover is equally split among the different land-cover categories. However, as in the case of the land-cover percentages, there is no variation in the diversity index for a region across individuals, and the only variation in the index is across the 16 destination regions. This is inadequate to estimate a parameter on the diversity index, and thus we introduce the diversity accessibility index by normalizing the diversity index by a function of travel time to obtain individual-specific diversity accessibility indices: $DA_{qk} = D_k / f(TT_{qk})$. As earlier, we test both a linear form and a logarithmic form for the effect of travel time in the denominator of this expression. The best data fit results were again obtained consistently with the linear form.

Another variable considered in the specifications was the travel cost to each destination region, with the expectation that a higher cost would deter visiting the corresponding region. Again, both a simple linear form as well as a logarithmic form were tested for this cost effect, with the linear form winning out as the preferred one in our empirical tests. In addition, we included a dummy variable for the presence of a ferry ride. This accommodates any positive leisure/relaxation value of the ferry ride itself, after accounting for the total travel time effect.

A continuous random coefficient specification is considered on all of the above variables in the baseline preference for each discrete mixture (that is, each latent segment).

Finally, there is one other important issue with regard to the baseline preference specification. As discussed earlier, we use an unlabeled system for the alternatives, which essentially means that we constrain the mean coefficients on the alternative specific constants to be zero in the baseline utility for each destination region alternative. That is, the elements of $\boldsymbol{b}_g$ (in the notation of Section 2.2) corresponding to the 15 alternative-specific dummy variables for each latent segment g in $\boldsymbol{z}_{qk}$ are set to zero. However, we allow random covariance about this mean of zero. That is, the 15 elements of $\tilde{\boldsymbol{\beta}}_{qg}$ corresponding to the alternative-specific constants

are included with a covariance matrix. Assume that the random coefficients on the alternative-specific constants (ASCs) are independent of the random coefficients on other independent variables. Let $\tilde{\boldsymbol{\beta}}_{g,ASC}$ be a vector that collects the random coefficients corresponding to the 15 ASCs for each segment g. Then, the simplest specification for the covariance matrix of the 15 ASCs (for each segment g) obtained as differences of the original 16 ASCs from the first ASC (corresponding to the Northland region) would be as below (which originates from a specification of independently and identically distributed (IID) random errors with a variance of 0.5 for each of the original 16 ASCs):

$$\tilde{\boldsymbol{\beta}}_{g,ASC} = \boldsymbol{\varepsilon}_{g,ASC} \sim f_{15}(\mathbf{0}, \boldsymbol{\Lambda}), \text{ where } \boldsymbol{\Lambda} = \begin{bmatrix} 1 & 0.5 & 0.5 & \dots & 0.5 \\ 0.5 & 1 & 0.5 & \dots & 0.5 \\ 0.5 & 0.5 & 1 & \cdots & 0.5 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0.5 & 0.5 & 0.5 & \cdots & 1 \end{bmatrix}_{15\times 15 \text{ matrix}} . \quad (2.15)$$

However, there is likely to be spatial correlation across the utilities of the different regions because of similarity in unobserved attributes across proximally located regions. But, we have to assume that one region is not spatially correlated with all the other regions (because only differences in the baseline utilities matter). In our analysis, the first region (that is, the Northland region) will play this base role. We then accommodate spatial correlation across other regions using a spatial autoregressive (SAR) error structure of order one for the random components of the ASCs of the other 15 regions as follows:

$$\tilde{\boldsymbol{\beta}}_{g,ASC} = \rho \mathbf{W} \tilde{\boldsymbol{\beta}}_{g,ASC} + \boldsymbol{\varepsilon}_{g,ASC}, \ (0 < \rho < 1), \quad (2.16)$$

where ρ is the spatial autoregressive coefficient, $\mathbf{W}$ is a distance-based spatial weight matrix with elements $w_{kk'}$ corresponding to regions k and k' (with $w_{kk} = 0$ and $\sum_{k'} w_{kk'} = 1$). With the specification above, and defining $\mathbf{S} = \left[\mathbf{IDEN}_{15} - \rho \mathbf{W}\right]^{-1}$ [15×15 matrix], where $\mathbf{IDEN}_K$ is the identity matrix of size K (K=15 in our case), we may then write:

$$\tilde{\boldsymbol{\beta}}_{g,ASC} = \mathbf{S}\boldsymbol{\varepsilon}_{g,ASC}, \text{and } \tilde{\boldsymbol{\beta}}_{g,ASC} \sim f_{15}(\mathbf{0}, \mathbf{S}\boldsymbol{\Lambda}\mathbf{S}') \quad (2.17)$$

In the above expression, technically, we can allow the distribution of $\tilde{\boldsymbol{\beta}}_{g,ASC}$ to vary across segments g by allowing a general specification for $\boldsymbol{\Lambda}$ that varies across segments (the only

normalization requirement is that the first element for the first segment be 1) and/or by allowing the spatial autoregressive coefficient to vary across segments. However, the first specification leads to proliferation in the number of parameters (especially given the number of alternatives), while the second one is not intuitive because there is no reason for the intensity of spatial correlation in unobserved attributes to vary across segments. Thus, from a pragmatic standpoint, we use the same simple covariance matrix across all segments for the $\tilde{\boldsymbol{\beta}}_{g,ASC}$ vector (as in Equation (2.15)). Doing so also allows a comparison of the magnitude of the mean of coefficients in the baseline preference across segments, as long as there are no substantial differences in the variance elements of the coefficients. A point to note in this discussion is that the expression in Equation (2.17) collapses to that of Equation (2.15) if there is no spatial correlation, as should be the case.

This leaves the specification of the weight matrix **W**. Several weight matrix specifications were considered in our empirical analysis to characterize the nature of the dynamics of the spatial dependence across regions. These included (1) a contiguity specification that generates spatial dependence between the destination region alternatives based on whether or not two regions are contiguous (we considered the Marlborough and Wellington regions as being contiguous because they are the ferry landing points for travel between the two islands), (2) the inverse of a continuous travel time specification where the time between regions is obtained from the skims discussed earlier, and (3) the inverse of the square of the continuous distance specification. In addition, for all the three specifications above, we also examined a specification that confines the spatial correlation to only the regions within each island (with zero spatial correlation between regions in different islands). Overall, the best data fit results were obtained consistently with the inverse of the continuous distance specification, which is the one used in the results discussed in the next section.

2.4.3.2 Satiation and segmentation specification

In our estimations, we considered both a γ-profile as well as an α-profile for introducing satiation. In all cases, the γ-profile provided superior results, so we will only discuss the specification for the γ-profile here. As discussed earlier, the parameter γ_{qgk} may be parameterized as $\exp(\boldsymbol{\theta}_g'\boldsymbol{a}_{qk})$, where $\boldsymbol{a}_{qk}$ is a vector of explanatory variables and $\boldsymbol{\theta}_g$ is a corresponding vector of parameters specific to segment g in the mixture model. It is the

specification of the $\boldsymbol{a}_{qk}$ vector that we discuss here. In addition to a constant, we considered all the other variables discussed in the previous section. We particularly examined the effect of wetland land-cover accessibility on satiation behavior, based on the suggestion from the descriptive statistics that a higher wetland cover percentage leads to higher satiation effects (less trips).

All the variables associated with demographics characteristics were considered for characterizing different discrete segments (see Equation 2.7 earlier). These demographic variables included respondent age, respondent's household income, respondent's household size by number of adults (>18 years of age) and number of children (18 years or less), respondent's household structure (single person, couple, nuclear family, single parent, multi-family household, and non-family household), and respondent gender. Of these, the respondent's household structure provided a very good indication of the travel group, because almost all trips were made with family members in couple, nuclear family, single parent, and multi-family households. Also, in our specifications, we considered respondent gender only for single person, single parent, and non-family households, because the decision in other households is likely to be jointly made (and gender simply provides information on which respondent happened to be picked in the survey in these households, and should not provide any preference information). All the segmentation variables were introduced as alternative-specific variables in the logit link function of Equation (2.7) with the first segment being the base.

2.4.4 Model estimation results

A number of different specifications were explored, with different sets of variables, different functional forms of variables, and different groupings. The final specification was based on having adequate observations in each category of categorical independent variables (such as for household structure), a systematic process of rejecting statistically insignificant effects, combining effects when they made sense and did not degrade fit substantially, and, of course, judgment and insights from earlier studies. To identify the optimal value for the number of latent segments (G), we estimated the model for increasing values of G (G =1,2,3,4,...) until we reached a point where an additional segment did not significantly improve model fit. The evaluation of model fit was based on the Bayesian Information Criterion (BIC):

$$\text{BIC} = -L(\boldsymbol{\theta}) + 0.5 \cdot R \cdot \ln(N).^{11} \quad (2.18)$$

The first term on the right side is the negative of the log-likelihood value at convergence; *R* is the number of parameters estimated and *N* is the number of observations (see Allenby, 1990, Bhat, 1997). As the number of segments, *G*, increases, the BIC value keeps declining till a point is reached where an increase in *G* results in an increase in the BIC value. Estimation is terminated at this point and the number of segments corresponding to the lowest value of BIC is considered the appropriate number for *G*. In our analysis, based on the Bayesian Information Criterion (BIC), the three-segment model was clearly the model with the best performance (the log-likelihood value at convergence for this model was -8,499.78 and, with 46 model parameters, the BIC was 8,687.52; the corresponding values for the model with one segment (that is, no latent segmentation), two segments, and four segments were 8,872.21, 8,711.36, and 8,780.14, respectively.

The estimation results for the three-segment mixture MDCP model are presented in Table 2.6. The first panel corresponds to the probabilistic assignment of individuals to each of the three segments (the first segment is the base segment). The second presents the parameter estimates on the independent variables in the baseline utility specifications of the MDCP model corresponding to each segment. The third provides the parameters in the satiation component. Each of these is discussed in turn in the next three sections.

2.4.4.1 Assignment of individuals to discrete (latent) segments

In the top panel of Table 2.6, the constants in the segmentation model contribute to the size of each segment and do not have any substantive interpretation. The other results in the top panel of Table 2.6 indicate that the second segment, relative to the other two segments, is more likely to consist of individuals with children (that is, the individuals are more likely to belong to nuclear

[11] Many measures have been suggested in the literature to evaluate model fit, especially in the context of the number of segments in latent segmentation models. These include the Akaike Information Criterion (AIC), the BIC, and many variants of both of these (see Fonseca, 2010 for a listing and description of these information criteria). In general, the criteria based on the AIC tend to favor complex models with many segments as the sample size increases, leading to potential overfit. On the other hand, the criteria based on the BIC tend to favor simpler models, with an adjustment for sample size (such as the ln*(N*) appearing in Equation (2.18)), to avoid overfit. More simply speaking, the BIC-based measures demand a higher strength of evidence to add complexity than do the AIC-based measures, and thus the BIC-based measures favor more parsimonious models with fewer segments than do the AIC-based measures (see Neath and Cavanaugh, 2012). In the context of latent segmentation models, where the number of parameters explodes as the number of segments increases, parsimony is a much desired property from an interpretation and simplicity perspective. Thus, most latent segment models adopt the BIC as the model selection criterion, as we also do in the current dissertation.

or single parent households) and low-income individuals. This second segment also is less likely to comprise single person households relative to the first segment. The third segment comprises individuals who tend to be in couple households of middle age (48 years) or older, the least likely to be single person households, and less likely to be in the "lower than NZ$50,000" annual income range relative to the second segment, but more likely to be in this income range relative to the first segment. A more intuitive way to characterize the different segments is to estimate the percentages of individuals in each category of the demographic variables in each segment (see Bhat, 1997 for the formula to do so). The results are presented in Table 2.7. For example, the first numerical value in the table indicates that 60.2% of individuals in the first segment are younger than 48 years, while the corresponding percentages are 61.8% and 35.7% in the second and third segments, respectively. In the overall sample, 46.4% of individuals are younger than 48 years. The figures in Table 2.7 support our previous observations regarding segment characteristics. Based on the relative characterizations of the segments, we will refer to the first segment as the "high-flyer low family commitments" (HFLFC) segment, the second as the "low income parents" (LIP) segment, and the third as the "couple baby-boomer" (CBB) segment (most individuals over 48 years of age in the sample were born between 1943 and 1964 and represent the post-war baby-boom generation of New Zealand). In terms of the relative sizes of the three segments, this can also be estimated in a straightforward way by aggregating the individual segment-level probabilities (Equation 2.7) across all individuals. The sizes are estimated to be 11.4%, 57.9% and 30.7% for the HFLFC, LIP, and CBB segments, indicating a domination of the LIP segment in the population.

Table 2.6. Three segments FDMN-MDCP model estimation results

Variable	First Segment		Second Segment		Third Segment	
	Estimate	t-stat[+]	Estimate	t-stat	Estimate	t-stat
Segment Probabilities						
Alternative specific constant	-	-	1.020	3.50	0.314	2.33
Age: 48 years or older	-	-	-	-	0.880	3.40
Single person household	-	-	-0.501	-3.69	-0.646	-3.50
Couple household	-	-	-	-	0.467	2.70
Nuclear family household	-	-	0.542	4.77	-	-
Single parent household	-	-	0.229	2.61	-	-
Income less than NZ $50,000	-	-	1.250	2.30	0.604	3.12
Baseline utilities						
Logarithm of the area (miles^2) – mean	0.797*	4.72*	0.797*	4.72*	0.797*	4.72*
Ferry (dummy) – mean	0.102	2.40	0.121	3.20	-	-
Travel cost ($/100) –mean	-0.700	-15.22	-0.821	-35.04	-0.780	-4.95
Travel cost ($/100) – standard deviation	0.501	3.00	0.573	2.89	0.442	3.14
Land cover accessibility measure specific to						
Urban (/10^4) –mean	0.431	2.43	0.429	2.09	0.457	2.64
Urban (/10^4) – standard deviation	0.119	2.28	0.100	2.23	0.091	2.17
Forest (/10^4) –mean	0.450	5.09	0.360	6.66	0.210	4.44
Wetland (/10^4) –mean	-4.210	-3.23	-4.195	-5.10	-4.030	-2.69
Agricultural (/10^4) –mean	-0.112	-4.91	-0.498	-9.15	0.212	3.59
Land-cover diversity accessibility index	0.270	2.69	-0.443	-2.32	-0.213	-2.16
Satiation parameters ($\tilde{\theta}_g$ parameters)						
Constant	1.802	27.20	1.789	25.42	1.672	23.11
Land cover accessibility measure specific to						
Wetland (/10^4) –mean	-2.535	-3.56	-2.367	-4.10	-2.055	2.17
Land-cover diversity accessibility index	0.770	2.09	-	-	-0.231	-2.04
Spatial autoregressive coefficient (t-stat)	0.096 (1.56)					
Log-Likelihood at Convergence	-8,499.78					

* The size coefficient (coefficient corresponding to the logarithm of the area in miles^2) is constrained to be equal across all segments. The t-statistic for this coefficient is with respect to the hypothesis that the coefficient is equal to one.

\+ All coefficients are different from zero (or different from one in the case of the size variable) at the 95% confidence level or higher (or a p-value of 0.05 or lower). The 95% confidence level corresponds to an absolute t-statistic value of 1.96.

Table 2.7. Quantitative characterization of the three segments

Segmentation Variable		First Segment	Second Segment	Third Segment	Overall Market
Age	Younger than 48	60.2%	61.8%	35.7%	46.4%
	48 years or older	39.8%	38.2%	64.3%	53.6%
Household structure	Single person	15.4%	13.8%	13.7%	14.0%
	Couple	22.9%	18.3%	41.5%	26.0%
	Nuclear family	45.0%	52.1%	31.6%	45.0%
	Single parent	5.1%	7.4%	4.7%	6.3%
	Multi family or non-family	11.6%	8.4%	8.5%	8.7%
Income	Less than NZ$50,000	22.9%	42.6%	34.5%	37.8%
	NZ$50,000 or more	77.1%	57.4%	65.5%	62.2%

2.4.4.2 Baseline utility parameters

Referring back to Table 2.6, the effect of size (see the second panel) in the baseline utility function is positive and less than one. We specified different size coefficients across the segments, but the coefficients were not statistically different and were constrained to be equal. This was also our theoretical expectation, because we saw no reason that the size coefficient (representing the magnitude of region-specific unobserved factors affecting all elemental opportunities within the region) should vary across segments. The coefficient is statistically different from one, indicating the inelastic effect of size growth on the baseline utility.

The effect of the ferry dummy variable in the baseline utility is positive for the HFLFC and LIP segments, but not significant for the CBB segment. The absence of effect on the CBB segment may be a reflection of the relative lack of families with children in this segment, and the possibly intrinsic and positive "adventure" value of a ferry ride for families with children. The effect of travel cost on baseline utility is, as expected and on average, negative in all the segments. The LIP segment is the most cost-sensitive, followed by the CBB segment, in an

inverse relationship of cost sensitivity to household income earnings of families across the segments. The results also show statistically significant heterogeneity (across individuals) in the responsiveness to cost within each latent segment, as manifested in the standard deviation estimates on the cost coefficient. The normal distribution assumption implies that some individuals do have a positive utility for cost, but the vast majority have a negative cost sensitivity. In particular, the mean and standard deviation estimates indicate that cost has a negative impact for 92% of individuals in the first and second segments, and for 96.5% of individuals in the final segment.

The land cover accessibility measures reinforce the findings from our descriptive analysis. Specifically, regions with high urban land cover "pull" leisure trips with about equal intensity from all three segments, though there is heterogeneity in the magnitude of the "pull" within each segment (as indicated by the statistically significant standard deviations on the urban land cover variable in Table 2.6). Cities clearly offer a much higher density of tourism opportunities from regional events and festivals during the year to gastronomic indulgence opportunities, art galleries, museums, theaters and shopping centers. The effect of forest land-cover on baseline utility is also positive, suggesting a preference for destination regions with high forest land cover. This preference varies across the three discrete segments, with the HFLFC segment having the highest preference for forest-oriented leisure pursuits and the CBB having the lowest. The high preference of the first segment for regions with forest land cover is presumably a reflection of young, single individuals (with relatively little familial commitments) seeking adventurous hiking and bicycling trails through New Zealand's rough and rugged forest terrain. On the other hand, the relatively older CBB segment group may not prefer such physically-intensive leisure pursuits to the extent that their younger counterparts do. Also, there is a clear and generic tendency across all segments to stay away from regions with high wetland land cover. This is not surprising, given that wetlands offer little attraction for tourism and, in New Zealand, are typically associated with negative externalities such as pollution, drainage problems, and presence of invasive plant species (see Peters and Clarkson, 2010). The effect of the agricultural land-cover accessibility varies across segments; while the individuals in the third segment are attracted to agricultural areas, the individuals in the first and second segments tend to avoid agricultural areas. This is perhaps an indication of couple baby-boomers (CBB) being drawn to activities such as visiting vineyards for a relaxed wine-tasting escapade, activities that

may not interest individuals with children (the LIP segment) or may be considered too "docile" by young individuals with little family commitments (the HFLFC segment).

The effects of the land-cover diversity accessibility index on the baseline function indicate that high-flying young individuals prefer regions with a good diversity of activities, while those in the LIP and CBB segments prefer regions with focused activities. Another interpretation is that those in the LIP and CBB segments are inclined to pursue very specific types of leisure activities (such as perhaps park entertainment for the LIP segment and wine tasting trips for CBBs), and then select regions that are heavily invested in opportunities of that specific leisure type.

Finally, the covariance estimate (not shown in Table 2.6) between the travel cost and urban accessibility random coefficients was 0.040 (t-statistic of 2.21), 0.035 (t-statistic of 2.28), and 0.042 (t-statistic of 2.03) for segments one, two and three respectively. This suggests that individuals who are less sensitive (more sensitive) to travel costs also prefer (dislike) urban destination zones. That is, individuals who prefer recreation pursuits based on man-made urban settings (amusement parks or leisure shopping complexes) appear not to mind spending additional time to get to their destinations, while those who prefer natural and pristine settings are the ones who would rather travel to close destinations to pursue their recreational interests.

2.4.4.3 Satiation effects

These effects are presented toward the bottom panel of Table 2.6. As indicated earlier, the satiation parameter is parameterized as $\gamma_{qgk} = \exp(\tilde{\boldsymbol{\theta}}_g' \boldsymbol{a}_{qk})$, and the satiation coefficients in Table 2.6 are the $\tilde{\boldsymbol{\theta}}_g$ parameters for each segment g. A positive parameter on a variable implies that an increase in the variable has the effect of increasing the γ_{qgk} parameter and decreasing satiation (that is, increasing repeat trips of the individual to a destination region), while a negative parameter has the effect of decreasing the γ_{qgk} parameter and increasing satiation (that is, decreasing repeat trips of the same individual to a destination region).

Everything else being equal, the constants indicate that satiation in the context of a destination region sets in fastest for the third CBB segment and slowest for the first HFLFC segment. That is, in general, individuals in the HFLFC segment are more willing to make repeat trips to a destination region than individuals in the LIP segment, and individuals in the LIP

segment are more willing to make repeat trips to the same destination region than individuals in the CBB segment. The wetland land-cover accessibility measure has a negative effect in all segments, *i.e.*, destinations with higher wetland land cover lead to a higher satiation effect (less repeat visits to such regions by the same individual) than destinations regions with a lower wetland land cover. This is not surprising, given the negative characteristics associated with wetland areas in New Zealand, Finally, among the satiation parameters, the effect of the land-cover diversity accessibility index variable indicates that individuals in the first HFLFC segment get less satiated with (willing to make more repeat visits to) destination regions with a high diversity in activity type opportunities (as proxied by land cover percentages), while individuals in the third CBB segment get satiated very quickly with (are unlikely to make repeat visits to) destination regions with a high diversity.

2.4.4.4 Spatial dependence

The spatial autoregressive coefficient, as expected, is positive, of the order of 0.10, and is different from zero at about the 7% level of significance for a one-tailed test.

2.4.4.5 Summary and implications for increasing destination competitiveness

A number of summary observations may be made from the model. First, the presence of a ferry leg appears to increase the attractiveness of a destination region for young single individuals and young parents (individuals in the HFLFC and LIP segments), but has relatively little attractive value for older baby-boomers. Of course, this is after controlling for the total cost of travel, which itself does have a very significant negative impact on destination region choice (especially for the LIP segment). Second, regions with high urban land cover are in general very attractive as a leisure trip destination. This is also true of regions with good forest cover; such regions have the highest attractive value for individuals in the first HFLFC segment and the least attractive value for individuals in the CBB segment. Third, regions with high wetland land cover lowers attractive value across the board, while regions with high agricultural land cover appeal substantially to middle-aged couples (individuals in the CBB segment) but "push away" young individuals in general and young parents in particular, presumably because agricultural lands in New Zealand correspond quite a bit to vineyards. Finally, the combined effects of the land-cover diversity index on the baseline and satiation function, as well as the constant coefficients in the satiation function, imply that individuals in the HFLFC segment place a premium on diversity of

opportunities in terms of the types of activities offered by a destination region, and are much more willing to be loyal to a destination region that offers that diversity (if they make multiple leisure trips). On the other hand, the LIP and CBB segments are much less interested in diversity of activity type opportunities within a destination region, though they also look more for diversity in terms of destination regions visited in general. The individuals in the CBB group in particular are averse to repeat-visiting regions with high diversity of activity opportunities.

The kinds of insights above offered by our proposed model can be valuable in branding and marketing campaigns. As a simple illustration, consider two of the most popular destination regions: Auckland and Nelson. Auckland has a higher diversity in activity opportunities as proxied by land-cover percentages (a diversity index of 0.38) than does Nelson (a diversity index of 0.29) which is heavily invested in forest land cover. Our results suggest that these two regions should use different strategies in their marketing and branding, as we discuss below.

Auckland should emphasize its "diversity uniqueness" when targeting the HFFLC group, perhaps by broadcasting customized media advertisements in high income neighborhoods all over New Zealand and having promotional flyers at bars and clubs where young singles spend quite a bit of time. This will serve Auckland well given that individuals in the HFFLC segment desire diversity and can be very loyal to regions that offer that diversity. While doing so, Auckland should also highlight its forest and urban land cover very specifically, because these will make the region more attractive in the perception map of individuals in the HFFLC group. At the same time, given the LIP and CBB segments are much larger in size, Auckland has to also target these segments appropriately. For the LIP and CBB groups, the strategy would be similar to the HFLFC group in its emphasis on urban and forest-related tourism opportunities. However, unlike promotions targeted at the HFLFC group, the Auckland promotion campaigns toward these two groups would do well not to speak about the diversity of types of activity opportunities, and retain a high intensity of coverage of the urban and forest-related tourism opportunities. For the CBB group, it would behoove Auckland campaigns to play up the vineyards and orchards for wine-tasting and consuming tours (Auckland, in addition to its diversity, has a large percentage of its land area invested in agricultural land-use).

Nelson is mainly invested in forest land-cover, with substantial opportunities for adventurous pursuits in rough and rugged forest terrain. This should be the main focus of promotional campaigns in all three segments as opposed to any diversity campaigns. In the CBB

segment, Nelson can play up its vineyards and wine-tasting tourism outlets. Another important marketing strategy for Nelson is to highlight its geographic proximity to the ferry landing in Picton, which is only a two-hour drive on the Queen Charlotte Drive that also happens to be one of the most picturesque drives in all of New Zealand. When promoting the region to the first HFLFC segment and the second LIP segment, Nelson should play up the ferry crossing experience, given that the ferry experience has a positive influence on destination region choice for the first two segments. Playing up the scenic experience also can temper negative travel time effects in general.

Of course, in addition to targeting appropriate individuals for promoting current destination attributes, each region can also consider enhancing the accessibility to opportunities located within the region. For instance, take the case of Waikato, and consider ways that Waikato can make itself more competitive. But before investing in changing the number and type of offerings, Waikato needs to undertake a cost-benefit analysis including an estimation of the additional tourism share that may be "pulled" to Waikato in response to such an investment. The proposed model can be used to provide information for such a cost-benefit analysis. Specifically, consider the case where Waikato realizes that it is not very much invested in urban activity opportunities, which, based on our model results, is a significant determinant of tourist "pull". The model can then be used to evaluate the increase that may be expected in total tourist trip share to Waikato (including repeat trips) due to a 20% increase in its urban land cover (through additional urban activity opportunities). To do so, for each individual in the sample, we predict the number of trips attracted to Waikato in the base case and in the case of an increased urban land cover in the following steps: (1) for the base case, draw 500 realizations for all the stochastic terms in the utility function of Equation (2.1), (2) predict the number of trips to Waikato for each of the realizations using the prediction method of Pinjari and Bhat (2011), (3) average the predicted trips across the 500 realizations to obtain the individual prediction of the number of trips to Waikato, and (4) for the scenario case, increase the urban land cover percentage by 20%, drawing away an equivalent amount from agricultural land-use, (5) redo steps (1), (2), and (3) using the scenario sample, keeping the same 500 realizations for all the stochastic terms as in the base case. Then, from the individual-level predictions for the base and scenario cases, obtain the total Waikato trips in the two cases by aggregating across all individuals in the sample. Finally, we can obtain a pseudo-elasticity effect by taking the change

in total trips to Waikato between the scenario and base cases as a percentage of the total trips to Waikato in the base case. This percentage turns out to be 16.1% (standard error of 1.7%) from the proposed model. As a point of reference, the corresponding percentage is estimated to be 13.3% (standard error of 1.2%) in the LC-MDCP model and 11.5% (standard error of 1.5%) in the RC-MDCP model.[12] Clearly, there are important differences among the models in the policy predictions, with the LC-MDCP and RC-MDCP models under-predicting the effectiveness of an increase in urban opportunities relative to the proposed FDMN-MDCP model. As we will see next, given that the proposed model fits the data much better than the other two models, the implication is that tourism policies to increase urban opportunities may be inappropriately discarded if the simpler LC-MDCP and RC-MDCP models were to be used.

2.4.5 Data fit comparisons with the LC-MDCP and the RC-MDCP models

The difference in policy sensitivity results between the FDMN-MDCP, LC-MDCP, and RC-MDCP models suggests the need to apply formal statistical tests to determine the structure that is most consistent with the data. In this section, we provide measures of fit for these models. For the RC-MDCP model, as we already indicated in a footnote earlier, we consider both observed and unobserved heterogeneity in the "strawman" specification

The LC-MDCP and the proposed model can be compared using the familiar likelihood ratio test, since the former is a restricted version of the latter with no continuous random heterogeneity in coefficients within each segment. For the test between the RC-MDCP and the proposed model, one can compute the adjusted likelihood ratio index with respect to the log-likelihood at equal shares:

$$\bar{\rho}^2 = 1 - \frac{\mathscr{L}(\hat{\boldsymbol{\theta}}) - M}{\mathscr{L}(c)}, \tag{2.19}$$

where $\mathscr{L}(\hat{\boldsymbol{\theta}})$ is the log-likelihood function at convergence, $\mathscr{L}(c)$ is the log-likelihood for the naïve unsegmented model with only the size measure in the baseline function, only the constant in the satiation function, no spatial dependence, and IID errors across regions as in Equation (2.15), and M is the number of parameters estimated in the model minus two (that is, minus the

[12] To be sure, in this part of the analysis, we did not just consider random coefficients on the variables in the baseline utility function, but also tested demographic variable interactions with the variables to obtain a RC-MDCP model that accommodates systematic heterogeneity (when found statistically significant) in the coefficients. This RC-MDCP specification is a much more appropriate "strawman" to compare with the proposed FDMN-MDCP model than a pure random coefficients specification that does not consider these interactions.

single size coefficient in the baseline utility and the single satiation constant estimated in the naïve unsegmented model). To test the performance of the two non-nested models (*i.e.* the proposed FDMN-MDCP and RC-MDCP models) statistically, the non-nested adjusted likelihood ratio test may be used. This test determines if the adjusted likelihood ratio indices of two non-nested models are significantly different. In particular, if the difference in the indices is $(\bar{\rho}_2^2 - \bar{\rho}_1^2) = \tau$, then the probability that this difference could have occurred by chance is no larger than $\Phi\left\{-[-2\tau\mathscr{L}(c) + (M_2 - M_1)]^{0.5}\right\}$ in the asymptotic limit. A small value of the probability of chance occurrence indicates that the difference is statistically significant and that the model with the higher value of adjusted likelihood ratio index is to be preferred.

The likelihood ratio test (for the comparison of the LC-MDCP and FDMN-MDCP models) and non-nested adjusted likelihood ratio test (for the comparison of the RC-MDCP and FDMN-MDCP models) constitute disaggregate measures of fit that consider performance at the multivariate and disaggregate level of all combinations of regions, While the best data fit measures, these are not very intuitive. So, we also evaluate the performance of the three models intuitively and informally at an aggregate level. However, since there are too many multivariate combinations possible of leisure trip-making to the destination regions and it is impossible to provide fit statistics for all these combinations, we compare the aggregate marginal bivariate predictions (with the true sample values) for combinations of two of the most visited regions – Waikato and Auckland. Specifically, we focus on the percentage of individuals who, during the four-week survey period, visit Waikato but not Auckland, Auckland but not Waikato, both Auckland and Waikato, and neither of the two. The prediction procedure is similar to the one used for undertaking the sensitivity analysis in the previous section, except that, for each individual, we compute the probability of visiting each of the four combinations of regions as the percentage of times in the 500 realizations that each of the combinations has a non-zero number of visits. The probabilities for each combination are added up across individuals to obtain the predicted number of individuals falling into each combination category and compared with the actual percentages using the mean absolute percentage error (MAPE) statistic.

The results of the data fit comparisons are presented in Table 2.8. The first row provides the log-likelihood for the naïve unsegmented model (that is, the $\mathscr{L}(c)$ value), which is, of course, the same across the three models. The second row indicates the superior performance of the proposed FDMN-MDCP model in terms of the convergent log-likelihood value, as does the

adjusted likelihood ratio index in the fifth row (note that the small magnitude of this index is not surprising, given the multitude of different possible multivariate combinations). The sixth row formally shows the likelihood ratio test result of the comparison of the FDMN-MDCP model with the LC-MDCP model, indicating the clear dominance of the FDMN-MDCP data fit. The same result is obtained in the next row through a non-nested adjusted likelihood ratio test comparing the FDMN-MDCP model with the RC-MDCP model; the probability that the adjusted likelihood ratio index difference between these models could have occurred by chance is literally zero. Finally, the last panel of the table first shows the actual percentages of individuals falling in each combination of visiting/not visiting the Waikato and Auckland regions, followed by the predicted percentages from the three different models. The MAPE values from the three models are provided in the last row of the table. The LC-MDCP models has a MAPE value that is about three times that of the FDMN-MDCP, while the RC-MDCP model has a MAPE that is about 3.5 times that of the FDMN-MDCP.

All the fit measures discussed thus far are based on model fit on the overall sample used in estimation. While taken together, these fit measures reveal the superiority of the proposed FDMN-MDCP model, there is still a small possibility that the better performance of our model is simply an artifact of overfitting and may not translate to predictive accuracy in other samples. To accommodate for this, we also evaluated the performance of the three models on various market segments of the estimation sample (such predictive fit tests are sometimes referred to as market segment prediction tests). The intent of using such predictive tests is to examine the performance of different models on sub-samples that do not correspond to the overall sample used in estimation. Effectively, the sub-samples serve a similar role as an out-of-sample for validation. The advantage of using the sub-sample approach rather than an out-of-sample approach to validation is that there is no reduction in the size of the sample for estimation. This is particularly an issue in models of the type estimated in this chapter because of the need to use as much information as possible given the number of parameters to be estimated. If a model shows superior performance in the subsamples in addition to the overall estimation sample, it is indication that the model indeed provides a better data fit.

Table 2.8. Measures of fit

Summary Statistic			Estimation Sample					
			FDMN-MDCP		LC-MDCP		RC-MDCP	
Log-likelihood of the naïve unsegmented model			-15,783.21					
Log-likelihood at convergence			-8,499.78		-8,550.03		-8,648.46	
Number of parameters			49		40		15	
Number of observations			3,508					
Adjusted likelihood ratio index			0.458		0.455		0.451	
Predictive likelihood ratio test between FDMN-MDCP and LC-MDCP models			Test statistic [-2*($LL_{LC\text{-}MDCP}$-$LL_{FDMN\text{-}MDCP}$)]=102 > Chi-Squared statistics with 9 degrees of freedom at any reasonable level of significance					
Non-nested adjusted likelihood ratio test between the FDMN-MDCP and RC-MDCP models			$\Phi[-16.22] << 0.0001$					
Percentage of individuals (trips) predicted to visit….	Actual percentage		Predicted percentage					
	Individuals	Trips	Individuals	Trips	Individuals	Trips	Individuals	Trips
Waikato but not Auckland	16.9	17.9	17.6	18.4	20.0	20.5	21.4	22.3
Auckland but not Waikato	10.8	8.1	12.3	9.5	14.0	11.8	14.5	12.4
Both Auckland and Waikato	5.6	6.4	6.7	7.7	8.6	9.4	9.3	9.7
Neither Auckland nor Waikato	66.7	67.6	63.4	64.4	57.4	58.3	54.8	55.6
Mean Absolute Percentage Error			10.7%	11.3%	28.9%	30.2%	36.2%	36.7%

To do so, we computed the mean absolute percentage error (MAPE) for the percentage of individuals predicted to visit the same four combinations of the two destinations as in Table 2.8 and for three segmentations of demographic variables: (1) income less than NZ$50,000 and income greater than NZ$50,000, (2) nuclear and non-nuclear households, and (3) age less than 48 years and age more than 48 years. The overall MAPE values for percentage of individuals predicted to visit the four destination combinations in the two income segments were 10.4% and 10.8% from the FDMN-MDCP model, 28.7% and 28.8% from the LC-MDCP model, and 36.1% and 36.3% from the RC-MDCP model. The corresponding values for the household structure segmentation were 10.7% and 10.6% from the FDMN-MDCP model, 29.0% and 28.7% from the LC-MDCP model, and 36.2% and 35.3% from the RC-MDCP model, and for the age segmentation were 9.8% and 9.5% from the FDMN-MDCP model, 28.3% and 27.9% from the LC-MDCP model, and 36.0% and 35.1% from the RC-MDCP model. All in all, the FDMN-MDCP model clearly outperforms the other two models even in such a predictive exercise.

2.5 CONCLUSIONS

This chapter has proposed a new econometric formulation and a complete blueprint of an associated estimation method for a finite discrete mixture of normals version of the multiple discrete-continuous probit (or FDMN-MDCP) model. The model allows consumers to choose multiple alternatives at the same time, along with the continuous dimension of the amount of consumption, and captures heterogeneity in the response coefficients of the baseline utility function. This is a very general way of including heterogeneity in the sensitivity to exogenous variables in the multiple discrete-continuous context, with the normally distributed random parameters approach and the latent class approach constituting special cases.

A simulation exercise is undertaken to evaluate the ability of the proposed approach to recover parameters from simulated datasets. The results from the experiments show that the proposed inference approach, which is computationally fast and straightforward to implement, does very well in recovering the true parameters used in the data generation. Also, the simulation results show that ignoring the continuous component of the mixing (as reflected in the LC-MDCP model) or ignoring the discrete component of the mixing (as in the RC-MDCP model) when the true data is generated using an FDMN MDCP structure leads to substantial parameter bias. The average absolute percentage bias (APB) for the LC-MDCP model is about 28.5%, and for the RC-MDCP model is 26%, relative to the APB for the correct FDMN-MDCP model which

is of the order of 3%. Clearly, the repercussion of imposing incorrect restrictions is very severe on parameter bias.

This chapter demonstrates the application of the proposed approach through a study of individuals' recreational (*i.e.*, long distance leisure trips of over 25 miles one-way) choice among alternative destination locations and the number of trips to each recreational destination location, using data drawn from the 2012 New Zealand Domestic Travel Survey (DTS). The Bayesian Information Criterion indicates that the preferred specification is a three-segment solution, with one segment loading on high flying low family commitment (HFLFC) individuals, the second on low income parents (LIP), and the third on couple baby-boomers (CBB). In a comparative empirical assessment of the FDMN-MDCP with the simpler LC-MDCP and RC-MDCP models, the FDMN-MDCP came out clearly as the winner in terms of data fit.

The results of the preferred three-segment solution showed heterogeneity (in the form of a continuous normal distribution) in sensitivity to cost and urban land cover within each latent segment, and differences (across the three latent segments) in the response to the presence of a ferry ride, travel cost, land cover accessibility measures, and the land cover diversity accessibility index. These differences, in combination with the socio-demographic characteristics of individuals in each segment, provide important information for effective targeting and strategic positioning to increase destination competitiveness. More generally, the FDMN-MDCP formulation appears to be a valuable methodology for marketing and positioning in markets that are characterized by multiple discreteness. Future research should focus on applying the FDMN-MDCP formulation to other multiple discrete contexts. Also, while the application to recreational destination choice demonstrates the value of the formulation, future work should consider a much richer set of destination region attributes.

CHAPTER 3: Multivariate Skew-Normal Distribution for Unobserved Heterogeneity in the Spatial MDC Model

This chapter proposes a new spatial MDC model with skew-normal kernel error terms and skew-normal distributed random response coefficients. To our knowledge, this is the first time a flexible and parametric skew-normal distribution for the kernel error term and/or random response coefficients has been used in both spatial- and aspatial-MDC models. The next section provides an overview of the multivariate skew-normal distribution (MVSN) and the properties of the distribution that are most helpful in the context of spatial MDC models. The third section presents the modeling methodology of the aspatial and spatial skew normal MDC models, along with the proposed estimation method. The fourth section offers an overview of the simulation exercises and the fifth section presents the empirical application (including a data description and model estimation results). The sixth and final section provides a discussion of the main findings together with concluding thoughts.

3.1 THE MULTIVARIATE SKEW-NORMAL DISTRIBUTION

In this section, we provide an overview of the multivariate skew-normal distribution, and briefly present the properties of the distribution that are most relevant in the context of application for MDC models. Most of the notation used in this section is extracted from Bhat and Sidharthan (2012).

In this study, we use the MVSN version originally proposed by Azzalini and Dalla Valle (1996).[13] The MVSN version used here is efficient in the number of additional parameters to be estimated and is closed under any affine transformation of the skew-normally distributed vector. At the same time, the cumulative distribution function of an L-variate skew normally distributed variable of the Azzalini and Dalla Valle type requires only the evaluation of an $(L+1)$-dimensional multivariate cumulative normal distribution function.

Consider an MVSN distributed random variable vector $\boldsymbol{\eta} = (\eta_1, \eta_2, \eta_3, \ldots, \eta_L)'$ with an $(L \times 1)$-location parameter vector $\mathbf{0}_L$ (that is, an $(L \times 1)$ vector with all elements being zero) and

[13] This version is also referred to by Lee and McLachlan, 2013 as the restricted multivariate skew normal distribution

an $(L \times L)$-symmetric positive-definite correlation matrix $\boldsymbol{\Omega}^*$. Then, the MVSN distribution for $\boldsymbol{\eta}$ implies that $\boldsymbol{\eta}$ is obtained through a latent conditioning mechanism on an $(L+1)$-variate normally distributed vector $(C_0^*, \boldsymbol{C}_1^{*\prime})'$, where C_0^* is a latent (1×1)-vector and $\boldsymbol{C}_1^{*\prime}$ is an $(L\times 1)$-vector:

$$\begin{pmatrix} C_o^* \\ \boldsymbol{C}_1^{*\prime} \end{pmatrix} \sim MVN_{L+1}\left(\begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix}, \boldsymbol{\Omega}_+^* \right), \text{ where } \boldsymbol{\Omega}_+^* = \begin{pmatrix} 1 & \boldsymbol{\rho}' \\ \boldsymbol{\rho} & \boldsymbol{\Omega}^* \end{pmatrix}. \tag{3.1}$$

$\boldsymbol{\rho}$ is an $(L\times 1)$-vector, each of whose elements may lie between –1 and +1. The matrix $\boldsymbol{\Omega}_+^*$ is also a positive-definite correlation matrix. Then, $\boldsymbol{\eta} = \boldsymbol{C}_1^{*\prime} \mid (C_0^* > 0)$ has the standard multivariate skew-normal (SMVSN) density function shown below:

$$\tilde{\varphi}_L(\boldsymbol{\eta} = \mathbf{z};\ \boldsymbol{\Omega}_+^*) = 2\varphi_L(\mathbf{z};\ \boldsymbol{\Omega}^*)\,\Phi(\boldsymbol{\alpha}'\mathbf{z}), \text{ where } \boldsymbol{\alpha} = \frac{(\boldsymbol{\Omega}^*)^{-1}\boldsymbol{\rho}}{\left(1-\boldsymbol{\rho}'(\boldsymbol{\Omega}^*)^{-1}\boldsymbol{\rho}\right)^{1/2}}, \tag{3.2}$$

where $\phi_L(.)$ and $\Phi(.)$ represent the standard multivariate normal density function of L dimensions and the standard univariate cumulative distribution function, respectively. We write $\boldsymbol{\eta} \sim \text{SMVSN}(\boldsymbol{\Omega}_+^*)$. To obtain the density function of the non-standardized multivariate skew-normal distribution, consider the distribution of $\mathbf{Y} = \boldsymbol{\zeta} + \boldsymbol{\omega}\boldsymbol{\eta}$. This MVSN distribution for $\mathbf{Y}$ implies that $\mathbf{Y}$ is obtained through a latent conditioning mechanism on an $(L+1)$-variate normally distributed vector $(C_0, \boldsymbol{C}_1')'$, where C_0 is a latent (1×1)-vector and $\boldsymbol{C}_1$ is an $(L\times 1)$-vector:

$$\begin{pmatrix} C_0 \\ \boldsymbol{C}_1 \end{pmatrix} \sim MVN_{L+1}\left(\begin{pmatrix} 0 \\ \boldsymbol{\zeta} \end{pmatrix}, \boldsymbol{\Omega}_+ \right), \text{ where } \boldsymbol{\Omega}_+ = \begin{pmatrix} 1 & \boldsymbol{\sigma}' \\ \boldsymbol{\sigma} & \boldsymbol{\Omega} \end{pmatrix}, \ \boldsymbol{\sigma} = \boldsymbol{\omega}\boldsymbol{\rho}, \text{ and } \boldsymbol{\Omega} = \boldsymbol{\omega}\boldsymbol{\Omega}^*\boldsymbol{\omega}. \tag{3.3}$$

Specifically, we write $\mathbf{Y} \sim \text{MVSN}(\boldsymbol{\zeta}, \boldsymbol{\omega}, \boldsymbol{\Omega}_+^*)$, and the conditioning-type stochastic representation of $\mathbf{Y}$ is obtained as $\mathbf{Y} = \boldsymbol{C}_1 \mid (C_0 > 0)$. The probability density function of the random variable $\mathbf{Y}$ may be written in terms of the SMVSN density function above as (see Bhat and Sidharthan, 2012):

$$f_L(\mathbf{Y}=\mathbf{y};\boldsymbol{\zeta},\boldsymbol{\omega},\boldsymbol{\Omega}_+^*)=\left(\prod_{j=1}^{L}\omega_j\right)^{-1}\tilde{\phi}_L(\mathbf{z};\boldsymbol{\Omega}_+^*),\ \text{where}\ \mathbf{z}=\boldsymbol{\omega}^{-1}(\mathbf{y}-\boldsymbol{\zeta}),\tag{3.4}$$

and ω_j is the j^{th} diagonal element of the matrix $\boldsymbol{\omega}$.

The cumulative distribution function for $\boldsymbol{\eta}$ may be obtained as:

$$P(\boldsymbol{\eta}<\mathbf{z})=\tilde{\Phi}_L(\mathbf{z};\boldsymbol{\Omega}_+^*)=2\Phi_{L+1}(\mathbf{0},\mathbf{z},\boldsymbol{\Omega}_-^*);\ \boldsymbol{\Omega}_-^*=\begin{pmatrix}1 & -\boldsymbol{\rho}'\\ -\boldsymbol{\rho} & \boldsymbol{\Omega}^*\end{pmatrix}.\tag{3.5}$$

The corresponding cumulative distribution function for $\mathbf{Y}$ is:

$$P(\mathbf{Y}<\mathbf{y})=\tilde{\Phi}_L\left(\boldsymbol{\omega}^{-1}(\mathbf{y}-\boldsymbol{\zeta});\boldsymbol{\Omega}_+^*\right)=2\Phi_{L+1}\left(\mathbf{0},\boldsymbol{\omega}^{-1}(\mathbf{y}-\boldsymbol{\zeta}),\boldsymbol{\Omega}_-^*\right).\tag{3.6}$$

It is important to notice that the notation $\mathbf{Y}\sim\text{MVSN}(\boldsymbol{\zeta},\boldsymbol{\omega},\boldsymbol{\Omega}_+^*)$ is convenient because all the components needed to express the cumulative distribution and density functions are distinguishable as the function arguments. Alternatively, we could also write $\mathbf{Y}\sim\text{MVSN}(\boldsymbol{\zeta},\boldsymbol{\rho},\boldsymbol{\Omega})$, where $\boldsymbol{\Omega}$ is the $(D\times D)$-symmetric positive-definite covariance matrix. Let $\boldsymbol{\omega}$ be a $(D\times D)$-diagonal matrix formed by the standard deviations of $\boldsymbol{\Omega}$ (ω_j is the jth diagonal element of the matrix $\boldsymbol{\omega}$). Then, we can write: $\boldsymbol{\Omega}^*=\boldsymbol{\omega}^{-1}\boldsymbol{\Omega}\boldsymbol{\omega}^{-1}$. Finally, we can construct $\boldsymbol{\Omega}_+^*$ as $\begin{pmatrix}1 & \boldsymbol{\rho}'\\ \boldsymbol{\rho} & \boldsymbol{\Omega}^*\end{pmatrix}$.

A couple of properties of the MVSN distribution are provided next. The proof for the first property is available in Arellano-Valle and Azzalini (2006) and Bhat and Sidharthan (2012). The proof for the second property is based on the marginal and conditional distribution properties of the multivariate normal distribution. Both properties will be useful in the development of the spatial skew-normal MDC model.

Property 1: The affine transformation of the MVSN distributed vector $\mathbf{Y}$ (dimension $L\times 1$) [$\mathbf{Y}\sim\text{MVSN}(\boldsymbol{\zeta},\boldsymbol{\omega},\boldsymbol{\Omega}_+^*)$] as $\mathbf{a}+\mathbf{B}\mathbf{Y}$, where $\mathbf{B}$ is a $(h\times L)$ matrix, is also an MVSN distributed vector of dimension $h\times 1$:

$$\mathbf{a}+\mathbf{B}\mathbf{Y}\sim\text{MVSN}(\mathbf{a}+\mathbf{B}\,\boldsymbol{\zeta},\tilde{\boldsymbol{\omega}},\tilde{\boldsymbol{\Omega}}_+^*)],\ \text{where}\ \tilde{\boldsymbol{\Omega}}_+^*=\begin{pmatrix}\mathbf{1} & \tilde{\boldsymbol{\rho}}'\\ \tilde{\boldsymbol{\rho}} & \tilde{\boldsymbol{\Omega}}^*\end{pmatrix},\ \tilde{\boldsymbol{\Omega}}^*=(\tilde{\boldsymbol{\omega}})^{-1}\tilde{\boldsymbol{\Omega}}(\tilde{\boldsymbol{\omega}})^{-1},\tilde{\boldsymbol{\Omega}}=\mathbf{B}\boldsymbol{\Omega}\mathbf{B}',$$

$\widetilde{\boldsymbol{\rho}} = (\widetilde{\boldsymbol{\omega}})^{-1}\mathbf{B}\boldsymbol{\omega}\boldsymbol{\rho}$, and $\widetilde{\boldsymbol{\omega}}$ is the diagonal matrix of standard deviations of $\widetilde{\boldsymbol{\Omega}}$.

Property 2:

If $[\mathbf{G} \sim \text{MVSN}(\boldsymbol{\xi}, \boldsymbol{\omega}, \boldsymbol{\Omega}^*_+)]$ is partitioned into two-subvectors $\mathbf{G}_1$ and $\mathbf{G}_2$, with corresponding partitions of the other vectors as follows:

$$\boldsymbol{\xi} = \begin{pmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \end{pmatrix},\ \boldsymbol{\omega} = \begin{pmatrix} \boldsymbol{\omega}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\omega}_2 \end{pmatrix},\ \text{and}\ \boldsymbol{\Omega}^*_+ = \begin{pmatrix} 1 & \boldsymbol{\rho}_1' & \boldsymbol{\rho}_2' \\ \boldsymbol{\rho}_1 & \boldsymbol{\Omega}^*_1 & \boldsymbol{\Omega}^*_{12} \\ \boldsymbol{\rho}_2 & \boldsymbol{\Omega}^*_{12} & \boldsymbol{\Omega}^*_2 \end{pmatrix},\ \text{then}$$

$$\mathbf{G}_1 \sim \text{MVSN}(\boldsymbol{\xi}_1, \boldsymbol{\omega}_1, \boldsymbol{\Omega}^*_{1+})], \boldsymbol{\Omega}^*_{1+} = \begin{pmatrix} 1 & \boldsymbol{\rho}_1' \\ \boldsymbol{\rho}_1 & \boldsymbol{\Omega}^*_1 \end{pmatrix},\ \text{and}\ \mathbf{G}_2 \sim \text{MVSN}(\boldsymbol{\xi}_2, \boldsymbol{\omega}_2, \boldsymbol{\Omega}^*_{2+})], \boldsymbol{\Omega}^*_{2+} = \begin{pmatrix} 1 & \boldsymbol{\rho}_2' \\ \boldsymbol{\rho}_2 & \boldsymbol{\Omega}^*_2 \end{pmatrix}.$$

That is, the rMSN distribution is closed under marginalization, which can be shown by straightforward integration (Azzalini, 2005).

3.2 MODELING FRAMEWORK

3.2.1 The aspatial Skew Normal MDC (or ASN-MDC) model

Let $q\,(q = 1, 2, ..., Q)$ be the index for individuals (or observation units in general) and let $k\,(k = 1, 2, ..., K)$ be the index for the alternatives. Following Bhat (2008), consider a vector $\boldsymbol{x}_q$ of dimension $K \times 1$ with elements x_{qk} $(x_{qk} > 0 \forall\, q)$, where x_{qk} is a specific consumption of good *k* by individual *q*. Consider the following utility-maximizing function subject to the binding budget constraint:

$$\max U_q(\boldsymbol{x}_q) = \left[\sum_{k=1}^{K-1} \gamma_k \psi_{qk} \ln\left(\frac{x_{qk}}{\gamma_k} + 1\right)\right] + \psi_{qK} \ln x_{qK} \tag{3.7}$$

$$s.t. \sum_{k=1}^{K} p_{qk} x_{qk} = E_q\,,$$

where the utility function $U_q(\boldsymbol{x}_q)$ is quasi-concave, increasing and continuously differentiable, $\boldsymbol{x}_q \geq 0$ is the consumption quantity (vector of dimension *K*×1 with elements x_{qk}), and γ_{qk} and ψ_{qk} are parameters associated with good *k* and consumer *q*. The constraint in Equation (3.7) is

the linear budget constraint, where E_q is the total expenditure (or income) of consumer q, and p_{qk} is the unit price of good k as experienced by consumer q. The utility function form in Equation (3.7) assumes that there is one essential outside good (all the individuals consume the K^{th} good), so that corner solutions (*i.e.*, zero consumptions) are allowed for all goods $k \neq K$. The parameter γ_{qk} in Equation (3.7) serves the role of a satiation parameter. ψ_{qk} represents the stochastic baseline marginal utility (the marginal utility at the point of zero consumption).

To complete the model structure, the baseline utility ψ_{qk}, which has to be non-negative, is parameterized as follows for each alternative:

$$\psi_{qk} = \exp(\tilde{\mathbf{z}}_{\mathbf{qk}}) = \exp(\boldsymbol{\beta}_{\mathbf{q}}' \tilde{\mathbf{z}}_{\mathbf{qk}}) \text{ or } \overline{\psi}_{qk}^{*} = \ln(\psi_{qk}) = \boldsymbol{\beta}_{\mathbf{q}}' \tilde{\mathbf{z}}_{\mathbf{qk}}, \tag{3.8}$$

where $\tilde{\mathbf{z}}_{\mathbf{qk}}$ is a D-dimensional vector of attributes that characterizes good k and individual q (including a dummy variable for each alternative except the last outside alternative, to capture intrinsic preferences for each alternative relative to the last alternative), and $\boldsymbol{\beta}_q$ is a individual-specific vector of coefficients (of dimension $D \times 1$), In order to allow heterogeneity in responsiveness to exogenous variables across individuals, $\boldsymbol{\beta}_q$ is assumed to be multivariate skew-normally distributed: $\boldsymbol{\beta}_q \sim \text{MVSN}(\boldsymbol{b}, \boldsymbol{\rho}, \boldsymbol{\Omega})$, where $\boldsymbol{\rho}$ is the vector of skew parameters and $\boldsymbol{\Omega}$ is the covariance matrix of size (D×D). As we discussed in Section 3.1, we alternatively can write $\boldsymbol{\beta}_q \sim \text{MVSN}(\boldsymbol{b}, \boldsymbol{\omega}, \boldsymbol{\Omega}_{+}^{*})$, where $\boldsymbol{\omega}$ is the diagonal matrix formed by the standard deviations of $\boldsymbol{\Omega}$, $\boldsymbol{\Omega}_{+}^{*} = \begin{pmatrix} 1 & \boldsymbol{\rho}' \\ \boldsymbol{\rho} & \boldsymbol{\Omega}^{*} \end{pmatrix}$, and $\boldsymbol{\Omega}^{*}$ is the correlation matrix associated to $\boldsymbol{\Omega}$. It is not necessary that all elements of $\boldsymbol{\beta}_q$ be random; that is, the analyst may specify fixed coefficients on some exogenous variables in the model, though it will be convenient in presentation to assume that all elements of $\boldsymbol{\beta}_q$ are random. For future reference, we also write $\boldsymbol{\beta}_q = \boldsymbol{b} + \tilde{\boldsymbol{\beta}}_q$, where $\tilde{\boldsymbol{\beta}}_q \sim S\text{MVSN}(\boldsymbol{\omega}, \boldsymbol{\Omega}_{+}^{*})$. Note that the randomness of the parameters (in the $\boldsymbol{\beta}_q$ vector) on the dummy variables specific to each alternative (except the last) represents the kernel error term and captures the idiosyncratic (unobserved) characteristics that impact the baseline utility of good k and individual q.

There is one important identification issue that still needs to be dealt with. We need to ensure the positive definiteness of the matrix $\boldsymbol{\Omega}^*_+$ (note that the positive definiteness of $\boldsymbol{\Omega}^*_+$ ensures the positive definiteness of $\boldsymbol{\Omega}^*$ and therefore $\boldsymbol{\Omega}$; this holds because of the property that any principal square sub-matrix of a positive definite matrix is also positive definite). To guarantee the positive definiteness of the correlation matrix $\boldsymbol{\Omega}^*_+$, we use the approach of Bhat and Srinivasan (2005). Specifically, let $\mathbf{L}$ be the Cholesky decomposition matrix for $\boldsymbol{\Omega}^*_+$. We need to guarantee that the parameters embedded within $\mathbf{L}$ are such that $\boldsymbol{\Omega}^*_+$ is a correlation matrix. This is done by parameterizing the diagonal terms of $\mathbf{L}$ as follows:

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ l_{21} & \sqrt{1-l_{21}^2} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ l_{D+1,1} & l_{D+1,2} & l_{D+1,3} & \cdots & \sqrt{1-l_{D+1,1}^2-l_{D+1,2}^2\cdots-l_{D+1,D}^2} \end{bmatrix} \tag{3.9}$$

In the estimation, the Cholesky elements in the matrix $\mathbf{L}$ are estimated, guaranteeing that $\boldsymbol{\Omega}^*_+$ is indeed a correlation matrix.

As in the multinomial probit model, only differences in the logarithm of the baseline utilities matter, not the actual logarithm of the baseline utility values (see Bhat, 2008). Thus, it will be easier to work with the logarithm of the baseline utilities of the first $K-1$ alternatives, and normalize the logarithm of the baseline utility for the last alternative to zero. That is, we write:

$$\begin{aligned} \overline{\psi}_{qk} &= \overline{\psi}^*_{qk} - \overline{\psi}^*_{qK} = \boldsymbol{\beta}'_{\boldsymbol{q}}(\widetilde{\boldsymbol{z}}_{\boldsymbol{qk}} - \widetilde{\boldsymbol{z}}_{\boldsymbol{qK}}) \\ &= \boldsymbol{\beta}'_{\boldsymbol{q}}\boldsymbol{z}_{\boldsymbol{qk}}, \ \ \boldsymbol{z}_{\boldsymbol{qk}} = \widetilde{\boldsymbol{z}}_{\boldsymbol{qk}} - \widetilde{\boldsymbol{z}}_{\boldsymbol{qK}} \ \ \forall\, k \neq K \\ \text{and} \quad \overline{\psi}_{qK} &= \overline{\psi}^*_{qK} - \overline{\psi}^*_{qK} = 0 \ . \end{aligned} \tag{3.10}$$

The optimal consumption vector $\boldsymbol{x}_q$ can be solved based on the constrained optimization problem of Equation (3.7) by forming the Lagrangian function and applying the KKT conditions. The Lagrangian function for the problem (substituting $\psi_{qk} = \exp(\overline{\psi}_{qk})$ in Equation (3.7)) is provided in Equation (3.11):

$$L_q = \sum_{k=1}^{K-1} \gamma_k \exp(\boldsymbol{b}'\boldsymbol{z}_{qk} + \widetilde{\boldsymbol{\beta}}'_q \boldsymbol{z}_{qk}) \ln\left(\frac{x_{qk}}{\gamma_k}+1\right) + \ln(x_{qK}) - \lambda_q\left[\sum_{k=1}^{K} x_{qk} - E_q\right], \tag{3.11}$$

where λ_q is the Lagrangian multiplier associated with the expenditure constraint. The KKT first-order conditions for the optimal consumption x_{qk}^* are as shown in Equation (3.12):

$$\exp(\boldsymbol{b}'\mathbf{z}_{qk} + \tilde{\boldsymbol{\beta}}_q'\mathbf{z}_{qk})\left(\frac{x_{qk}^*}{\gamma_k}+1\right)^{-1} - \lambda_q = 0 \text{ , if } x_{qk}^* > 0,\ k = 1, 2, ..., K-1 \tag{3.12}$$

$$\exp(\boldsymbol{b}'\mathbf{z}_{qk} + \tilde{\boldsymbol{\beta}}_q'\mathbf{z}_{qk})\left(\frac{x_{qk}^*}{\gamma_k}+1\right)^{-1} - \lambda_q < 0 \text{ , if } x_{qk}^* = 0,\ k = 1, 2, ..., K-1.$$

Substituting $\lambda_q = \left(x_{qK}^*\right)^{-1}$ into the above Equation, and taking logarithms, we can rewrite the KKT conditions as:

$$y_{qk}^* = (V_{qk} - V_{qK}) + \tilde{\boldsymbol{\beta}}_{\boldsymbol{q}}'\mathbf{z}_{\boldsymbol{qk}} = 0, \text{ if } x_{qk}^* > 0,\ k = 1, 2, ..., K-1 \tag{3.13}$$

$$y_{qk}^* = (V_{qk} - V_{qK}) + \tilde{\boldsymbol{\beta}}_{\boldsymbol{q}}'\mathbf{z}_{\boldsymbol{qk}} < 0, \text{ if } x_{qk}^* = 0,\ k = 1, 2, ..., K-1,$$

where $V_{qk} = \boldsymbol{b}'\mathbf{z}_{\boldsymbol{qk}} - \ln\left(\frac{x_{qk}^*}{\gamma_k}+1\right)$ for $k = 1, 2, ..., K-1$, and $V_{qK} = -\ln\left(x_{qK}^*\right)$.

3.2.2 The Spatial Skew Normal MDC (or SSN-MDC) model

The derivation thus far is based on the assumption that individuals are spatially independent. In this section, we include spatial dependency between observations. We begin the formulation of the spatial model from Equation (3.10), and write the logarithm of the baseline utilities (taken as the difference from the logarithm of the baseline utility of the last alternative) for the alternatives as follows:

$$\begin{aligned} \overline{\psi}_{qk} &= \boldsymbol{\beta}_{\boldsymbol{q}}'\mathbf{z}_{\boldsymbol{qk}} + \delta_k \sum_{q'} w_{qq'}\,\overline{\psi}_{q'k},\ \text{for } k = 1, 2, ..., K-1 \\ \overline{\psi}_{qK} &= 0 \text{ for } k = K. \end{aligned} \tag{3.14}$$

where $w_{qq'}$ is a distance-based spatial weight corresponding to individuals q and q' (with $w_{qq} = 0$ and $\sum_{q'} w_{qq'} = 1$) for every q, and δ_k $(0 < \delta_k < 1)$ is the spatial lag autoregressive parameter specific to good k $(k = 1, 2, ..., K-1)$. This formulation takes the typical spatial lag specification used extensively in spatial econometrics, and causes the logarithm of the baseline utilities to be spatially interdependent across individuals based on the spatial proximity (for

another example of an MDC model with a spatial lag formulation see Bhat *et al.*, 2015). The weight $w_{qq'}$ are assumed to converge to zero as the spatial distance between individuals q and q' tends to infinity.

We now set out additional notation to write the baseline utility in a compact form. Define the following:

$\overline{\boldsymbol{\psi}}_q = (\overline{\psi}_{q1}, \overline{\psi}_{q2}, ..., \overline{\psi}_{q,K-1})'$, $[(K-1)\times 1 \text{ vector}]$

$\overline{\boldsymbol{\psi}} = (\overline{\boldsymbol{\psi}}_1', \overline{\boldsymbol{\psi}}_2', ..., \overline{\boldsymbol{\psi}}_Q')'$, $[Q(K-1)\times 1 \text{ vector}]$

$\boldsymbol{z}_q = (\boldsymbol{z}_{q1}, \boldsymbol{z}_{q2}, ..., \boldsymbol{z}_{q,K-1})'$ $[(K-1)\times D \text{ matrix}]$, $\mathbf{z} = (\boldsymbol{z}_1', \boldsymbol{z}_2', ..., \boldsymbol{z}_Q')'$ $[Q(K-1)\times D \text{ matrix}]$, and

$\tilde{\boldsymbol{\beta}} = (\tilde{\boldsymbol{\beta}}_1', \tilde{\boldsymbol{\beta}}_2', .., \tilde{\boldsymbol{\beta}}_Q')'$ $(QD\times 1 \text{ vector})$.

Let $\mathbf{IDEN}_K$ be the identity matrix of size K. Also, define the following matrices:

$$\ddot{\mathbf{z}} = \begin{bmatrix} \boldsymbol{z}_1 & 0 & 0 & \dots & 0 \\ 0 & \boldsymbol{z}_2 & 0 & \dots & 0 \\ 0 & 0 & \boldsymbol{z}_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \boldsymbol{z}_Q \end{bmatrix} \quad [Q(K-1)\times QD \text{ matrix}], \text{ and} \tag{3.15}$$

$$\boldsymbol{\delta} = \begin{bmatrix} \delta_1 & 0 & 0 & \dots & 0 \\ 0 & \delta_2 & 0 & \dots & 0 \\ 0 & 0 & \delta_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \delta_{K-1} \end{bmatrix} \quad [(K-1)\times(K-1) \text{ matrix}]. \tag{3.16}$$

Let $\tilde{\boldsymbol{W}}$ be the $(Q\times Q)$ weight matrix with weight $w_{qq'}$ as its elements, and let $\boldsymbol{1}_{QQ}$ be a $(Q\times Q)$ matrix with each element taking the value of one. Next, define $\mathbf{W} = (\boldsymbol{1}_{QQ} \otimes \boldsymbol{\delta}).*(\tilde{\boldsymbol{W}} \otimes \boldsymbol{IDEN}_{K-1})$, where "$\otimes$" is the kronecker product and "$.*$" stands for the element-by-element multiplication of two matrices. Let $\mathbf{S} = \left[\mathbf{IDEN}_{Q(K-1)} - \mathbf{W}\right]^{-1}$ $[Q(K-1)\times Q(K-1) \text{ matrix}]$. Then, we can write Equation (3.14) for all goods k $(k = 1, 2, ..., K-1)$ and all individuals $q = 1, 2, ..., Q$ in matrix notation as:

$$\overline{\boldsymbol{\psi}} = \mathbf{S}\left(\mathbf{z}\boldsymbol{b} + \ddot{\mathbf{z}}\tilde{\boldsymbol{\beta}}\right) = \mathbf{S}\mathbf{z}\boldsymbol{b} + \mathbf{S}\ddot{\mathbf{z}}\tilde{\boldsymbol{\beta}} \quad [Q(K-1)\times 1 \text{ vector}]. \tag{3.17}$$

Let $[.]_e$ indicate the e^{th} element of the column vector $[.]$, and let $d_{qk} = (q-1)K + k$. Equation (3.17) can be equivalently written as:

$$\overline{\psi}_{qk} = [\mathbf{Szb}]_{d_{qk}} + [\mathbf{S\tilde{z}\tilde{\beta}}]_{d_{qk}},\ k = 1, 2, ..., K-1. \tag{3.18}$$

Using the same approach as for the aspatial case, the KKT conditions take the same form for y^*_{qk} as in Equation (3.13):

$$y^*_{qk} = (V_{qk} - V_{qK}) + [\mathbf{S\tilde{z}\tilde{\beta}}]_{d_{qk}} = 0, \text{ if } x^*_{qk} > 0,\ k = 1, 2, ..., K-1 \tag{3.19}$$

$$y^*_{qk} = (V_{qk} - V_{qK}) + [\mathbf{S\tilde{z}\tilde{\beta}}]_{d_{qk}} < 0, \text{ if } x^*_{qk} = 0,\ k = 1, 2, ..., K-1,$$

where $V_{qk} = [\mathbf{Szb}]_{d_{qk}} - \ln\left(\frac{x^*_{qk}}{\gamma_k} + 1\right)$ for $k = 1, 2, ..., K-1$, and $V_{qK} = -\ln(x^*_{qK})$.

Now, stack the elements $y^*_{qk}\ (k = 1, 2, ..., K-1)$ in the following order:

$$\mathbf{y}^*_q = (y_{q1}, y_{q2}, ..., y_{q,K-1})',\ \text{a } (K-1)\times 1 \text{ vector, and} \tag{3.20}$$

$$\mathbf{y}^* = (\mathbf{y}^{*'}_1, \mathbf{y}^{*'}_2, ..., \mathbf{y}^{*'}_{\mathbf{Q}})',\ \text{a } (Q\times(K-1))\times 1 \text{ vector.}$$

Define the following additional matrices:

$$\boldsymbol{B}_q = (V_{q1} - V_{qK}, V_{q2} - V_{qK}, ..., V_{q,K-1} - V_{qK})'\ [(K-1)\times 1 \text{ vector}], \tag{3.21}$$

$$\boldsymbol{B} = (\boldsymbol{B}'_1, \boldsymbol{B}'_2, ..., \boldsymbol{B}'_Q)'\ [Q(K-1)\times 1 \text{ vector}].$$

It is easy to see that $\mathbf{y}^*$ has a mean vector of $\boldsymbol{B}$. To determine the covariance matrix of $\mathbf{y}^*$, define the following additional matrices:

$$\tilde{\boldsymbol{\Omega}} = \tilde{\mathbf{z}}(\mathbf{IDEN}_{\mathbf{Q}} \otimes \boldsymbol{\Omega})\tilde{\mathbf{z}}'\ \ [Q(K-1)\times Q(K-1) \text{ matrix}],^{14} \text{ and} \tag{3.22}$$

$$\boldsymbol{\Sigma} = \mathbf{S}\tilde{\boldsymbol{\Omega}}\mathbf{S}'\ \ [Q(K-1)\times Q(K-1) \text{ matrix}].$$

Based on property 2 earlier in Section 3.1, we can derive the location and other parameters of the vector $\mathbf{y}^*$, which is also skew-normally distributed. Specifically, by successive applications of property 2, we obtain the following important result:

[14] This way to construct the correlation matrix imposes restrictions for the **ρ** that can substantially constrain the values of its components in order to ensure the positive semidefiniteness of the matrix.

$\mathbf{y}^*$ ~ MVSN ($\boldsymbol{B}, \breve{\mathbf{z}}\breve{\boldsymbol{\rho}}, \boldsymbol{\Sigma}$). (3.23)

where $\breve{\boldsymbol{\rho}}$ is the stacked vector $\breve{\boldsymbol{\rho}} = (\boldsymbol{\rho}', \boldsymbol{\rho}', ..., \boldsymbol{\rho}')'$ [$QD \times 1$ vector].

3.2.3 Model Estimation

Part of the notation used in this section has been extracted from Bhat *et al.*'s (2015) formulation. The parameter vector to be estimated in the model is denoted $\boldsymbol{\theta} = (\boldsymbol{b}', \boldsymbol{\rho}', \boldsymbol{\gamma}', \text{Vech}(\boldsymbol{\delta}), \text{Vech}(\boldsymbol{\Omega}))'$. where $\text{Vech}(\boldsymbol{\Omega})$ represent the column vector of upper triangle elements and $\text{Vech}(\boldsymbol{\delta})$ represents the column vector of diagonal elements of $\boldsymbol{\delta}$. Several restrictive models are obtained from the spatial model formulation developed here. If $\boldsymbol{\delta} = \mathbf{0}$, but $\boldsymbol{\rho} \neq \mathbf{0}$, the result is the aspatial skew-normal MDC (ASN-MDC) model. If $\boldsymbol{\delta} \neq \mathbf{0}$, but $\boldsymbol{\rho} = \mathbf{0}$, the result is the spatial MDCP (S-MDCP) model that has been proposed by Bhat *et al.* (2015). If both $\boldsymbol{\delta} = \mathbf{0}$ and $\boldsymbol{\rho} = \mathbf{0}$, the result is the aspatial MDCP (A-MDCP) model.

3.2.3.1. Development of the Maximum Likelihood Estimator

Let $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, ..., \gamma_{K-1})'$. The parameters to estimate in the spatial skew-normal MDC model include the $\boldsymbol{\gamma}$ parameter vector, the $\boldsymbol{b}$ vector, the elements of the spatial lag parameter matrix $\boldsymbol{\delta}$, the skew parameter vector $\boldsymbol{\rho}$, and the covariance matrix $\boldsymbol{\Omega}$. The data provide information on the vector of exogenous variables $\boldsymbol{z_q}$ for each individual q, the spatial weight matrix $\widetilde{\boldsymbol{W}}$, and the observed goods consumption vector across the alternatives for each individual q: $\boldsymbol{x}_q^* = (x_{q1}^*, x_{q2}^*, ... x_{q,K}^*)'$. Note that a specific individual may not consume some alternatives, in which case the corresponding x_{qk}^* values take a value of zero.

Next, partition the vector $\mathbf{y}^*$ into a sub-vector $\widetilde{\boldsymbol{y}}_{NC}^*$ of length $L_{NC} \times 1$ ($[0 \leq L_{NC} \leq Q(K-1)]$) corresponding to the individual and good type combinations in which there is no consumption, and another sub-vector $\widetilde{\boldsymbol{y}}_C^*$ of length $L_C \times 1$ ($[0 \leq L_C \leq Q(K-1)]$) for the individual and good type combinations in which there is consumption ($[L_{NC} + L_C = Q(K-1)]$). In forming the sub-vector $\widetilde{\boldsymbol{y}}_C^*$, the outside alternative is not included. Let $\widetilde{\boldsymbol{y}}^* = \left(\left[\widetilde{\boldsymbol{y}}_{NC}^* \right]', \left[\widetilde{\boldsymbol{y}}_C^* \right]' \right)'$, which may be obtained from $\mathbf{y}^*$ as $\widetilde{\boldsymbol{y}}^* = \mathbf{R}\mathbf{y}^*$, where $\mathbf{R}$ is a re-

arrangement matrix of dimension $Q(K-1)\times Q(K-1)$ with zeroes and ones. For example, consider the case of three individuals and five goods. The last alternative is the outside alternative. Among the remaining four alternatives, let individual 1 be consuming alternatives 1 and 4 (not consuming =alternatives 2 and 3), let individual 2 be consuming alternatives 2 and 3 (not consuming alternatives 1 and 4), and let individual 3 be consuming alternative 1 (not consuming alternatives 2, 3, and 4). In this case, $L_{NC}=7$ and $L_C=5.$ Then, the re-arrangement matrix **R** is:

$$\mathbf{R} == \left[\begin{array}{cccc|cccc|cccc} 0&1&0&0&0&0&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&1&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&1&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&0&0&0&1&0\\ 0&0&0&0&0&0&0&0&0&0&0&1\\ \hline 1&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&1&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&1&0&0&0&0&0&0\\ 0&0&0&0&0&0&1&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&1&0&0&0 \end{array}\right] = \left[\begin{array}{c} \mathbf{R}_{NC}\\ \hline \mathbf{R}_C \end{array}\right], \tag{3.24}$$

where the upper sub-matrix $\mathbf{R}_{NC}$ corresponds to the individual and good alternative combinations with no consumption (of dimension $L_{NC}\times Q(K-1)$) and the lower sub-matrix $\mathbf{R}_C$ corresponds to the individual and good alternative combinations (excluding the outside alternative for each individual) with positive consumption (of dimension $L_C\times Q(K-1)$). Note also that $\tilde{\mathbf{y}}^*_{NC}=\mathbf{R}_{NC}\mathbf{y}^*$ and $\tilde{\mathbf{y}}^*_C=\mathbf{R}_C\mathbf{y}^*$.[15]

[15] $\mathbf{R}_{NC}$ has as many rows and columns as the number of individual and good alternative combinations with no consumption (each column corresponds to an alternative except the K^{th} alternative). Then, for each row, $\mathbf{R}_{NC}$ has a value of "1" in one of the columns corresponding to a individual- alternative combination that is not consumed (starting from the first alternative that is not consumed for the first individual and working down to the last alternative that is not consumed for the last individual). Each row has strictly one column with a value of "1" and the value of "0" everywhere else. A similar construction is involved in creating the $\mathbf{R}_C$ matrix.

Consistent with the above re-arrangement, define $\tilde{\boldsymbol{H}} = \mathbf{R}\boldsymbol{B}$, $\tilde{\boldsymbol{H}}_{\boldsymbol{NC}} = \mathbf{R}_{NC}\boldsymbol{B}$, $\tilde{\boldsymbol{H}}_{\boldsymbol{C}} = \mathbf{R}_{C}\boldsymbol{B}$, and $\tilde{\boldsymbol{\Sigma}} = \mathbf{R}\boldsymbol{\Sigma}\mathbf{R}'$. Define $\boldsymbol{x}^* = \left(\left[\boldsymbol{x}_1^*\right]', \left[\boldsymbol{x}_2^*\right]', \ldots \left[\boldsymbol{x}_Q^*\right]'\right)'$. Then, the maximum likelihood function may be obtained as:

$$L_{ML}(\boldsymbol{\theta}) = \text{Prob}(\boldsymbol{x}^*) = \det(\mathbf{J}) \int_{\boldsymbol{h}_{NC}=-\infty}^{\mathbf{0}_{\mathbf{L_{NC}}}} f_{Q(K-1)}\left(\mathbf{h}_{\mathbf{NC}}, \mathbf{0}_{\mathbf{L_C}} \mid \tilde{\boldsymbol{H}}, \tilde{\boldsymbol{\Sigma}}\right) d\boldsymbol{h}_{\boldsymbol{NC}}, \tag{3.25}$$

$$L_{ML}(\boldsymbol{\theta}) = \text{Prob}(\boldsymbol{x}^*) = \det(\mathbf{J}) \int_{\boldsymbol{h}_{NC}=-\infty}^{\mathbf{0}_{\mathbf{L_{NC}}}} \left(\prod_{q=1}^{Q} \tilde{\omega}_j\right)^{-1} \left[2\phi\left(\tilde{\boldsymbol{\omega}}^{-1}(\mathbf{h}_{\text{NC}} - \tilde{\boldsymbol{H}}); \tilde{\boldsymbol{\Sigma}}_{+}^{*}\right)\right]\left[\Phi(\boldsymbol{\alpha}'\mathbf{h}_{\text{NC}}; \tilde{\boldsymbol{\Sigma}}_{+}^{*})\right] d\mathbf{h}_{\text{NC}},$$

where $\tilde{\boldsymbol{\Sigma}}_{+}^{*} = \begin{pmatrix} 1 & \tilde{\boldsymbol{\rho}}' \\ \tilde{\boldsymbol{\rho}} & \tilde{\boldsymbol{\Sigma}} \end{pmatrix}$, $\boldsymbol{\alpha} = \dfrac{(\tilde{\boldsymbol{\Omega}}^*)^{-1}\tilde{\boldsymbol{\rho}}}{\left(1 - \tilde{\boldsymbol{\rho}}'(\tilde{\boldsymbol{\Omega}}^*)^{-1}\tilde{\boldsymbol{\rho}}\right)^{1/2}}$ and $\tilde{\boldsymbol{\rho}} = \mathbf{S}\ddot{\boldsymbol{z}}\breve{\boldsymbol{\rho}}$ The likelihood function in Equation (3.25) involves integration of dimension L_{NC}. This is of very high dimensionality in the typical case of sample sizes of 500 observations or more. The lower bound of L_{NC} is equal to zero, corresponding to the case when each individual consumes each good alternative. The upper bound is equal to $(K-1)*Q$, corresponding to the case when each individual consumes only the outside alternative. Of course, in practice, the situation will be somewhere between these two extreme values for L_{NC}, but the value for L_{NC} will be sufficient to render maximization of the likelihood function using traditional simulation methods almost impractical. In particular, existing estimation methods, including the Maximum Simulated Likelihood (MSL) method and the Bayesian Inference method, become cumbersome and encounter convergence problems even for moderately sized Q (Bhat *et al.*, 2010). In this research, we instead use Bhat's Maximum Approximate Composite Marginal Likelihood (MACML) inference approach for estimation.

To write the pairwise CML function, let $L_{qq',NC} = L_{q,NC} + L_{q',NC}$ and $L_{qq',C} = L_{qC} + L_{q'C}$. Define a vector $\boldsymbol{y}_{qq'}^*$ of size $\left[2(K-1)\times 1\right]$ as follows:

$$\boldsymbol{y}_{\boldsymbol{qq'}}^* = \left(\left[\mathbf{y}_{\mathbf{q}}^*\right]', \left[\mathbf{y}_{\mathbf{q'}}^*\right]'\right)'. \tag{3.26}$$

Let $\boldsymbol{\Delta}_{\boldsymbol{qq'}}$ be a selection matrix of size $2\times Q$. This matrix has the value of "1" in the top row and the column q, and the value of "1" in the bottom row and column q'. All other cells of this

matrix are filled with values of zero. Then, $\boldsymbol{y}^{*}_{qq'} \sim MVN_{2\times(K-1)}(\boldsymbol{B}_{\boldsymbol{qq'}}, \boldsymbol{\Sigma}_{\boldsymbol{qq'}})$, where $\boldsymbol{B}_{\boldsymbol{qq'}} = (\boldsymbol{\Delta}_{\boldsymbol{qq'}} \otimes \mathbf{IDEN}_{K-1})\boldsymbol{B}$, and $\boldsymbol{\Sigma}_{\boldsymbol{qq'}} = (\boldsymbol{\Delta}_{\boldsymbol{qq'}} \otimes \mathbf{IDEN}_{K-1})\boldsymbol{\Sigma}(\boldsymbol{\Delta}_{\boldsymbol{qq'}} \otimes \mathbf{IDEN}_{K-1})'$. Next, define the re-arrangement matrices $\mathbf{R}_{qq'}$ (of dimension $2(K-1)\times 2(K-1)$), $\mathbf{R}_{qq',NC}$ (of dimension $L_{qq',NC}\times 2(K-1)$), and $\mathbf{R}_{qq',C}$ (of dimension $L_{qq',C}\times 2(K-1)$) similar to the corresponding re-arrangement matrices defined on the entire sample for the maximum likelihood approach. Also, define $\tilde{\boldsymbol{B}}_{\boldsymbol{qq',NC}} = \mathbf{R}_{qq',NC}\boldsymbol{B}_{\boldsymbol{qq'}}$, $\tilde{\boldsymbol{B}}_{\boldsymbol{qq',C}} = \mathbf{R}_{qq',C}\boldsymbol{B}_{\boldsymbol{qq'}}$, and $\tilde{\boldsymbol{\Sigma}}_{qq'} = \mathbf{R}_{qq'}\boldsymbol{\Sigma}_{qq'}\mathbf{R}'_{qq'} = \begin{bmatrix} \tilde{\boldsymbol{\Sigma}}_{qq',NC} & \tilde{\boldsymbol{\Sigma}}'_{qq',NC,C} \\ \tilde{\boldsymbol{\Sigma}}_{qq',NC,C} & \tilde{\boldsymbol{\Sigma}}_{qq',C} \end{bmatrix}$, where $\tilde{\boldsymbol{\Sigma}}_{qq',NC} = \mathbf{R}_{qq',NC}\boldsymbol{\Sigma}_{qq'}\mathbf{R}'_{qq',NC}$, $\tilde{\boldsymbol{\Sigma}}_{qq',C} = \mathbf{R}_{qq',C}\boldsymbol{\Sigma}_{qq'}\mathbf{R}'_{qq',C}$, and $\tilde{\boldsymbol{\Sigma}}_{qq',NC,C} = \mathbf{R}_{qq',NC}\boldsymbol{\Sigma}_{qq'}\mathbf{R}'_{qq',C}$. Let $\breve{\boldsymbol{B}}_{qq',NC} = \tilde{\boldsymbol{B}}_{\boldsymbol{qq',NC}} + \tilde{\boldsymbol{\Sigma}}'_{qq',NC,C}(\tilde{\boldsymbol{\Sigma}}_{qq',C})^{-1}(-\tilde{\boldsymbol{B}}_{\boldsymbol{qq',C}})$, $\breve{\boldsymbol{\Sigma}}_{qq',NC} = \tilde{\boldsymbol{\Sigma}}_{qq',NC} - \tilde{\boldsymbol{\Sigma}}'_{qq',NC,C}(\tilde{\boldsymbol{\Sigma}}_{qq',C})^{-1}\tilde{\boldsymbol{\Sigma}}_{qq',NC,C}$, $\tilde{\boldsymbol{B}}^{*}_{\boldsymbol{qq',C}} = \boldsymbol{\omega}^{-1}_{\tilde{\boldsymbol{\Sigma}}_{qq',C}}\left(-\tilde{\boldsymbol{B}}_{\boldsymbol{qq',C}}\right)$, $\breve{\boldsymbol{B}}^{*}_{qq',NC} = \boldsymbol{\omega}^{-1}_{\breve{\boldsymbol{\Sigma}}_{qq',NC}}\left(-\breve{\boldsymbol{B}}_{\boldsymbol{qq',NC}}\right)$, $\breve{\boldsymbol{\Sigma}}^{*}{}_{qq',NC} = \boldsymbol{\omega}^{-1}_{\breve{\boldsymbol{\Sigma}}_{qq',NC}}\breve{\boldsymbol{\Sigma}}_{qq',NC}\boldsymbol{\omega}^{-1}_{\breve{\boldsymbol{\Sigma}}_{qq',NC}}$, and $\tilde{\boldsymbol{\Sigma}}^{*}{}_{qq',C} = \boldsymbol{\omega}^{-1}_{\tilde{\boldsymbol{\Sigma}}_{qq',C}}\tilde{\boldsymbol{\Sigma}}_{qq',C}\boldsymbol{\omega}^{-1}_{\tilde{\boldsymbol{\Sigma}}_{qq',C}}$ where $\boldsymbol{\omega}_{\breve{\boldsymbol{\Sigma}}_{qq',NC}}$ is the diagonal matrix of standard deviations of $\breve{\boldsymbol{\Sigma}}_{qq',NC}$ and $\boldsymbol{\omega}_{\tilde{\boldsymbol{\Sigma}}_{qq',C}}$ is the diagonal matrix of standard deviations of $\tilde{\boldsymbol{\Sigma}}_{qq',C}$. Let $\overline{\boldsymbol{\omega}}_{\tilde{\boldsymbol{\Sigma}}_{qq',C}}$ be the product of the diagonal elements of $\boldsymbol{\omega}_{\tilde{\boldsymbol{\Sigma}}_{qq',C}}$, and write the determinant of the Jacobian corresponding to individuals q and q' as $\det(\mathbf{J}_{qq'}) = \prod_{l=q,q'}\left[\left\{\prod_{k\in\tilde{L}_{lC}}\frac{1-\alpha_k}{x^{*}_{lk}+\gamma_k}\right\}\left\{\sum_{k\in\tilde{L}_{lC}}\left(\frac{x^{*}_{lk}+\gamma_k}{1-\alpha_k}\right)\right\}\right]$. Then, using the marginal and conditional distribution properties of the multivariate normal distribution, the pairwise CML function for the SSN-MDC model can be written as:

$$
\begin{aligned}
L_{CML}(\boldsymbol{\theta}) &= \text{Prob}(\boldsymbol{x}^{*}_{q}, \boldsymbol{x}^{*}_{q'}) \\
&= \prod_{q=1}^{Q-1}\prod_{q'=q+1}^{Q}\det(\mathbf{J}_{qq'})\times\left(\overline{\boldsymbol{\omega}}_{\tilde{\boldsymbol{\Sigma}}_{qq',C}}\right)^{-1}\left[2\phi\left(\tilde{\boldsymbol{\omega}}^{-1}\tilde{\boldsymbol{B}}^{*}_{qq',C};\tilde{\boldsymbol{\Sigma}}^{*}{}_{qq',C}\right)\right]\left[\Phi(\boldsymbol{\alpha}'\tilde{\boldsymbol{B}}^{*}_{qq',C};\tilde{\boldsymbol{\Sigma}}^{*}{}_{qq',C})\right].
\end{aligned}
\tag{3.27}
$$

The CML function above requires the computation of the multivariate normal cumulative distribution (MVNCD) function that is utmost of dimension $(K-1)*2$ integrals (instead of $(K-1)*Q$ in the full maximum likelihood case). Such integrals may be computed easily using the MVNCD approximation method embedded in the MACML method (the MVNCD function approximates the pairwise probabilities in Equation (3.27) using only univariate and bivariate cumulative normal distribution functions; see Bhat, 2011 and Bhat and Sidharthan, 2012).

The MACML estimator is obtained by maximizing the logarithm of the function in Equation (3.27) after evaluating the MVNCD function using the analytic approximation. Since the MACML estimator entails only the computation of univariate and bivariate cumulative normal distribution functions, it is extremely quick to evaluate. The covariance matrix is given by the inverse of Godambe's (1960) sandwich information matrix (see Zhao and Joe, 2005). Bhat (2011) exploits the fading spatial dependence pattern implied by the spatial lag structure (due to the decaying nature of the distance weight matrix, combined with the spatial lag parameter being less than 1) to propose a specific implementation of Heagerty and Lumley's (2000) windows sampling procedure to estimate this sandwich information matrix.

The pairwise CML function of Equation (3.27) comprises $Q(Q-1)/2$ individual pairs of probability computations. To further accelerate the estimation, one can reduce the number of individual pairs because spatial dependency drops quickly with inter-observations distance. In fact, as demonstrated by Bhat *et al.* (2010) and Varin and Czado (2010), retaining all pairs not only increases computational costs, but may also reduce estimator efficiency. We examine this issue by creating different distance bands and, for each specified distance band, we consider only those pairings in the CML function that are within the spatial distance band. Then, we develop the asymptotic variance matrix $\mathbf{V}_{CML}(\hat{\boldsymbol{\theta}})$ for each distance band and select the threshold distance value that minimizes the total variance across all parameters as given by $tr[\mathbf{V}_{CML}(\hat{\boldsymbol{\theta}})]$ (*i.e.*, the trace of the matrix $[\mathbf{V}_{CML}(\hat{\boldsymbol{\theta}})]$).

A final issue regarding estimation. To ensure the constraints on the δ_k $(k=1,2,...,K-1)$ autoregressive terms, we parameterize these as $\delta_k = 1/[1+\exp(\tilde{\delta}_k)]$. Once estimated, the $\tilde{\delta}_k$ estimates can be translated back to estimates of δ_k.

3.3 SIMULATION STUDY

The simulation exercises undertaken in this section examine the ability of the MACML estimator to recover parameters from finite samples in a SSN-MDC model by generating simulated data sets with known underlying model parameters. We also examine the effects of (a) assuming normality in the random coefficients and kernel error term when they are distributed skew-normally, and (b) ignoring spatial dependence.

3.3.1 Experimental Design

In the design, we generate 750 observations (*i.e.*, Q = 750) using pre-specified values for the $\boldsymbol{\theta}$ vector. We consider the case with three alternatives (K=3). The last alternative is assumed to be the outside alternative. We allow for two independent variables in the z_{qk} vector in the baseline utility for each alternative. The values of the two independent variables for each alternative are drawn from standard univariate normal distributions. Additionally, we generate budget amounts $E_q \ \ (q = 1,2,...,Q)$ from a univariate normal distribution with a mean of 150, and truncated between the values of 100 and 200. The prices of all goods are fixed at the value of one across all consumers. In the simulations, the 750 individuals are located on a rectangular grid with the longer side containing 50 locations spaced 1 unit apart and the shorter side containing 15 locations spaced 1 unit apart. The spatial weight matrix **W** (of size 750×750) is created using the inverse of the distance on the coordinate plane between observational units. Once generated, the independent variable values, the spatial configuration and weights, and the total budget are held fixed in the rest of the simulation exercise.

The first two components of the coefficient vector $\boldsymbol{\beta_q}$ correspond to the alternative specific constants for alternatives 1 and 2 (the inside goods).[16] In order to represent the kernel error terms, we will allow these two constants to be random. The third and fourth components of vector $\boldsymbol{\beta_q}$ correspond to the coefficients associated to the two independent variables. We consider the third coefficient random, but the last one is fixed. In summary, $\boldsymbol{\beta_q}$ is allowed to be random according to a trivariate skew-normal distribution for the first three coefficients, but set to be fixed for the fourth coefficient. The mean vector for $\boldsymbol{\beta_q}$ is assumed to be $\boldsymbol{b}$ = (0.5, 1.0, –1.0, 0.8).

In the simulations, we set the γ_k parameters for the first two alternatives to the value of one. Finally, to examine the potential impact of different levels of spatial dependence and skewness on the ability of the MACML approach to recover model parameters, we consider two sets of values of the spatial autoregressive coefficients corresponding to low dependence $(\delta_1 = 0.1,\ \delta_2 = 0.2)$ and high dependence $(\delta_1 = 0.7,\ \delta_2 = 0.8)$, as well as two sets of values for the skew parameters corresponding to low rightward skew $(\rho_1 = 0.3,\ \rho_2 = 0.3, \rho_3 = 0.3)$ and

[16] Readers should remember that the constant of the last alternative is normalized to 0.

high rightward skew $(\rho_1 = 0.7,\ \rho_2 = 0.7, \rho_3 = 0.7)$. Thus, in total, there are four possible combinations of the spatial lag and skew coefficients considered in the simulations.

The correlation matrix $\mathbf{\Omega}_+^*$ (and the corresponding $\boldsymbol{\omega}$ matrix)for the three random coefficients are specified as follows:

$$\boldsymbol{\omega} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1.25 \end{pmatrix}, \text{ and } \mathbf{\Omega}_+^* = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & 1 & 0.49 & 0.49 \\ \rho_2 & 0.49 & 1 & 0.49 \\ \rho_3 & 0.49 & 0.49 & 1 \end{pmatrix} \text{ for the high skew case and}$$

$$\boldsymbol{\omega} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1.25 \end{pmatrix}, \text{ and } \mathbf{\Omega}_+^* = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & 1 & 0.09 & 0.09 \\ \rho_2 & 0.09 & 1 & 0.09 \\ \rho_3 & 0.09 & 0.09 & 1 \end{pmatrix} \text{ for the low skew case.}$$

The correlation matrix $\mathbf{\Omega}_+^*$ above is constructed in a specific manner so that the off-diagonal elements of the corresponding Cholesky matrix are all zero, except for the first column which now contains the skew parameters as its elements. The Cholesky matrix of $\mathbf{\Omega}_+^*$ for the high skew case is

$$\mathbf{L} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0.7 & 0.7141 & 0 & 0 \\ 0.7 & 0 & 0.7141 & 0 \\ 0.7 & 0 & 0 & 0.7141 \end{pmatrix}$$

and for the low skew case is

$$\mathbf{L} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0.3 & 0.9539 & 0 & 0 \\ 0.3 & 0 & 0.9539 & 0 \\ 0.3 & 0 & 0 & 0.9539 \end{pmatrix}$$

The set-up above is used to develop the $[Q(K-1)\times 1]$ vector $\boldsymbol{Szb}$ and the covariance matrix $\boldsymbol{\Sigma}$. Since $\overline{\boldsymbol{\psi}} \sim MVSN(\boldsymbol{Szb}, \boldsymbol{\Sigma})$, a specific realization of the $[Q(K-1)\times 1]$ vector for $\overline{\boldsymbol{\psi}}$ is drawn from the multivariate skew-normal distribution. The method to generate realizations from

the MVSN distribution is based on first drawing a multivariate standard normal vector in the usual way. This constitutes a draw for the latent underlying 4-variate normally distributed vector $\tilde{\mathbf{M}} = (\tilde{M}_1, \tilde{\mathbf{M}}_2')'$, where $\tilde{M}_1$ is a latent (1×1)-vector and $\tilde{\mathbf{M}}_2$ is a (3×1)-vector. From this multivariate standard normal draw, a 3-variate vector from the multivariate standard skew normal distribution is generated as follows:

$$\mathbf{Z} = \begin{cases} \tilde{\mathbf{M}}_2 & \text{if } \tilde{M}_1 > 0 \\ -\tilde{\mathbf{M}}_2 & \text{if } \tilde{M}_1 \le 0. \end{cases} \tag{3.28}$$

Then, using subsets of this $\overline{\boldsymbol{\psi}}$ vector corresponding to each individual, and the specified γ vector, we generate the investment quantity vector $\boldsymbol{x}_q^*$, using the forecasting algorithm proposed by Pinjari and Bhat (2011). The above data generation process is undertaken 30 times with different realizations of the $\overline{\boldsymbol{\psi}}$ vector to generate 30 different data sets each for the four possible combinations of spatial dependence and skewness.

The MACML estimator is applied to each data set to estimate data specific values of $\boldsymbol{\theta} = (\boldsymbol{b}', \boldsymbol{\delta}', \boldsymbol{\rho}', \boldsymbol{\gamma}', \text{Vech}(\boldsymbol{\omega}), \text{Vech}(\boldsymbol{\Omega}))'$. A single random permutation is generated for each individual (the random permutation varies across individuals, but is the same across iterations for a given individual) to decompose the multivariate normal cumulative distribution (MVNCD) function into a product sequence of marginal and conditional probabilities (see Section 2.1 of Bhat, 2011).[17] All the $Q(Q-1)/2$ pairings of individuals are considered in the MACML estimator. The estimator is applied to each dataset 10 times with different permutations to obtain the approximation error, computed as the standard deviation of estimated parameters among the 10 different estimates on the same data set.

3.3.2 Performance Evaluation

The MACML estimation procedure is applied to each data set to estimate data-specific values. The Godambe information-based covariance matrix and the corresponding standard errors are also computed. For each combination of spatial and skew parameters, the performance of the MACML estimation procedure is assessed based on the following performance characteristics:

[17] Technically, the MVNCD approximation should improve with a higher number of permutations in the MACML approach. However, when we investigated the effect of different numbers of random permutations per individual, we noticed little difference in the estimation results between using a single permutation and higher numbers of permutations, and hence we settled with a single permutation per individual.

(1) First, estimate the MACML parameters for each data set and for each of 10 independent sets of permutations. Estimate the standard errors (s.e.) using the Godambe (sandwich) estimator.

(2) Second, the mean estimate for each model parameter across the 10 random permutations is obtained and labeled as MED. Then take the mean of the MED values across the 30 data sets to obtain a mean estimate. The parameter-specific mean absolute percentage (finite sample) bias or APB value (relative to the "true" value of the parameter) is computed as $APB = \left| \frac{\text{mean estimate - true value}}{\text{true value}} \right| \times 100$.

(3) Third, the mean standard error for each model parameter is computed across the 10 permutations. This is labeled as MSED, and then the mean of the MSED values across the 30 data sets is computed. This is labeled as the asymptotic standard error (ASE) for the parameter (essentially this is the standard error of the distribution of the estimator as the sample size gets large). Also, the standard deviation of the MED values across the 30 datasets is computed, and labeled as the finite sample standard error or FSEE (essentially, this is the empirical standard error). Then, to evaluate the accuracy of the asymptotic standard error formula as computed using the MACML inference approach for the finite sample size used, we the relative efficiency (RE) of the estimator is computed as follows: $\text{RE} = \frac{\text{ASE}}{\text{FSEE}}$. Relative efficiency values in the range of 0.75-1.25 indicate that the ASE, as computed using the Godambe matrix in the MACML method, does provide a good approximation of the FSSE.

(4) Fourth, a Monte Carlo estimate of the coverage probability or COVP (that is, the probability that the true estimate lies within the 95% confidence interval) is obtained by determining the percentage of times (across the 30 data sets and the corresponding 10 permutations for each data set) that the true parameter lies within the 95% coverage bound of the CML estimator (see Koehler *et al.*, 2009). That is, for each data set and for each parameter, an indicator value of '1' is assigned if the true value is within a range of ± 1.96 times the ASE value of the estimated parameter on that data set. If not, an indicator value of '0' is assigned. Then, for each parameter, the percentage of '1' values across the 300 (30x10) draws is designated as the COVP value.

3.3.3 Comparison with More Restrictive Models

The main purpose of our proposed model is to accommodate spatial dynamics and skewed random distributions. Therefore, to examine the potential problems that could arise from ignoring spatial dynamics and skewness, we estimate three additional models on the 30 data sets generated for each combination of spatial and skew levels. The first model ignores the spatial autocorrelation coefficients δ_k (that is, assumes $\boldsymbol{\delta} = \mathbf{0}$, leading to the ASN-MDC model), the second model assumes away any skewness (that is, assumes that $\boldsymbol{\rho} = \mathbf{0}$, leading to the S-MDCP model), and the third model ignores both the spatial coefficients and the skew parameters (that is, assumes $\boldsymbol{\delta} = \mathbf{0}$ and $\boldsymbol{\rho} = \mathbf{0}$, leading to the A-MDCP model). We compare these three restrictive formulations with the general SSN-MDC model based on the mean APB measure across all parameters and the adjusted composite log-likelihood ratio test (ADCLRT) value (see Pace *et al.*, 2011 and Bhat, 2011 for more details on the ADCLRT statistic, which is the equivalent of the log-likelihood ratio test statistic when a composite marginal likelihood inference approach is used; this statistic has an approximate chi-squared asymptotic distribution). The ADCLRT statistic needs to be computed for each data set separately, and compared with the chi-squared table value with the appropriate degrees of freedom. Here we identify the number of times (out of the 30 model runs, one run for each of the 30 data sets) that the ADCLRT value rejects the restrictive models in favor of the proposed SSN-MDC model.

3.3.4 Simulation Results

3.3.4.1 Recoverability of Parameters in the SSN-MDC Model

Tables 3.1a, 3.1b, 3.1c and 3.1d present the results of the simulation exercises. Table 3.1a corresponds to the low spatial dependency and low skewness case, Table 3.1b corresponds to the low spatial dependency and high skewness case, Table 3.1c corresponds to the high spatial dependency and low skewness case, and Table 3.1d corresponds to the high spatial dependency and high skewness case. The first column indicates the notation of each parameter based on the simulation design presented in Section 3.3.1. The second column presents the true values used in generating the data samples. The third broad column labeled "Parameter Estimates" provides the mean value (across the data sets) of each parameter as well as the corresponding absolute bias, APB measure, and Monte-Carlo coverage probability, while the fourth broad column labeled

"Standard Error Estimates" provides the FSSE, ASE, and the relative efficiency values for the parameter standard errors.

Overall, the MACML inference approach does well in accurately and precisely recovering parameters in the four cases. The overall mean APB value across all parameters is smaller than a 4% in all cases. The APB values are in general higher for higher spatial dependency. For example, within the low skewness cases, the mean APB of the low spatial dependency case is 3.63%, which is smaller than the mean APB in the case of high spatial dependency (3.69%). The same happens with the high skewness cases, where the low spatial dependency case and high spatial dependency case have a mean APB value of 3.62% and 3.79% respectively. Another way to corroborate this is the COVP, which is smaller for the high spatial dependency case (in comparison with the low spatial dependency case) within the low skewness cases. We witness the same phenomenon within the high skewness cases. In general, across the parameters, the APB values are relatively high for the skew parameters in all four cases. Previous studies (see, for example, Bhat *et al.*, 2016c) have also shown that the skew parameters are difficult to pin down. The finite sample standard errors and the asymptotic standard errors are close; the relative efficiency is always between 0.75 and 1.25 in all four cases and for any parameter.

3.3.4.2 Comparison between the Proposed Model and More Restrictive MDC Models

Table 3.2 presents the results for the simulation exercise focusing on the comparison between the proposed SSN-MDC model and three other, more restrictive versions of the model: the ASN-MDC, the S-MDCP, and the A-MDCP models. For the sake of brevity, only the high spatial dependency and high skewness case is presented. The APB values of the parameters are in general higher relative to the APB values of the parameters in the original model. The overall mean APB values across parameters are 5.5%, 7.1%. and 18.4% for the ASN-MDC, S-MDCP, and A-MDCP models, respectively—significantly higher in comparison with the mean APB value of 3.8% in the proposed model. The superior performance of the SSN-MDC model is also evidenced in the higher log-likelihood value, on average, for the SSN-MDC model across the 30 estimations (on the 30 data sets). In addition, for each of the 30 data sets, a likelihood ratio test comparing the SSN-MDC model with the three other models clearly rejects the alternative models in favor of the SSN-MDC model.

Table 3.1a: Simulation results for the four-alternative case with 30 datasets for low spatial dependency and low skewness (based on a total of 30×10 runs/dataset=300 runs)

Parameter	True Value	Parameter Estimates				Standard Error Estimates		
		Mean Est.	Abs. Bias	Absolute Percentage Bias (APB)	Coverage Probability (COVP)	Finite Sample St. Err. (FSSE)	Asymptotic St. Err. (ASE)	Relative Efficiency
b_1	0.50	0.480	0.020	4.00%	99.34%	0.021	0.022	1.07
b_2	1.00	1.020	0.020	2.00%	99.00%	0.018	0.018	1.00
b_3	-1.00	-0.980	0.020	-2.00%	98.67%	0.021	0.025	1.19
b_4	0.80	0.790	0.010	1.25%	99.67%	0.009	0.009	0.96
γ_1	1.00	0.970	0.030	3.00%	100.00%	0.031	0.036	1.19
γ_2	1.00	0.980	0.020	2.00%	99.34%	0.019	0.017	0.89
ω_1	1.00	0.980	0.020	2.00%	99.67%	0.018	0.020	1.07
ω_2	1.00	0.950	0.050	5.00%	99.00%	0.054	0.055	1.01
ω_3	1.25	1.210	0.040	3.20%	98.67%	0.039	0.045	1.18
δ_1	0.10	0.106	0.006	6.00%	100.00%	0.006	0.006	0.95
δ_2	0.20	0.202	0.002	1.00%	99.77%	0.002	0.002	1.08
ρ_1	0.30	0.280	0.020	6.67%	99.34%	0.018	0.017	0.94
ρ_2	0.30	0.270	0.030	10.00%	99.00%	0.027	0.024	0.87
ρ_3	0.30	0.280	0.020	6.67%	98.67%	0.019	0.020	1.04

Overall mean value across parameters	0.025	3.63%	99.30%	0.022	0.023	1.03

Table 3.1b: Simulation results for the four-alternative case with 30 datasets for low spatial dependency and high skewness (based on a total of 30×10 runs/dataset=300 runs)

Parameter	True Value	Parameter Estimates				Standard Error Estimates		
		Mean Est.	Abs. Bias	Absolute Percentage Bias (APB)	Coverage Probability (COVP)	Finite Sample St. Err. (FSSE)	Asymptotic St. Err. (ASE)	Relative Efficiency
b_1	0.50	0.478	0.022	4.40%	99.67%	0.022	0.023	1.04
b_2	1.00	1.025	0.025	2.50%	100.00%	0.025	0.021	0.84
b_3	-1.00	-0.978	0.022	-2.20%	98.34%	0.024	0.021	0.85
b_4	0.80	0.786	0.014	1.75%	99.34%	0.015	0.017	1.15
γ_1	1.00	0.967	0.033	3.30%	99.67%	0.035	0.032	0.90
γ_2	1.00	0.979	0.021	2.10%	99.67%	0.019	0.019	0.96
ω_1	1.00	0.970	0.030	3.00%	99.34%	0.030	0.033	1.08
ω_2	1.00	0.947	0.053	5.30%	98.67%	0.053	0.060	1.13
ω_3	1.25	1.200	0.050	4.00%	100.00%	0.053	0.057	1.07
δ_1	0.10	0.107	0.007	7.00%	98.67%	0.007	0.008	1.10
δ_2	0.20	0.204	0.004	2.00%	99.00%	0.004	0.003	0.94
ρ_1	0.70	0.751	0.051	7.29%	100.00%	0.051	0.051	1.00
ρ_2	0.70	0.748	0.048	6.86%	99.34%	0.045	0.046	1.02
ρ_3	0.70	0.676	0.024	3.43%	98.67%	0.022	0.026	1.20
Overall mean value across parameters			0.029	3.62%	99.31%	0.029	0.030	1.02

Table 3.1c: Simulation results for the four-alternative case with 30 datasets for high spatial dependency and low skewness (based on a total of 30×10 runs/dataset=300 runs)

Parameter	True Value	Parameter Estimates				Standard Error Estimates		
		Mean Est.	Abs. Bias	Absolute Percentage Bias (APB)	Coverage Probability (COVP)	Finite Sample St. Err. (FSSE)	Asymptotic St. Err. (ASE)	Relative Efficiency
b_1	0.50	0.477	0.023	4.60%	98.67%	0.021	0.024	1.15
b_2	1.00	1.032	0.032	3.20%	99.34%	0.031	0.028	0.91
b_3	-1.00	-0.969	0.031	-3.10%	98.00%	0.033	0.030	0.91
b_4	0.80	0.778	0.022	2.75%	99.00%	0.023	0.023	1.03
γ_1	1.00	0.962	0.038	3.80%	99.34%	0.039	0.036	0.93
γ_2	1.00	0.972	0.028	2.80%	99.34%	0.027	0.029	1.08
ω_1	1.00	1.037	0.037	3.70%	99.00%	0.040	0.042	1.04
ω_2	1.00	0.951	0.049	4.90%	98.34%	0.050	0.059	1.18
ω_3	1.25	1.198	0.052	4.16%	99.67%	0.049	0.046	0.94
δ_1	0.70	0.680	0.020	2.86%	98.34%	0.021	0.022	1.05
δ_2	0.80	0.781	0.019	2.38%	98.67%	0.019	0.021	1.11
ρ_1	0.30	0.326	0.026	8.67%	100.00%	0.026	0.028	1.08
ρ_2	0.30	0.287	0.013	4.33%	99.67%	0.012	0.011	0.88
ρ_3	0.30	0.320	0.020	6.67%	98.67%	0.021	0.023	1.09
Overall mean value across parameters			0.029	3.69%	99.00%	0.029	0.030	1.03

Table 3.1d: Simulation results for the four-alternative case with 30 datasets for high spatial dependency and high skewness (based on a total of 30×10 runs/dataset=300 runs)

Parameter	True Value	Parameter Estimates				Standard Error Estimates		
		Mean Est.	Abs. Bias	Absolute Percentage Bias (APB)	Coverage Probability (COVP)	Finite Sample St. Err. (FSSE)	Asymptotic St. Err. (ASE)	Relative Efficiency
b_1	0.50	0.470	0.030	6.00%	98.00%	0.032	0.038	1.20
b_2	1.00	1.039	0.039	3.90%	99.67%	0.040	0.039	0.98
b_3	-1.00	-0.960	0.040	-4.00%	97.67%	0.041	0.039	0.94
b_4	0.80	0.768	0.032	4.00%	99.00%	0.033	0.038	1.16
γ_1	1.00	0.959	0.041	4.10%	99.00%	0.038	0.042	1.11
γ_2	1.00	0.964	0.036	3.60%	99.67%	0.039	0.037	0.95
ω_1	1.00	1.041	0.041	4.10%	98.34%	0.044	0.036	0.82
ω_2	1.00	0.951	0.049	4.90%	98.00%	0.049	0.056	1.14
ω_3	1.25	1.188	0.062	4.96%	99.34%	0.059	0.049	0.83
δ_1	0.70	0.676	0.024	3.43%	98.00%	0.022	0.022	1.04
δ_2	0.80	0.774	0.026	3.25%	97.67%	0.024	0.024	0.99
ρ_1	0.70	0.669	0.031	4.43%	99.67%	0.028	0.030	1.07
ρ_2	0.70	0.732	0.032	4.57%	99.34%	0.030	0.024	0.80
ρ_3	0.70	0.741	0.041	5.86%	98.34%	0.043	0.048	1.12
Overall mean value across parameters			0.037	3.79%	98.69%	0.037	0.037	1.01

Table 3.2: Effects of ignoring spatial autocorrelation and skewness when present (for the high spatial dependence and high skewness case)

Parameters	True Value	ASN-MDC[2]		S-MDCP[3]		A-MDCP[4]	
		Mean Est.	Absolute Percentage Bias (APB)	Mean Est.	Absolute Percentage Bias (APB)	Mean Est.	Absolute Percentage Bias (APB)
b_1	0.50	0.515	2.94%	0.452	9.66%	0.451	19.48%
b_2	1.00	0.991	0.90%	1.093	9.28%	1.003	0.62%
b_3	-1.00	-0.997	0.34%	-0.866	13.40%	-0.912	17.64%
b_4	0.80	0.756	5.56%	0.816	1.95%	0.703	24.31%
γ_1	1.00	0.893	10.70%	0.911	8.85%	0.974	5.18%
γ_2	1.00	1.001	0.10%	1.016	1.56%	0.999	0.23%
ω_1	1.00	1.123	12.29%	0.993	0.72%	1.200	40.05%
ω_2	1.00	0.991	0.91%	0.885	11.51%	0.799	40.29%
ω_3	1.25	1.173	6.18%	1.121	10.30%	1.142	17.34%
δ_1	0.70	— [a]	—	0.731	4.43%	—	—
δ_2	0.80	—	—	0.851	6.35%	—	—
ρ_1	0.70	0.636	9.18%	—	—	—	—
ρ_2	0.70	0.718	2.61%	—	—	—	—
ρ_3	0.70	0.800	14.22%	—	—	—	—
Overall mean value across parameters		0.717	5.49%	0.727	7.09%	0.707	18.35%
Mean composite log-likelihood value at convergence		-94,724.15		-97,001.22		-99,574.42	
Number of times the adjusted composite likelihood ratio test (ADCLRT) statistic favors the SSN-MDC[1] model[b]		All thirty times when compared with $\chi^2_{2,0.99} = 9.21$ value (mean ADCLRT statistic is 18.3)		All thirty times when compared with $\chi^2_{3,0.99} = 11.34$ value (mean ADCLRT statistic is 38.6)		All thirty times when compared with $\chi^2_{5,0.99} = 15.09$ value (mean ADCLRT statistic is 41.7)	

[1] SSN-MDC: Spatial skew-normal MDC model
[2] ASN-MDC: Aspatial skew-normal MDC model
[3] S-MDCP: Spatial MDCP model
[4] A-MDCP: Aspatial MDCP model
[a] A "—"entry in a cell indicates that the corresponding parameter is not estimated and is fixed to the value mentioned in Section 4.3
[b] The mean composite log-likelihood value for the SSN-MDC model at converged parameters is -93,618.63.

3.4 APPLICATION DEMONSTRATION

In our empirical application, we study land-use over an entire urban region of Austin, Texas. The units of analysis are quarter-of-a-mile square grid cells. Our MDC variable corresponds to land-use type and intensity.

3.4.1 Background

Transportation and land-use are mutually dependent and inseparable. Transportation supply has affected land-use patterns, particularly how people choose to locate their homes and businesses. Conversely, spread out land-use patterns further increase the demand for transportation because of greater travel distances, and this has become an eternal cycle. Land-use change models have been used to meet land management needs, and to better understand the role of land-use change in the functioning of the land-use system. Land-use change models offer the possibility to test the sensitivity of land-use patterns to changes in selected variables and therefore, they can also be helpful for exploring future land-use changes under different scenario conditions. Land-use change models have not only been applied in transportation (or urban) planning, but also in many other fields, such as ecological science, climate science, environmental science, geography, watershed hydrology, and political science.

There are many types of land-use models. The reader is referred to Verburg *et al.* (2004), Irwin (2010), and Bhat *et al.* (2015) for comprehensive reviews on the characterization and classification of land-use change models. In our analysis, we borrow elements from three different types of models: pattern-based models, process-based models, and spatial-based models. A detailed explanation of these three model types can be found in Bhat *et al.* (2015). The important thing is that we use an aggregate spatial unit of analysis of a quarter-of-a-mile square grid cell, and we decide to model each grid as they have the "option" of investing (and converting) from one package of land-uses to another alternative package of land-uses. The grid-level land-use is obtained by aggregating underlying parcel-level land-use information. Additionally, we consider a rich set of population demographics of the citizenry of each aggregate grid to approximate a collective decision-making process for that grid. By using a grid size that is not too aggregate, we retain some of the process-based model characteristics of having a connection between the spatial unit of analysis and human decision-makers. The hybrid

model just discussed is further enhanced by considering spatial analysis aspects considered in spatial-based models, specifically, we consider spatial dependency between decisions makers.

3.4.2 Data and Sample Formation

A detailed description of the study area and the data preparation process can be found in Bhat *et al.* (2015). Not only the source data set, but also the estimation sample used in our application is the same as that in Bhat *et al.* (2015). In this section, we briefly summarizes the characteristics of the sample that are most relevant for our current analysis.

The data used in the empirical application of this study is drawn from parcel-level land-use inventory data for the year 2010, from Austin, Texas. The land-use type for each parcel is aggregated into the following four mutually exclusive land-use categories: (1) commercial land-use (including commercial, office, hospitals, government services, educational services, cultural services, and parking), (2) industrial land-use (including manufacturing, warehousing, resource extraction (mining), landfills, and miscellaneous industrial uses), (3) residential land-use (including single family, duplexes, three/four-plexes, apartments, condominiums, mobile homes, group quarters, and retirement housing), and (4) undeveloped land-use (including open and undeveloped spaces, preserves, parks, golf courses, and agricultural open spaces). The last among these alternatives serves as an "essential outside good" in that all grid cells inevitably will have at least some of their land area that remains undeveloped. In this case, there is no price variation (all the prices are equal to 1).

For the current analysis, an area measuring 145.94 mi^2 covering the central business district (CBD) and important surrounding areas is considered. The study area is divided into 2,383 square grids, each of size 0.25 mi × 0.25 mi. Figure 3.1 depicts the main elements of our area of study. The two major highways in Austin are the Interstate Highway 35 and Loop-1 MoPac. These two highways are "parallel" to each other. The other major thoroughfares in the Austin area that can be seen in the picture are: (1) Ben White Blvd (State Highway 71) that forms the southern boundary of the study region, (2) US-290, (3) US-183 that runs diagonally from the northwest to the southeast at the north end of the study area and then directly south at the south end of the study area, (4) Loop 360, and (5) FM-2222. In addition, several major arterials are also shown in the figure, including Lamar Blvd (roughly parallel to IH-35 and MoPac, and between these two highways), Parmer Lane (toward the north), Cesar Chavez Street

(just south of the downtown area), Martin Luther King Jr (MLK) Blvd (just north of the downtown area), Congress Avenue, and Dessau Road.

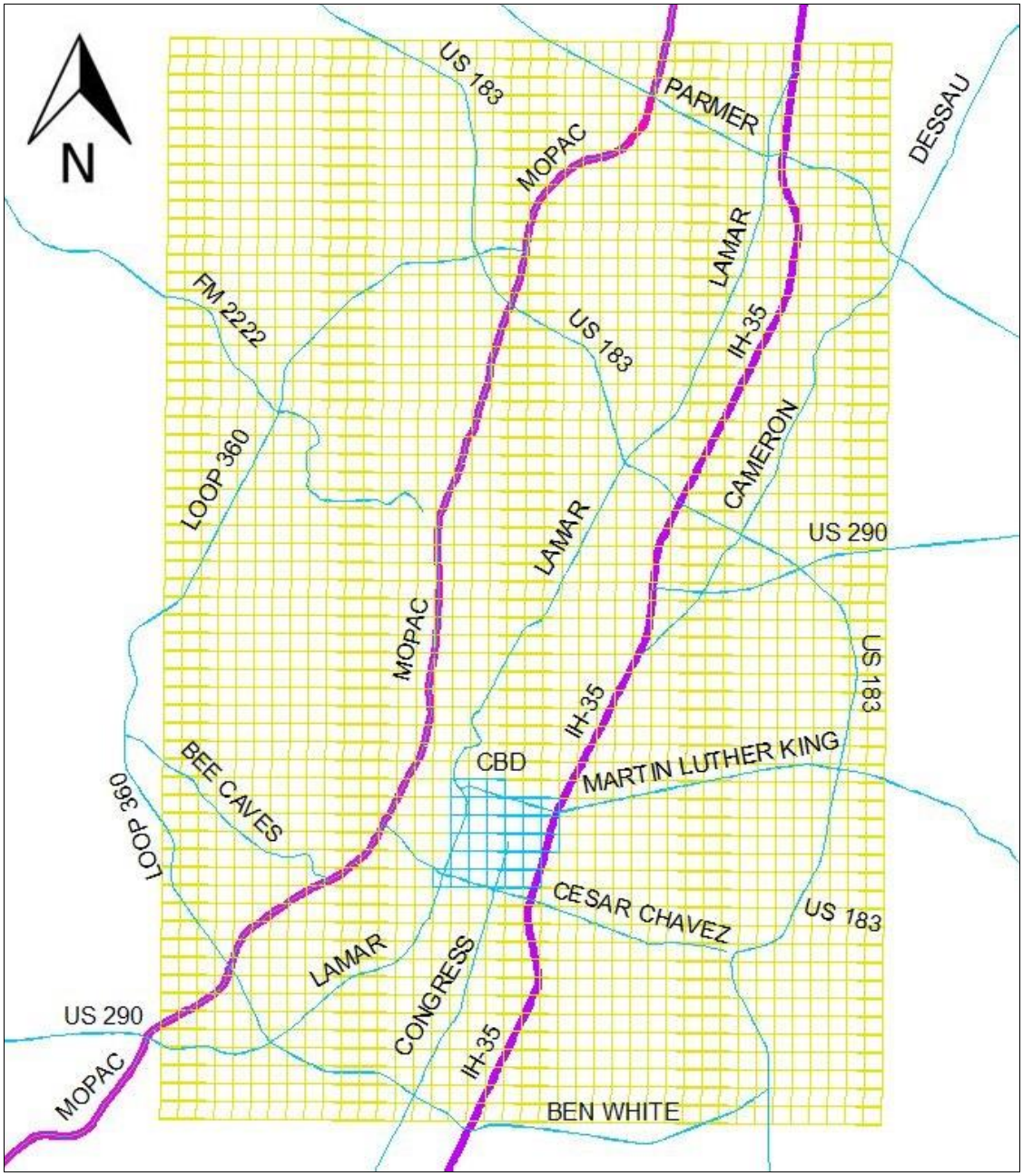


Figure 3.1 Highways, thoroughfares, and CBD location in the analysis area

The label "major thoroughfares" is used in the rest of this chapter to refer to both the major thoroughfares as well as the major arterials identified above. Finally, the Austin Central Business District (CBD) zone is defined as the "square" bounded by Lamar Blvd to the west, MLK Blvd to the north, IH-35 to the east, and Cesar Chavez to the south.

3.4.3 Model Specification

The variables that characterize the land-use intensity of each grid are: (1) road access measures (distance to main highways, and distance to other nearest major thoroughfares), (2) distance to nearest school, (3) distance to the nearest hospital, (4) fraction of grid area that is under a floodplain (area susceptible to flooding), (5) an interaction term of proximity to road access with proximity to the floodplain (distance to nearest road divided by distance to the nearest floodplain), (6) average elevation of the grid, and (7) whether the grid is in the Austin CBD zone or not. These are the variables that are used to define the baseline utilities ($\boldsymbol{\beta}$ vector), in addition to the alternative specific constants (the constant for the undeveloped alternative is fixed to 0). All the components of $\boldsymbol{\beta}$ are considered random (and skew-normally distributed), but only a few of them result in random parameters based on the estimation results (see next section).

The different weight matrix specifications that were tested include (1) a contiguity specification that generates spatial dependence based on whether or not two grids are contiguous, (2) another contiguity specification but based on shared boundary length, (3) the inverse of a continuous distance specification where the distance is measured as the Euclidean distance (crow fly distance) from the centroids of each grid, (4) the inverse of the square of the continuous distance specification, and (5) the inverse of the root of the continuous distance specification. For the last three continuous distance-based specifications, we also explored alternative distance bands to select the pairs of observations for inclusion in the composite marginal likelihood (CML) estimation. This distance band determination may be based on minimizing the trace of the variance matrix of parameters given by $tr[\mathbf{V}_{CML}(\hat{\boldsymbol{\theta}})]$. Our results did not show substantial variations in the trace value for different distance bands (regardless of the specific continuous functional form used to represent the distance separation and the variable specification used), though the best estimator efficiency was obtained at about 0.25 miles for all the three continuous distance specifications formulations and all variable specifications we attempted. Further, the results indicated that for all variable specifications, the best spatial weight matrix specification

was consistently the inverse of the continuous distance specification with the 0.25 mile distance band. This determination was based on the composite likelihood information criterion (CLIC) statistic, which may be used to compare the data fit of non-nested formulations (see Varin and Vidoni, 2005). This CLIC statistic takes the form shown below:

$$\text{CLIC} = \log L_{CML}(\hat{\boldsymbol{\theta}}) - tr\left[\hat{\mathbf{J}}(\hat{\boldsymbol{\theta}})\hat{\mathbf{H}}(\hat{\boldsymbol{\theta}})^{-1}\right], \tag{3.29}$$

where $\hat{\boldsymbol{\theta}}$ is the estimated model parameter vector, and $\hat{\mathbf{J}}(\hat{\boldsymbol{\theta}})$ and $\hat{\mathbf{H}}(\hat{\boldsymbol{\theta}})$ are the "vegetable" and "bread" matrices used in the estimation of the asymptotic variance matrix $\mathbf{V}_{CML}(\hat{\boldsymbol{\theta}})$ (see Bhat, 2011 for details of how these matrices may be estimated in a spatial context). In the current context, the weight specification that provides the highest value of the CLIC statistic is preferred over the other competing weight specifications.

As we discussed earlier, the covariance matrix of the random coefficients (and kernel error term, which are represented as the randomness of the alternative specific constants) must be parameterized during the estimation. However, only the unparamaterized coefficients are reported in the result analysis.

3.4.4 Estimation Results

Estimation results corresponding to our proposed SSN-MDC model can be found in Table 3.3a. We have listed as "columns" the three land-use categories ("undeveloped" is the base) and each row represents the effect of one explanatory variable. For those coefficients that are found to be random, we have included also the corresponding skew parameter. The first row of variables in Table 3.3a corresponds to the alternative specific constants for each land-use alternative. These constant terms do not have any substantive interpretations, and simply represent adjustments to the baseline utilities of alternatives after accommodating the other variables in the model. The results in Table 3.3a show that grids in the proximity of MoPac are more likely to be invested in commercial and residential land uses and less likely to be invested in industrial land use relative to being in an undeveloped state On the other hand, grids that are close to IH-35 have, on average, a higher propensity of being invested in commercial and industrial land use than residential land use. The rest of the baseline utility parameters can be interpreted following the same logic. The detailed explanation of the effects of each explanatory variables can be found in Bhat *et al.* (2015).

The satiation parameter corresponding to residential land-use is notably larger than the satiation parameters corresponding to commercial and industrial land-use. This indicates that there is investment in each of the commercial, industrial, and residential land uses, but the residential land-use investment intensity in a grid generally exceed that of the commercial and industrial land use intensities.

The spatial dependency parameters are considerable in magnitude and highly statistically significant, supporting the hypothesis that land-use investments in grids located in close proximity of each other are indeed positively correlated.

The skew parameters came out to be statistically significant, except for the coefficient associated to the high elevation indicator (specific to Commercial) and the coefficients associated to the distance to nearest thoroughfare/ distance to floodplain (specific to Industrial and Residential). It is important to notice that the skew parameters related to all the alternative specific constants are significant. Thus, the hypothesis of the kernel error term being normal (as has been the norm in all previous studies) is soundly rejected. The difference in shape between the implied skew-normal distribution and the normal distribution make substantial differences in the impacts of specific policy actions, as we study in detail in the next section. In particular, using the normal distribution when the skew-normal is the appropriate distribution leads to an underestimation in the land-use changes due to different scenarios. This is particularly so when the skew-normal distribution is combined with spatial dependence effects, because consideration of spatial dependence leads to a "spillover" or "multiplier" effect.

The estimated variance-covariance structure among the random parameters (the $\mathbf{\Omega}$ matrix) is presented in Table 3.3b. The variance term (*i.e.*, the diagonal element) indicates a high variance in the baseline utility of industrial land-use. There is also a significant and high covariance between the baseline utilities for commercial and industrial land use, indicating the presence of common unobserved grid-specific factors that increase (or decrease) the propensity of a grid to be invested in these two land-uses. A similar positive and significant covariance exists in the investments in residential and commercial land-uses, and between industrial and residential baseline utilities.

Table 3.3a. Estimation results of the SSN-MDC model

Variables	Land-use alternatives (base is Undeveloped)					
	Commercial		Industrial		Residential	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
Baseline utility parameters						
Alternative specific constant	-0.390	-1.08	1.275	2.41	-1.816	-1.82
Skew parameter	-0.124	-6.12	-0.182	-5.53	-0.450	-4.18
Distance to MoPac (miles)	-0.072	-4.23	0.165	2.76	-0.068	-4.03
Distance to IH-35 (miles)	-0.120	-4.16	-0.374	-4.99	0.085	3.74
Skew parameter					-0.239	-6.24
Distance to US-183 (miles)	—[+]	—	-0.257	-6.15	—	—
Distance to nearest thoroughfare (miles)	-0.398	-2.31	-1.276	-2.96	0.276	4.15
Skew parameter			0.167	8.10		
Distance to school (miles)	-0.215	-3.78	0.540	3.14	-0.462	-6.81
Distance to hospital (miles)	-0.261	-5.80	0.198	2.84	0.041	1.78
Fraction of grid area under floodplain	-0.018	-8.16	-0.025	-4.76	-0.012	-8.64
Distance to nearest thoroughfare/Distance to floodplain	-0.411	-7.99	-0.396	-3.35	0.107	4.63
Skew parameter	0.281	5.32	0.0	(fixed[*])	0.0	(fixed[*])
High elevation indicator	-0.272	-5.16	-1.326	-6.09	0.217	3.55
Skew parameter	0.0	(fixed[*])				
CBD indicator	—	—	-0.968	-2.73	-0.813	-5.00
Satiation parameters	8.751	17.43	3.495	9.63	39.64	12.39
Spatial lag parameters	0.298	2.15	0.614	2.03	0.458	3.41

[+] A "—"entry in the table indicate that the variable is not statistically significant

[*] Fixed because the parameter was not significantly different from zero at not even a 20% level of confidence

Note: The covariance matrix is presented in the next table

Table 3.3b: Variance-covariance matrix (Ω) estimates (t-statistics in parenthesis)

		Alternative specific constant			**Distance to IH-35**	**Distance to nearest thoroughfare**	**Distance to nearest thoroughfare/Distance to floodplain**			**High elevation indicator**
		Commercial	Industrial	Residential	Residential	Industrial	Commercial	Industrial	Residential	Commercial
Alternative specific constant	Commercial	1.00 (fixed)	1.426 (4.12)	0.267 (2.53)	0.00*	0.00*	0.00*	0.00*	0.00*	0.00*
	Industrial		4.176 (4.95)	0.228 (3.06)	0.00*	0.00*	0.00*	0.00*	0.00*	0.00*
	Residential			0.625 (5.02)	0.00*	0.00*	0.00*	0.00*	0.00*	0.00*
Distance to IH-35	Residential				0.278 (4.27)	0.00*	0.00*	0.00*	0.00*	0.00*
Distance to nearest thoroughfare	Industrial					2.169 (3.10)	0.00*	0.00*	0.00*	0.00*
Distance to nearest thoroughfare/Distance to floodplain	Commercial						0.253 (2.88)	0.00*	0.00*	0.00*
	Industrial							0.307 (2.76)	0.00*	0.00*
	Residential								0.188 (4.11)	0.00*
High elevation indicator	Commercial									1.108 (5.29)

* Fixed because the parameter was not significantly different from zero at not even a 20% level of confidence

‘

3.4.5 Measures of Data Fit

In this section, we examine the data fit of the proposed spatial skew-normal MDC (SSN-MDC) model with its more restricted versions: (1) the simple aspatial MDCP (A-MDCP) model, (2) the spatial MDCP (S-MDCP) model, and (3) the aspatial skew-normal MDC (ASN-MDC) model. As discussed in Section 3.4.2, all of these four models are restrictive versions of the SSN-MDC model, and may be tested against the SSN-MDC model using the adjusted composite likelihood ratio test (ADCLRT) statistic, which is asymptotically chi-squared distributed similar to the likelihood ratio test statistic for the maximum likelihood approach. The reader is referred to Bhat (2011) for details regarding the ADCLRT test statistic. Table 3.4a shows the ADLCRT results. The fifth row compares the SSN-MDC model with its restrictive versions using ADCLRT tests, and indicates the clear superior performance of the SSN-MDC model relative to other models.

In addition to testing the models using the ADCLRT statistic, we also predict the percentage of grids that invest in the different land-use alternatives (Table 3.4b). For these predictions, we use procedure discussed in Bhat *et al.* (2015). We can predict the shares for each land-use category with each model and compare the predicted shares with the actual sample shares using the Mean Absolute Percentage Error (MAPE) measure. Shares incorporate both the discrete and continuous elements of the MDC framework, because the share of a specific land-use type in a grid can decrease because that land-use type is not invested in anymore or because of a decrease in the intensity of investment in that land-use type even if it is still invested in. MAPE computations confirm that the SSN-MDC model is superior to its restrictive versions.

3.4.6 Policy Implications

Bhat *et al.* (2015) proved that the S-MDCP model could be used to compute the elasticity effects of explanatory variables (indicating the magnitude and direction of variable effects on acreage in each land use category). Due to positive spatial autocorrelation parameters, the S-MDCP model is able to capture how a change in a variable for one grid directly influences the land use of that grid, as well as the land use of neighboring grids. Bhat *et al.* (2015) used a prediction procedure to test this hypothesis and noticed that the A-MDCP model underestimates several of the elasticities. Our proposed SSN-MDC model can also be used to compute elasticities, in fact, the procedure is similar to Bhat *et al.*'s, with the only difference is that we draw $\overline{\boldsymbol{\psi}}$ from a SMVN distribution instead of MVN.

The elasticity computations are summarized in Table 3.5. Results show the difference in terms of prediction of the different models. For example, the second numerical row of Table 3.5 shows that a 25% increase in distance to IH35 (for a random grid) induces an elasticity for the residential land-use alternative of 3.79 when the A-MDCP model is applied, but the same elasticity corresponds to 13.02 when the S-MDCP model is used. Using our proposed SSN-MDC model for the computations, the elasticity of the distance to IH35 variable on residential land-use investment is 17.15. That is, the A-MDCP predicts an elasticity of only 22% (=100*3.79/17.15) of the corrected elasticity. The remaining 78% is attributable to two factors: spatial dependency and presence of skewness. 69% (=100*(13.02-3.79)/(17.15-3.79)) of the remaining is attributable to spatial dependency and 31% is attributable to the skewness. If square-footage prices data were available, we would be able to compute the marginal willingness to pay for being one mile closer to IH-35. With a traditional A-MDCP model we would be estimating only 22% of the actual willingness to pay. The same analysis can be made to study the changes of the land-use investments due to changes in any of the independent variables. Another important application of our model would be an urban growth analysis (see for example Deng and Srinivasan, 2016, Cao and Chatman, 2016, and Liu *et al.*, 2017). All the inside alternatives in our specification can be aggregated in the "developed area" of each grid. Therefore, the urban development of the study area can also be described by our model specification.

Table 3.4a. Measures of fit

Summary Statistic	Model			
	SSN-MDC[1]	**ASN-MDC[2]**	**S-MDCP[3]**	**A-MDCP[4]**
Number of observations	2,383			
Composite log-likelihood at convergence of the naïve model	-138,587.10			
Predictive log-likelihood at convergence	-76,239.87	-76,243.32	-76,255.11	-76,280.08
Number of parameters	49	46	43	40
Adjusted composite likelihood ratio test (ADCLRT) between SSN-MDC model and the corresponding model (at any reasonable level of significance)	Not applicable	$[-2*(LL_{SSN-MDC} - LL_{ASN-MDC})]=6.9>$ Chi-Squared statistics with 3 degrees of freedom	$[-2*(LL_{SSN-MDC} - LL_{S-MDCP})]=30.5 >$ Chi-Squared statistics with 6 degrees of freedom	$[-2*(LL_{SSN-MDC} - LL_{A-MDCP})]=80.4 >$ Chi-Squared statistics with 9 degrees of freedom
Spatial lag parameters (t-stat)				
Commercial land-use	0.298 (2.15)	0.0 (fixed)	0.300 (2.36)	0.0 (fixed)
Industrial land-use	0.614 (2.03)	0.0 (fixed)	0.623 (2.09)	0.0 (fixed)
Residential land-use	0.458 (3.41)	0.0 (fixed)	0.477 (4.95)	0.0 (fixed)
Skew parameters (t-stat)				
Commercial land-use specific constant	-0.124 (-6.12)	-0.148 (-7.18)	0.0 (fixed)	0.0 (fixed)
Industrial land-use specific constant	-0.182 (-5.53)	-0.251 (-5.76)	0.0 (fixed)	0.0 (fixed)
Residential land-use specific constant	-0.450 (-4.18)	-0.477 (-4.87)	0.0 (fixed)	0.0 (fixed)
Distance to IH-35 *specific to Residential*	-0.239 (-6.24)	-0.272 (-6.17)	0.0 (fixed)	0.0 (fixed)
Distance to nearest thoroughfare *specific to Industrial*	0.167 (8.10)	0.192 (7.53)	0.0 (fixed)	0.0 (fixed)
Distance to nearest thoroughfare/ distance to floodplain *specific to Commercial*	0.281 (5.32)	0.253 (5.14)	0.0 (fixed)	0.0 (fixed)
Distance to nearest thoroughfare/ distance to floodplain *specific to Industrial*	0.0 (fixed*)	0.0 (fixed*)	0.0 (fixed)	0.0 (fixed)
Distance to nearest thoroughfare/ distance to floodplain *specific to Residential*	0.0 (fixed*)	0.0 (fixed*)	0.0 (fixed)	0.0 (fixed)
High elevation indicator *specific to Commercial*	0.0 (fixed*)	0.0 (fixed*)	0.0 (fixed)	0.0 (fixed)

*: Fixed because the parameter was not significantly different from zero at not even a 20% level of confidence

[1] SSN-MDC: Spatial skew-normal MDC model

[2] ASN-MDC: Aspatial skew-normal MDC model

[3] S-MDCP: Spatial MDCP model

[4] A-MDCP: Aspatial MDCP model

Table 3.4b. Measures of fit – Predicted shares

Summary Statistic			Model							
			SSN-MDC[1]		ASN-MDC[2]		S-MDCP[3]		A-MDCP[4]	
Percentage of grids predicted to invest in…	Actual sample		Predicted %	Predicted average Investment	Predicted %	Predicted average Investment	Predicted %	Predicted average Investment	Predicted %	Predicted average Investment
	Grids (%)	Average investment (sq mi)								
Commercial	54.7	0.0136	58.5	0.0144	59.8	0.0154	60.1	0.0157	62.8	0.0160
Industrial	24.3	0.0134	28.1	0.0148	29.1	0.0161	29.5	0.0163	31.3	0.0169
Residential	82.0	0.0267	78.2	0.0235	76.8	0.0223	76.3	0.0219	75.6	0.0215
Mean absolute percentage error (MAPE)			9.1	9.4	11.8	16.6	12.7	18.4	17.1	21.1
Percentage of grids predicted to invest in…	Actual sample Grids (%)		Predicted percentage		Predicted percentage		Predicted percentage		Predicted percentage	
Commercial but not Residential	8.2		8.1		8.0		7.9		7.9	
Residential but not Commercial	37.7		36.3		36.0		35.9		35.6	
Both Commercial and Residential	51.7		51.4		51.2		51.0		50.8	
Neither Commercial nor Residential	2.4		4.2		4.8		5.2		5.7	
Mean absolute percentage error (MAPE)			20.1		27.0		31.6		37.1	

[1] SSN-MDC: Spatial skew-normal MDC model

[2] ASN-MDC: Aspatial skew-normal MDC model

[3] S-MDCP: Spatial MDCP model

[4] A-MDCP: Aspatial MDCP model

Table 3.5. Aggregate level elasticity effects of the A-MDCP, S-MDCP and SSN-MDC models (standard-error in parenthesis)

Scenario	Commercial			Industrial			Residential			Undeveloped		
	A-MDCP	**S-MDCP**	**SSN-MDC**	**A-MDCP**	**S-MDCP**	**SSN-MDC**	**A-MDCP**	**S-MDCP**	**SSN-MDC**	**A-MDCP**	**S-MDCP**	**SSN-MDC**
A 25% increase in distance to MoPac	-4.92 (1.82)	-9.87 (1.68)	-12.34 (2.75)	6.95 (1.28)	19.86 (5.24)	20.16 (5.41)	-5.91 (2.09)	-8.34 (2.36)	-9.17 (2.74)	-0.11 (0.38)	0.08 (0.32)	0.05 (0.16)
A 25% increase in distance to IH35	-6.80 (4.17)	-0.15 (5.47)	-0.78 (3.79)	-7.26 (1.37)	-15.40 (5.23)	-18.13 (6.12)	3.79 (2.76)	13.02 (2.79)	17.15 (2.75)	0.65 (0.58)	0.14 (1.10)	0.19 (0.23)
A 25% increase in distance to US-183	2.86 (1.15)	6.13 (3.21)	7.45 (4.68)	-5.78 (1.75)	-16.47 (4.11)	-19.89 (3.75)	2.40 (0.83)	3.29 (1.81)	3.75 (1.63)	0.42 (0.32)	0.01 (0.59)	0.06 (0.16)
A 25% increase in distance to nearest thoroughfare	-2.15 (0.38)	-3.09 (5.44)	-3.18 (5.12)	5.64 (1.19)	-7.84 (6.80)	-5.12 (5.38)	4.37 (1.03)	9.05 (1.29)	10.12 (1.35)	1.03 (0.86)	0.07 (0.64)	0.08 (0.53)
A 25% increase in distance to nearest school	-1.69 (1.04)	-9.57 (1.58)	-10.17 (1.86)	5.67 (2.16)	8.61 (2.15)	10.15 (2.48)	-3.64 (1.68)	-15.12 (3.24)	-16.38 (3.11)	0.30 (0.23)	0.34 (0.61)	0.39 (0.27)
A 25% increase in distance to nearest hospital	-9.19 (1.84)	-11.30 (1.80)	-12.89 (2.11)	9.68 (3.14)	23.45 (5.44)	27.18 (6.31)	4.64 (3.22)	0.79 (0.89)	1.13 (0.67)	-0.56 (0.47)	-0.05 (0.21)	-0.10 (0.24)
A 25% increase in fraction of grid area under floodplain	-2.73 (1.06)	-3.20 (1.11)	-3.48 (1.07)	-1.79 (0.93)	-2.76 (1.16)	-2.85 (0.96)	-1.13 (0.61)	-5.66 (1.83)	-7.38 (2.00)	0.61 (0.07)	0.57 (0.82)	0.52 (0.08)
A 25% increase in distance to nearest thoroughfare and a 25% decrease in distance to floodplain	-4.31 (1.43)	-6.81 (3.50)	-6.92 (2.87)	-1.56 (0.49)	-6.05 (3.32)	-6.86 (2.11)	6.27 (0.99)	12.41 (1.35)	15.32 (1.57)	-0.34 (0.86)	-1.34 (1.11)	-1.59 (0.82)
A switch of the grid location from lower elevation to higher elevation	32.44 (4.46)	46.93 (21.34)	51.73 (19.25)	-49.26 (2.65)	-43.21 (13.04)	-40.12 (9.53)	41.77 (3.78)	41.80 (22.78)	41.09 (6.32)	0.36 (0.52)	0.39 (1.73)	0.47 (0.62)
A switch of the grid location from non CBD zone to CBD zone	13.40 (2.78)	28.16 (22.78)	33.41 (24.63)	-53.39 (3.55)	-70.84 (18.45)	-75.08 (16.24)	-42.91 (3.09)	-47.88 (27.85)	-52.84 (19.17)	1.56 (1.27)	7.06 (2.97)	9.16 (1.70)

3.5 CONCLUSIONS

This chapter proposes a new spatial skew-normal multiple discrete-continuous (or SSN-MDC) model and an associated estimation method. While the use of an incorrect kernel distribution in aspatial models will, in general, lead to inconsistent estimates of the choice probabilities as well as the effects of exogenous variables, the situation gets exacerbated in spatial models because of the multiplier effect. To our knowledge, this is the first time a flexible and parametric skew-normal distribution for the kernel error term and/or random response coefficients has been used in both spatial- and aspatial-MDC models. The resulting model is estimated by using Bhat's (2011) maximum approximate composite marginal likelihood (MACML) inference approach. Simulation exercises are undertaken to examine the ability of this estimation method to recover parameters from finite samples.

Our sophisticated modeling framework can be applied to any MDC context that needs to consider spatial issues and a-not-so-restrictive distribution for unobserved heterogeneity. As an empirical demonstration, the proposed approach is applied to land-use-change decisions using the city of Austin's parcel-level land-use data. The results highlight the importance of introducing social dependence effects and non-normal kernel error terms from a policy standpoint. The empirical results of our empirical application provide important insights regarding land-use investment in multiple types of land-uses simultaneously. The variables that characterize the land-use intensity of each grid are: (1) road access measures (distance to main highways, and distance to other nearest major thoroughfares), (2) distance to nearest school, (3) distance to the nearest hospital, (4) fraction of grid area that is under a floodplain (area susceptible to flooding), (5) an interaction term of proximity to road access with proximity to the floodplain (distance to nearest road divided by distance to the nearest floodplain), (6) average elevation of the grid, and (7) whether the grid is in the Austin CBD zone or not.

Due to positive spatial autocorrelation parameters, our model is able to capture how a change in a variable for one grid directly influences the land use of that grid, as well as the land use of neighboring grids. We also compare the results of our proposed model with other more restrictive specifications that ignore the spatial dependency and/or unobserved heterogeneity. The results should indicate the superiority, in terms of data fit, of the spatial skew-normal MDC model relative to its restrictive variants.

CHAPTER 4: Incorporating a MDC Outcome in the Estimation of Joint Mixed Models

The material in this chapter is drawn substantially from the following published paper:

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.

The purpose of this chapter is to propose a new econometric approach for the estimation of joint mixed models that include an MDC outcome. Section 4.1 motivates the need for an integrated model system capable to include an MDC outcome. Section 4.2 presents the methodological framework to formulate and estimate a joint mixed model with an MDC outcome and a nominal discrete outcome, in addition to count, binary/ordinal outcomes and continuous outcomes. Section 4.3 describes an empirical application of the proposed model in which we analyze residential location choice, household vehicle ownership choice, as well as time-use choices, and investigate the extent of association versus causality in the effects of residential density on activity participation and mobility choices. Finally, Section 4.4 summarizes the main findings of this chapter.

4.1 JOINT MIXED MODELS

The joint modeling of multiple outcomes is of substantial interest in several fields. In econometric terminology, this jointness may arise because of the impact (on the multiple choice outcomes) of common underlying exogenous observed variables, or common underlying exogenous unobserved variables, or a combination of the two. For instance, consider the choice of residential location, motorized vehicle ownership (or simply auto ownership from hereon), and activity time-use in recreational pursuits (such as going to the movies/opera, going to the gym, playing sports, and camping). In this setting, it is possible (if not very likely) that individuals from households who have a high green lifestyle propensity (an unobserved variable)

may search for locations that are relatively dense (with good non-motorized and public transportation facilities and high accessibility to activity locations), may own fewer cars, may travel less and so pursue more in-home (IH) activities, and pursue less of what they may perceive as activities that correlate with extravagant living and indulgence such as out-of-home (OH) personal care/grooming, shopping, and dining out. In this case, when one or more unobserved factors (for example, green lifestyle) affect(s) the multiple outcomes, independently modeling the outcomes results in the inefficient estimation of covariate effects for each outcome (because such an approach fails to borrow information on other outcomes; see Teixeira-Pinto and Harezlak, 2013). But, more importantly, if some of the endogenous outcomes are used to explain other endogenous outcomes (such as examining the effect of density of residence on auto ownership, or the effect of density of residence on OH activity time-use, or the effect of auto ownership on time-use in activities), and if the outcomes are not modeled jointly in the presence of unobserved exogenous variable effects, the result is inconsistent estimation of the effects of one endogenous outcome on another (see Bhat and Guo, 2007, and Mokhtarian and Cao, 2008). In the next section, we position the current research within this broader methodological context of modeling multiple outcomes jointly.

The joint modeling of multiple outcomes has been a subject of interest for many years, dominated by the joint modeling of multiple continuous outcomes (see de Leon and Chough, 2013). However, in many cases, the outcomes of interest are not all continuous, and will be non-commensurate (that is, a mix of continuous, count, and discrete variables). The joint modeling of non-commensurate outcomes makes things more difficult because of the absence of a convenient multivariate distribution to jointly (and directly) represent the relationship between discrete and continuous outcomes. This is particularly the case when one of the dependent outcomes is of a multiple discrete-continuous (MDC) nature. An outcome is said to be of the MDC type if it exists in multiple states that can be jointly consumed to different continuous amounts. In the example presented in the earlier paragraph, activity time-use is an MDC variable, assuming a daily or weekly or monthly period of observation. Thus, in a given day, an individual may participate in multiple types of non-work activities (shopping, personal business, child-care, recreation, and so on) and invest different amounts of time in each activity types (see Bhat *et al.*, 2009 and Pinjari and Bhat, 2014 for detailed reviews of MDC contexts).

In this chapter, we introduce a joint mixed model that includes an MDC outcome and a

nominal discrete outcome, in addition to count, ordinal, and continuous outcomes. Each non-continuous outcome is cast in the form of a latent underlying variable regression, wherein the latent "dependent" stochastic variable is assumed to manifest itself through an *a priori* transformation rule in the observed non-continuous outcomes. Next, the continuous observed outcome and the latent continuous manifestations of the non-continuous dependent outcomes themselves are tied together using a second layer of common latent underlying unobserved decision-maker variables (such as individual lifestyle, personality, and attitudinal factors) that impact the outcomes. The presence of this second layer of latent "independent" is what generates jointness among the outcomes. Reported subjective ordinal attitudinal indicators for the latent "independent" variables help provide additional information and stability to the model system. In this manner, we build on Bhat's (2015) Generalized Heterogeneous Data Model (GHDM) that expressly acknowledges the presence of latent "independent" variables (or sometimes referred to as latent psychological constructs in the social sciences and in this research as well) affecting choice, and assumes that these latent "independent" variables get manifested in observed psychological indicators as well as the observed dependent outcomes. In particular, we develop a powerful and parsimonious way of jointly analyzing mixed outcomes including an MDC outcome. In addition, we formulate and implement a practical estimation approach for the resulting GHDM (GHDM including an MDC outcome) model using Bhat's (2011) maximum approximate composite marginal likelihood (MACML) inference approach. This approach is not simulation-based (see Bhat, 2000 and Bhat, 2001 for such simulation approaches, but which can lead to convergence issues as well as be computationally intensive). Rather, the MACML approach requires only the evaluation of bivariate or univariate cumulative normal distribution functions regardless of the number of latent variables or the number and type of dependent variable outcomes. Many structural equation models (SEMs) and similar models in the past, on the other hand, are estimated using simulation-based methods or, alternatively, sequential estimation methods (see Temme *et al.*, 2008 and Katsikatsou *et al.*, 2012 for discussions of these sequential methods). The problem with the latter sequential methods is that they do not account for sampling variability induced in earlier steps in the later steps, leading to inefficient estimation. In addition, the use of such sequential methods will, in general, also lead to inconsistent estimation (see Daziano and Bolduc, 2013 for discussions of the reasons). The MACML approach is a practical way to obtain consistent estimators even in high dimensional

mixed multivariate model systems.

To our knowledge, this is the first formulation and application of such an integrated model system in the econometric and statistical literature. The model should be applicable in a wide variety of fields where MDC variables appear as elements of package choices of different types of outcomes of interest. For example, in the health field, in addition to binary, count, and continuous variables related to the occurrence, frequency, and intensity, respectively, of specific health problems, it is not uncommon to obtain ordinal information on quality of life outcomes/perceptions and there may be interest in associating these variables with an MDC variable representing the type and intensity of participation in different types of physical activities and the durations in each participated physical activity. Other fields where the proposed model should be of interest include biology, developmental toxicology, finance, economics, epidemiology, and social science (see a good synthesis of potential applications of mixed models in De Leon and Chough, 2013). However, to make clear the application potential of the methodology presented here, we will further motivate the methodology with a specific application context originating in the land use-trasnportation domain, as we discuss in Section 4.3.

4.2 METHODOLOGY

For ease in notation, consider a cross-sectional model. As appropriate and convenient, we will suppress the index q for decision-makers (q=1,2,…,Q) in parts of the presentation.

4.2.1. Latent variable structural equation model

In the usual structural equation model set-up, we specify the latent "independent" variable or latent construct z_l^* (l=1,2,…,L) as a linear function of covariates:

$$z_l^* = \tilde{\boldsymbol{\alpha}}_l' \boldsymbol{w} + \eta_l, \tag{4.1}$$

where $\boldsymbol{w}$ is a $(\tilde{D} \times 1)$ vector of observed covariates (not including a constant), $\tilde{\boldsymbol{\alpha}}_l$ is a corresponding $(\tilde{D} \times 1)$ vector of coefficients, and η_l is a random error term assumed to be standard normally distributed for identification purposes (see Stapleton, 1978). Next, define the $(L \times \tilde{D})$ matrix $\tilde{\boldsymbol{\alpha}} = (\tilde{\boldsymbol{\alpha}}_1, \tilde{\boldsymbol{\alpha}}_2, ... \tilde{\boldsymbol{\alpha}}_L)'$, and the $(L \times 1)$ vectors $\boldsymbol{z}^* = (z_1^*, z_2^*, ..., z_L^*)'$ and

$\boldsymbol{\eta} = (\eta_1, \eta_2, \eta_3, \ldots, \eta_L)'$. Let $\boldsymbol{\eta} \sim MVN_L[\mathbf{0}_L, \boldsymbol{\Gamma}]$, where $\mathbf{0}_L$ is an $(L \times 1)$ column vector of zeros, and $\boldsymbol{\Gamma}$ is an $(L \times L)$ correlation matrix. In matrix form, we may write Equation (4.1) as:

$$\boldsymbol{z}^* = \tilde{\boldsymbol{\alpha}}\boldsymbol{w} + \boldsymbol{\eta}. \quad (4.2)$$

4.2.2. Latent variable measurement equation model components

Consider a combination of continuous, ordinal, count, nominal, and MDC outcomes of the underlying latent variable vector $\boldsymbol{z}^*$. Note that, in the GHMD, the actual mixed outcomes of interest ("endogenous" variables, including continuous, count, nominal, and MDC outcomes) as well as any subjective indicators (all ordinal in the current research) of the latent vector $\boldsymbol{z}^*$ are together (and simultaneously) used to estimate the structural Equation (4.2) that relates the latent constructs with exogenous covariates (through a reduced form of the measurement equation system; see Appendix B). That is, the fact that we have additional ordinal indicators of the latent constructs helps provide stability to the estimation of Equation (4.2) in the model system, but does not play a central role in identifying the latent constructs per se. In other words, there is no distinction between the traditional subjective indicators (usually ordinal) and other actual endogenous variables of interest in the GHDM. All of these indicators/outcomes together are treated identically as marker manifestations of the underlying latent construct vector $\boldsymbol{z}^*$. Thus, in the GHDM, there is even no need for any subjective indicators, since the actual endogenous outcomes themselves serve as indicators of the latent constructs. The latent constructs are identified based on theory and earlier studies, as in all earlier land use-transportation studies that incorporate latent psychological constructs in the modeling framework (please see Section 4.3.4 for a more complete discussion of this point). Once estimated, the relationship between the latent constructs and the subjective indicators can be discarded (these purely help in efficiently estimating Equation (4.2), and in identifying Equation (4.2) if the number of endogenous outcomes present are not adequate). The focus is on (a) the measurement relationship between the actual endogenous outcomes with (i) exogenous covariates, (ii) other actual endogenous outcomes, and (iii) the latent constructs, and (b) the structural equation system of Equation (4.2). In the former relationship, the inter-relationships among the endogenous variables are "uncorrupted causal" influences after controlling for error correlations across the many

dimensions (engendered by the latent effects). These endogenous effects correspond to recursive influences among the dependent variable outcomes.[18]

In the following presentation, we will use the term "outcome" to refer to both the actual endogenous outcomes of interest as well as subjective ordinal indicators of the latent constructs. We also allow more than one outcome for the continuous and ordinal variable types, but confine attention to only one outcome each for the count, nominal and MDC variable types. This is purely for ease in presentation, and is by no means methodologically restrictive. Indeed, the extension to more than one count, and/or one nominal and/or one MDC outcome is straightforward.

Let there be H continuous outcomes $(y_1, y_2, ..., y_H)$ with an associated index h $(h = 1, 2, ..., H)$. Let $y_h = \boldsymbol{\gamma}_h' \boldsymbol{x} + \boldsymbol{d}_h' \boldsymbol{z}^* + \varepsilon_h$ in the usual linear regression fashion, where $\boldsymbol{x}$ is an $(A \times 1)$-vector of exogenous variables (including a constant) as well as the observed values of other endogenous outcomes. $\boldsymbol{\gamma}_h$ is the corresponding compatible coefficient vector. $\boldsymbol{d}_h$ is an $(L \times 1)$ vector of latent variable loadings on the h^{th} continuous outcome, and ε_h is a normally distributed measurement error term. Define the following two $(H \times 1)$ vectors: $\boldsymbol{y} = (y_1, y_2, ..., y_H)'$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_H)'$, with $\boldsymbol{\varepsilon} \sim MVN_H(\boldsymbol{0}_H, \boldsymbol{\Sigma})$ (that is, the vector $\boldsymbol{\varepsilon}$ is assumed to be H-variate normally distributed with zero means for all its elements and a covariance matrix $\boldsymbol{\Sigma}$). $\boldsymbol{\Sigma}$ is restricted to be diagonal to aid in identification because the latent variable vector $\boldsymbol{z}^*$ already serves as a vehicle to generate covariance between the outcome variables. Define the $(H \times A)$ matrix $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, ..., \boldsymbol{\gamma}_H)'$ and the $(H \times L)$ matrix of latent variable loadings $\boldsymbol{d} = (\boldsymbol{d}_1, \boldsymbol{d}_2, ..., \boldsymbol{d}_H)'$. Then, one may write the following vector measurement equation for the continuous outcomes:

$$\boldsymbol{y} = \boldsymbol{\gamma} \boldsymbol{x} + \boldsymbol{d} \boldsymbol{z}^* + \boldsymbol{\varepsilon}. \tag{4.3}$$

Next, let there be N ordinal outcomes (indicator variables in this research) for the individual, and let n be the index for the ordinal outcomes $(n = 1, 2, ..., N)$. Also, let J_n be the

[18] 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 current research that has discrete dependent, count, and MDC variables, the structural effects of one limited-dependent variable on another can only be in a single direction. See Maddala, (1983) and Bhat (2015) for a more detailed explanation.

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 $\tilde{y}_n^*$ be the latent underlying variable whose horizontal partitioning leads to the observed outcome for the n^{th} ordinal variable. Assume that the individual under consideration chooses the a_n^{th} ordinal category. Then, in the usual ordered response formulation, we may write:

$$\tilde{y}_n^* = \tilde{\boldsymbol{\gamma}}_n' \boldsymbol{x} + \tilde{\boldsymbol{d}}_n' \boldsymbol{z}^* + \tilde{\varepsilon}_n, \text{ and } \tilde{\psi}_{n,a_n-1} < \tilde{y}_n^* < \tilde{\psi}_{n,a_n}, \tag{4.4}$$

where $\boldsymbol{x}$ is as defined earlier, $\tilde{\boldsymbol{\gamma}}_n$ is a corresponding vector of coefficients to be estimated, $\tilde{\boldsymbol{d}}_n$ is an $(L \times 1)$ vector of latent variable loadings on the n^{th} continuous outcome, the $\tilde{\psi}$ terms represent thresholds (for each n, $\tilde{\psi}_{n,0} < \tilde{\psi}_{n,1} < \tilde{\psi}_{n,2} ... < \tilde{\psi}_{n,J_n-1} < \tilde{\psi}_{n,J_n}$; $\tilde{\psi}_{n,0} = -\infty, \tilde{\psi}_{n,1} = 0$, and $\tilde{\psi}_{n,J_n} = +\infty$), and $\tilde{\varepsilon}_n$ is the standard normal random error for the n^{th} ordinal outcome. For later use, let $\tilde{\boldsymbol{\psi}}_n = (\tilde{\psi}_{n,2}, \tilde{\psi}_{n,3}, ..., \tilde{\psi}_{n,J_n-1})'$ and $\tilde{\boldsymbol{\psi}} = (\tilde{\boldsymbol{\psi}}_1', \tilde{\boldsymbol{\psi}}_2', ..., \tilde{\boldsymbol{\psi}}_N)'$. Stack the N underlying continuous variables $\tilde{y}_n^*$ into an $(N \times 1)$ vector $\tilde{\boldsymbol{y}}^*$, and the N error terms $\tilde{\varepsilon}_n$ into another $(N \times 1)$ vector $\tilde{\boldsymbol{\varepsilon}}$. Define $\tilde{\boldsymbol{\gamma}} = (\tilde{\boldsymbol{\gamma}}_1, \tilde{\boldsymbol{\gamma}}_2, ..., \tilde{\boldsymbol{\gamma}}_H)'$ [$(N \times A)$ matrix] and $\tilde{\boldsymbol{d}} = \left(\tilde{\boldsymbol{d}}_1, \tilde{\boldsymbol{d}}_2, ..., \tilde{\boldsymbol{d}}_N\right)$ [$(N \times L)$ matrix], and let $\mathbf{IDEN}_N$ be the identity matrix of dimension N representing the correlation matrix of $\tilde{\boldsymbol{\varepsilon}}$; $\tilde{\boldsymbol{\varepsilon}} \sim MVN_N(\mathbf{0}_N, \mathbf{IDEN}_N)$. Finally, stack the lower thresholds for the decision-maker $\tilde{\psi}_{n,a_n-1}(n = 1, 2, ..., N)$ into an $(N \times 1)$ vector $\tilde{\boldsymbol{\psi}}_{low}$ and the upper thresholds $\tilde{\psi}_{n,a_n}(n = 1, 2, ..., N)$ into another vector $\tilde{\boldsymbol{\psi}}_{up}$. Then, in matrix form, the measurement equation for the ordinal outcomes (indicators) for the decision-maker may be written as:

$$\tilde{\boldsymbol{y}}^* = \tilde{\boldsymbol{\gamma}} \boldsymbol{x} + \tilde{\boldsymbol{d}} \boldsymbol{z}^* + \tilde{\boldsymbol{\varepsilon}}, \quad \tilde{\boldsymbol{\psi}}_{low} < \tilde{\boldsymbol{y}}^* < \tilde{\boldsymbol{\psi}}_{up}. \tag{4.5}$$

For the count variable, let the index be g for the count categories $(g = 0, 1, 2, ..., \infty)$ and let r be the actual observed count value for the household. Then, a generalized version of the negative binomial count model may be written as (see Castro, Paleti, and Bhat, or CPB, 2012b and Bhat *et al*., 2013b):

$$\breve{y}^* = \breve{\boldsymbol{d}}' \boldsymbol{z}^* + \breve{\varepsilon}, \; \breve{\psi}_{r-1} < \breve{y}^* < \breve{\psi}_r, \tag{4.6}$$

$$\breve{\psi}_r = \Phi^{-1}\left[\frac{(1-\upsilon)^{\theta}}{\Gamma(\theta)} \sum_{t=0}^{r} \left(\frac{\Gamma(\theta + t)}{t!} (\upsilon)^t \right) \right] + \varphi_r, \; \upsilon = \frac{\lambda}{\lambda + \theta}, \text{ and } \lambda = e^{\vec{\gamma}' \boldsymbol{x}}. \tag{4.7}$$

In the above equation, $\breve{y}^*$ is a latent continuous stochastic propensity variable that maps into the observed count *r* through the $\breve{\boldsymbol{\psi}}$ vector (which is a vertically stacked column vector of thresholds $(\breve{\psi}_{-1}, \breve{\psi}_0, \breve{\psi}_1, \breve{\psi}_2, ...)'$. $\breve{\boldsymbol{d}}$ is an $(L \times 1)$ vector of latent variable loadings on the count outcome, and $\breve{\varepsilon}$ is a standard normal random error term. $\breve{\boldsymbol{\gamma}}$ is a column vector corresponding to the vector $\boldsymbol{x}$ (including a constant) of exogenous observable covariates and endogenous outcomes. Φ^{-1} in the threshold function of Equation (4.7) is the inverse function of the univariate cumulative standard normal. θ is a parameter that provides flexibility to the count formulation, and is related to the dispersion parameter in a traditional negative binomial model ($\theta > 0$; if $\theta \to \infty$, the general negative binomial structure collapses to a general Poisson structure). $\Gamma(\theta)$ is the traditional gamma function; $\Gamma(\theta) = \int_{\tilde{t}=0}^{\infty} \tilde{t}^{\theta-1} e^{-\tilde{t}}\, d\tilde{t}$. The threshold terms in the $\breve{\boldsymbol{\psi}}$ vector satisfy the ordering condition (*i.e.*, $\breve{\psi}_{-1} < \breve{\psi}_0 < \breve{\psi}_1 < \breve{\psi}_2 .... < \infty$) as long as $\varphi_{-1} < \varphi_0 < \varphi_1 < \varphi_2 .... < \infty$. The presence of the φ terms in the thresholds provides substantial flexibility to accommodate high or low probability masses for specific count outcomes (see CPB, 2012b for a detailed discussion). For identification, set $\varphi_{-1} = -\infty$ and $\varphi_0 = 0$. In addition, we identify a count value e^* ($e^* \in \{0, 1, 2, ...\}$) above which φ_g ($g \in \{1, 2, ...\}$) is held fixed at φ_{e^*}; that is, $\varphi_g = \varphi_{e^*}$ if $g > e^*$, where the value of e^* can be based on empirical testing. Doing so is the key to allowing the count model to predict beyond the count range available in the estimation sample. For later use, let $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \ldots \varphi_{e^*})'$ ($e^* \times 1$ vector) (assuming $e^* > 0$).

Next, consider the nominal (unordered-response) outcome for the individual, and let *i* be the corresponding index (*i* = 1, 2, 3,…, *I*). Let the individual under consideration choose the alternative m. Also, assume the usual random utility structure for each alternative *i*.

$$U_i = \tilde{\boldsymbol{b}}_i' \boldsymbol{x} + \vartheta_i'(\boldsymbol{\beta}_i \boldsymbol{z}^*) + \tilde{\varsigma}_i, \tag{4.8}$$

where $\boldsymbol{x}$ is the same fixed vector of exogenous variables as earlier, $\tilde{\boldsymbol{b}}_i$ is an $(A \times 1)$ column vector of corresponding coefficients, and $\tilde{\varsigma}_i$ is a normal random error term. $\boldsymbol{\beta}_i$ is a $(N_i \times L)$ matrix of variables interacting with latent variables to influence the utility of alternative

i, and $\boldsymbol{\vartheta}_i$ is an $(N_i \times 1)$ column vector of coefficients capturing the effects of latent variables and their interaction effects with other exogenous variables. If each of the latent variables impacts the utility of the alternatives for each nominal variable purely through a constant shift in the utility function, $\boldsymbol{\beta}_i$ will be an identity matrix of size L, and each element of $\boldsymbol{\vartheta}_i$ will capture the effect of a latent variable on the constant specific to alternative *i* (see Bhat and Dubey, 2014). To move forward, let $\tilde{\boldsymbol{\varsigma}} = (\tilde{\varsigma}_1, \tilde{\varsigma}_2, ..., \tilde{\varsigma}_I)'$ ($I \times 1$ vector), and $\tilde{\boldsymbol{\varsigma}} \sim MVN_I(\mathbf{0}_I, \boldsymbol{\Lambda})$. Taking the difference with respect to the first alternative, only the elements of the covariance matrix $\breve{\boldsymbol{\Lambda}}$ of the covariance matrix of the error differences, $\breve{\boldsymbol{\varsigma}} = (\breve{\varsigma}_2, \breve{\varsigma}_3, ..., \breve{\varsigma}_I)$ (where $\breve{\varsigma}_i = \tilde{\varsigma}_i - \tilde{\varsigma}_1$, $i \neq 1$), is estimable.[19] Further, the variance term at the top left diagonal of $\breve{\boldsymbol{\Lambda}}$ is set to one to account for scale invariance. $\boldsymbol{\Lambda}$ is constructed from $\breve{\boldsymbol{\Lambda}}$ by adding an additional row on top and an additional column to the left. All elements of this additional row and column are filled with values of zeros. Next, define $\boldsymbol{U} = (U_1, U_2, ..., U_I)'$ ($I \times 1$ vector), $\tilde{\boldsymbol{b}} = (\tilde{\boldsymbol{b}}_1, \tilde{\boldsymbol{b}}_2, \tilde{\boldsymbol{b}}_3, ..., \tilde{\boldsymbol{b}}_I)'$ ($I \times A$ matrix), and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', ..., \boldsymbol{\beta}_I')'$ $\left(\sum_{i=1}^{I} N_i \times L\right)$ matrix. Also, define the $\left(I \times \sum_{i=1}^{I} N_i\right)$ matrix $\boldsymbol{\vartheta}$ which is initially filled with all zero values. Then, position the $(1 \times N_1)$ row vector in the first row to occupy columns 1 to N_1, position the $(1 \times N_2)$ row vector in the second row to occupy columns N_1+1 to $N_1 + N_2$, and so on until the $(1 \times N_I)$ row vector is appropriately positioned. Further, define $\tilde{\boldsymbol{\varpi}} = (\boldsymbol{\vartheta}\boldsymbol{\beta})$ ($I \times L$ matrix). Then, in matrix form, we may write:

$$\boldsymbol{U} = \tilde{\boldsymbol{b}}\boldsymbol{x} + \tilde{\boldsymbol{\varpi}}\boldsymbol{z}^* + \tilde{\boldsymbol{\varsigma}}. \tag{4.9}$$

Next, note that, under the utility maximization paradigm, $u_{im} = U_i - U_m$ must be less than zero for all $i \neq m$, since the individual chose alternative *m*. Stack the latent utility differentials into a vector $\boldsymbol{u} = \left[\left(u_{1m}, u_{2m}, ..., u_{Im}\right)'; i \neq m\right]$. To write this utility differential vector compactly in terms of the original utilities, define a matrix **M** of size $[I-1] \times [I]$. Insert an identity matrix of size

[19] Also, in MNP models, identification is tenuous when only individual-specific covariates are used in the vector $\boldsymbol{x}$ (see Keane, 1992 and Munkin and Trivedi, 2008). In particular, exclusion restrictions are needed in the form of at least one individual characteristic being excluded from each alternative's utility in addition to being excluded from a base alternative (but appearing in some other utilities). But these exclusion restrictions are not needed when there are alternative-specific variables.

$(I_1 - 1)$ after supplementing with a column of '-1' values in the column corresponding to the chosen alternative *m*. Then, we may write the following:

$$\boldsymbol{u} = \mathbf{M}\boldsymbol{U} = \mathbf{M}\tilde{\boldsymbol{b}}\boldsymbol{x} + \mathbf{M}\tilde{\boldsymbol{\varpi}}\boldsymbol{z}^* + \mathbf{M}\tilde{\boldsymbol{\varsigma}} = \boldsymbol{b}\boldsymbol{x} + \boldsymbol{\varpi}\boldsymbol{z}^* + \boldsymbol{\varsigma}, \text{ with } \boldsymbol{b} = \mathbf{M}\tilde{\boldsymbol{b}}, \boldsymbol{\varpi} = \mathbf{M}\tilde{\boldsymbol{\varpi}}, \text{ and } \boldsymbol{\varsigma} = \mathbf{M}\tilde{\boldsymbol{\varsigma}}.$$

Finally, consider the MDC outcome. Following Bhat (2005) and Bhat (2008), consider a choice scenario where the decision maker maximizes his/her time utility subject to a binding time budget constraint:

$$\max \tilde{U}(\boldsymbol{t}) = \sum_{k=1}^{K-1} \frac{\tau_k}{\alpha_k} \psi_k \left(\left(\frac{t_k}{\tau_k} + 1 \right)^{\alpha_k} - 1 \right) + \frac{1}{\alpha_K} \psi_K (t_K)^{\alpha_K} \tag{4.10}$$

$$s.t. \sum_{k=1}^{K} t_k = T,$$

where the utility function $\tilde{U}(\boldsymbol{t})$ is quasi-concave, increasing and continuously differentiable, $\boldsymbol{t}$ is the time investment vector of dimension *K*×1 with elements t_k $(t_k \geq 0)$, τ_k, α_k, and ψ_k are parameters associated with activity purpose *k*, and *T* represents the time budget to be allocated among the *K* activity purposes. The utility function form in Equation (4.10) allows corner solutions (*i.e.*, zero consumptions) for activity purposes 1 through $K-1$ through the parameters τ_k, which allow corner solutions for these alternatives while also serving the role of satiation parameters $(\tau_k > 0 : k = 1, 2, ..., K-1)$. On the other hand, the functional form for the final activity purpose ensures that some time is invested in activity purpose *K* (for example, activity purpose *K* may refer to in-home activities such as eating, watching TV, and relaxing; activity purpose *K* is usually referred to as an *essential outside good* in the microeconomics literature; see Bhat, 2008). The role of α_k is to capture satiation effects, with a smaller value of α_k implying higher satiation for activity purpose *k*. ψ_k represents the stochastic baseline marginal utility; that is, it is the marginal utility at the point of zero time investment for alternative *k*.

As we indicated in Section 1.4, the utility function in Equation (4.10) constitutes a valid utility function if, in addition to the constraints on the τ_k parameters as discussed above, $\alpha_k \leq 1$, and $\psi_k \geq 0$ for all *k*. Also, as indicated earlier, τ_k and α_k influence satiation, though in quite different ways: τ_k controls satiation by translating consumption quantity, while α_k controls satiation by exponentiating consumption quantity. Empirically speaking, it is difficult to

disentangle the effects of τ_k and α_k separately. We suggested earlier to estimate a τ - and an α - profile, and choose the profile that provides a better statistical fit.[20] However, we will retain the utility form of Equation (4.10) to keep the presentation general. Next, to complete the model structure, the baseline utility is specified to be a function of the latent variable vector, the *A*-dimensional exogenous variable vector $\boldsymbol{x}$, and a random error term as follows:

$$\psi_k = \exp(\boldsymbol{x}, \boldsymbol{z}^*, \tilde{\xi}_k) = \exp(\tilde{\boldsymbol{\delta}}_k' \boldsymbol{x} + \tilde{\boldsymbol{\mu}}_k' \boldsymbol{z}^* + \tilde{\xi}_k) \quad \text{or} \quad \overline{\psi}_k^* = \ln(\psi_k) = \tilde{\boldsymbol{\delta}}_k' \boldsymbol{x} + \tilde{\boldsymbol{\mu}}_k' \boldsymbol{z}^* + \tilde{\xi}_k, \tag{4.11}$$

where $\tilde{\boldsymbol{\delta}}_k$ and $\tilde{\boldsymbol{\mu}}_k$ are *A*-dimensional and *L*-dimensional column vectors, respectively, and $\tilde{\xi}_k$ captures the idiosyncratic characteristics that impact the baseline utility of activity purpose *k*. We assume that the error terms $\tilde{\xi}_k$ are multivariate normally distributed across alternatives: $\tilde{\boldsymbol{\xi}} = (\tilde{\xi}_1, \tilde{\xi}_2, ..., \tilde{\xi}_K)' \sim MVN_K(\boldsymbol{0}_K, \tilde{\boldsymbol{\Omega}})$. But only differences in the logarithm of the baseline utilities matter, not the actual logarithm of the baseline utility values (see section 1.4). Thus, it will be easier to work with the logarithm of the baseline utilities of the first $K-1$ alternatives, and normalize the logarithm of the baseline utility for the last alternative to zero. That is, we write:

$$\begin{aligned} \overline{\psi}_k &= \overline{\psi}_k^* - \overline{\psi}_K^* = (\tilde{\boldsymbol{\delta}}_k - \tilde{\boldsymbol{\delta}}_K)' \boldsymbol{x} + (\tilde{\boldsymbol{\mu}}_k - \tilde{\boldsymbol{\mu}}_K)' \boldsymbol{z}^* + (\tilde{\xi}_k - \tilde{\xi}_K) \\ &= \boldsymbol{\delta}_k' \boldsymbol{x} + \boldsymbol{\mu}_k' \boldsymbol{z}^* + \xi_k, \ \boldsymbol{\delta}_k = (\tilde{\boldsymbol{\delta}}_k - \tilde{\boldsymbol{\delta}}_K), \ \boldsymbol{\mu}_k = (\tilde{\boldsymbol{\mu}}_k - \tilde{\boldsymbol{\mu}}_K), \ \xi_k = (\tilde{\xi}_k - \tilde{\xi}_K) \ \forall k \neq K \end{aligned} \tag{4.12}$$

$\overline{\psi}_K = \overline{\psi}_K^* - \overline{\psi}_K^* = 0$ for $k = K$.

It should be clear from above that only the covariance matrix, say $\boldsymbol{\Omega}$ of the error differences $\xi_k = (\tilde{\xi}_k - \tilde{\xi}_K)$ is estimable, and not the covariance matrix $\tilde{\boldsymbol{\Omega}}$ of the original error terms. Further, with the formulation as in Equation (4.10), where the sum of the time investments across activity purposes is equal to the total time budget, an additional scale normalization needs to be imposed (see Section 1.4). A convenient normalization is to set the first element of $\boldsymbol{\Omega}$ (that is, $\boldsymbol{\Omega}_{11}$ to one). Further, for ease in interpretation of the covariance matrix $\boldsymbol{\Omega}$, we assume that the error term of the "outside" alternative ξ_K is independent of the error terms of the "inside" alternatives ξ_k $(k = 1, 2, ..., K-1)$. With this assumption, each covariance matrix element of $\boldsymbol{\Omega}$ can then

---

[20] The τ-profile equivalent of Equation (4.10) is $\tilde{U}(\boldsymbol{t}) = \sum_{k=1}^{K-1} \tau_k \psi_k \ln\left(\frac{t_k}{\tau_k} + 1\right) + \psi_K \ln\{t_K\}$, and the α-profile equivalent is $\tilde{U}(\boldsymbol{t}) = \sum_{k=1}^{K-1} \frac{1}{\alpha_k} \psi_k \left\{(t_k + 1)^{\alpha_k} - 1\right\} + \frac{1}{\alpha_K} \psi_K t_K^{\alpha_K}$.

immediately be interpreted as a direct indicator of the extent of variance and covariance in the utilities of the inside alternatives.[21]

The analyst can solve for the optimal consumption allocations corresponding to Equation (4.10) by forming the Lagrangian and applying the Karush-Kuhn-Tucker (KKT) conditions. The Lagrangian function for the problem, after substituting $\psi_k = \exp(\overline{\psi}_k)$ (equal to $\exp(\boldsymbol{\delta}_k'\boldsymbol{x} + \boldsymbol{\mu}_k'\boldsymbol{z}^* + \xi_k)$ for $k = 1, 2, ..., K-1$ and equal to $\exp(0) = 1$ for $k = K$) in Equation (4.10) is:

$$L = \sum_{k=1}^{K-1} \frac{\tau_k}{\alpha_k} \exp(\boldsymbol{\delta}_k'\boldsymbol{x} + \boldsymbol{\mu}_k'\boldsymbol{z}^* + \xi_k)\left(\left(\frac{t_k}{\tau_k} + 1\right)^{\alpha_k} - 1\right) + \frac{1}{\alpha_K}(t_K)^{\alpha_K} - \tilde{\lambda}\left[\sum_{k=1}^{K} t_k - T\right] \tag{4.13}$$

where $\tilde{\lambda}$ is the Lagrangian multiplier associated with the time budget constraint (that is, it can be viewed as the marginal utility of total time). The KKT first-order condition for the "optimal" investment t_K^* in the last activity purpose (which is always positive) implies the following: $\left(t_K^*\right)^{\alpha_K - 1} - \tilde{\lambda} = 0$; that is, $\tilde{\lambda} = \left(t_K^*\right)^{\alpha_K - 1}$. The KKT first-order conditions for the optimal time investments for the inside alternatives (the t_k^* values for $k = 1, 2, ..., K-1$) are given by:

$$\exp(\boldsymbol{\delta}_k'\boldsymbol{x} + \boldsymbol{\mu}_k'\boldsymbol{z}^* + \xi_k)\left(\frac{t_k^*}{\tau_k} + 1\right)^{\alpha_k - 1} - \tilde{\lambda} = 0 \text{ , if } t_k^* > 0,\ k = 1, 2, ..., K-1 \tag{4.14}$$

$$\exp(\boldsymbol{\delta}_k'\boldsymbol{x} + \boldsymbol{\mu}_k'\boldsymbol{z}^* + \xi_k)\left(\frac{t_k^*}{\tau_k} + 1\right)^{\alpha_k - 1} - \tilde{\lambda} < 0, \text{ if } t_k^* = 0,\ k = 1, 2, ..., K-1$$

Substitute $\tilde{\lambda} = \left(t_K^*\right)^{\alpha_K - 1}$ into the above equations, take logarithms, and rewrite the KKT conditions as:

---

[21] In particular, assume that the variance of ξ_K is 0.5. Then, to normalize $\boldsymbol{\Omega}_{11}$ to one, we should have that the variance of ξ_1 is also 0.5. Let the variance of ξ_k $(k = 2, 3, ..., K-1)$ be σ_k^2 and the covariance between ξ_k and $\xi_{k'}(k, k' = 1, 2, 3, ..., K-1; k \neq k')$ be $\sigma_{kk'}$. Then, the matrix $\boldsymbol{\Omega}$ of the error differences $\xi_k = (\tilde{\xi}_k - \tilde{\xi}_K)$ is:

$$\boldsymbol{\Omega} = \begin{bmatrix} 1 & 0.5+\sigma_{12} & 0.5+\sigma_{13} & \cdots & 0.5+\sigma_{1,K-1} \\ 0.5+\sigma_{12} & 0.5+\sigma_2^2 & 0.5+\sigma_{23} & \cdots & 0.5+\sigma_{2,K-1} \\ 0.5+\sigma_{13} & 0.5+\sigma_{23} & 0.5+\sigma_3^2 & \cdots & 0.5+\sigma_{3,K-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0.5+\sigma_{1,K-1} & 0.5+\sigma_{2,K-1} & 0.5+\sigma_{3,K-1} & \cdots & 0.5+\sigma_{K-1}^2 \end{bmatrix}$$

$\tilde{u}_k = V_k + \boldsymbol{\mu}_k' \mathbf{z}^* + \xi_k = 0$, if $t_k^* > 0$, $k = 1, 2, ..., K-1$ (4.15)

$\tilde{u}_k = V_k + \boldsymbol{\mu}_k' \mathbf{z}^* + \xi_k < 0$, if $t_k^* = 0$, $k = 1, 2, ..., K-1$,

where $V_k = \boldsymbol{\delta}_k' \mathbf{x} + (\alpha_k - 1) \ln\left(\frac{t_k^*}{\tau_k} + 1\right) - (\alpha_K - 1) \ln\left(t_K^*\right)$ for $k = 1, 2, ..., K-1$.

Define $\tilde{\boldsymbol{u}} = (\tilde{u}_1, \tilde{u}_2, ..., \tilde{u}_{K-1})'$ [$(K-1) \times 1$ vector], $\boldsymbol{\delta} = (\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, ..., \boldsymbol{\delta}_{K-1})'$ [$(K-1) \times A$ vector], $\boldsymbol{V} = (V_1, V_2, ..., V_{K-1})'$ [$(K-1) \times 1$ vector], $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, ..., \boldsymbol{\mu}_{K-1})'$ [$(K-1) \times A$ matrix], $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_K)'$, $\boldsymbol{\tau} = (\tau_1, \tau_2, ..., \tau_K)'$ and $\boldsymbol{\xi} = (\xi_1, \xi_2, ..., \xi_{K-1})' \sim MVN_{K-1}(\mathbf{0}_{K-1}, \boldsymbol{\Omega})$. Then, we may write, in matrix form, the following equation:

$\tilde{\boldsymbol{u}} = \boldsymbol{V} + \boldsymbol{\mu} \mathbf{z}^* + \boldsymbol{\xi}$, (4.16)

with the elements of $\tilde{\boldsymbol{u}}$ adhering to the conditions in Equation (4.15). Also, for later use, let F_C be the set of consumed alternatives not including the last alternative (with cardinality $\tilde{F}_C$), and F_{NC} be the set of non-consumed alternatives (with cardinality $\tilde{F}_{NC}$).

The parameter vector to be estimated is $\vec{\boldsymbol{\theta}} = [\text{vech}(\tilde{\boldsymbol{\alpha}}), \text{vech}(\boldsymbol{\Gamma}), \text{vech}(\boldsymbol{\gamma}), \text{vech}(\boldsymbol{d}), \text{vech}(\tilde{\boldsymbol{\gamma}}),$
$\text{vech}(\tilde{\boldsymbol{d}}), \text{vech}(\boldsymbol{\Sigma}), \text{vech}(\breve{\boldsymbol{d}}), \text{vech}(\breve{\boldsymbol{\gamma}}), \theta, \boldsymbol{\varphi}, \text{vech}(\tilde{\boldsymbol{b}}), \text{vech}(\boldsymbol{\mu}), \text{vech}(\boldsymbol{\delta}), \boldsymbol{\alpha}$ or $\boldsymbol{\tau}, \text{vech}(\boldsymbol{\Omega})]$, where vech($\boldsymbol{\Delta}$) implies a row vector of all the unique and non-fixed elements of matrix $\boldsymbol{\Delta}$. The maximum likelihood estimation of the model involves the evaluation of an $(N + I + \tilde{F}_{NC})$-dimensional rectangular integral for each decision-maker, which can be computationally expensive. So, we use the Maximum Approximate Composite Marginal Likelihood (MACML) approach of Bhat (2011). The estimation approach is very notation-intensive, and so we relegate the details of the approach to Appendix B. Also, in Appendix C, we provide a diagrammatic representation of the entire model system, including the notations used in this section for easy association.

4.3 APPLICATION TO RESIDENTIAL SELF-SELECTION EFFECT ANALYSIS

In this section, we apply the proposed model to examine households' residential location (characterized by commute distance and the density or number of households per square mile in the Census block group of the household's residence, as obtained from the 2010 decennial

Census data), auto ownership level, and time spent on a typical weekday on (a) in-home (IH) non-work, non-educational, and non-sleep activities and (b) out-of-home (OH) non-work non-educational pursuits. In the analysis, the OH activities are classified into one of six types: personal business (including family or personal obligations, going to day care, and medical appointments), shopping (including buying food and goods), eating out, social activities (including visiting friends or relatives and attending parties), recreation (including visiting cultural/arts centers, going to the movies, attending sports events, going to the gym, pursuing physical activities such as running, walking, swimming, and playing sports), and "other" activities (including picking up or dropping off someone, and "other" non-work, non-education, and non-sleep activities. A further investigation of this "other" activity category indicated that it was dominated by serve passenger activity. Specifically, 80% of the "other" activities corresponded to serve passenger activity. Hence, to make our labeling easy and comprehensible, we will refer to the "other" category as the "serve passenger" category in the rest of this section.

4.3.1 Empirical context

An issue that has received particular attention within the broad land use-transportation literature is whether any effect of the BE on travel demand is causal or merely associative (or some combination of the two; see Bhat and Guo, 2007, Mokhtarian and Cao, 2008, Pinjari *et al.*, 2008, Bohte *et al.*, 2009, Van Wee, 2009, and Van Acker *et al.*, 2014). Commonly labeled as the residential self-selection problem, the underlying problem is that the data available to assess the potential effects of land-use on activity-travel (AT) patterns is typically of a cross-sectional nature. In such observational data, the residential location of households and the activity-travel patterns of household members are jointly observed at a given point in time. Thus, the data reflects household residential location preferences co-mingled with the AT preferences of the households. On the other hand, from a policy perspective, the emphasis is on analyzing whether (and how much) a neo-urbanist design (compact BE design, high bicycle lane and roadway street density, good land-use mix, and good transit and non-motorized mode accessibility/facilities) would help in reducing motorized travel. To do so, the conceptual experiment that reveals the "true" effect of the BE features of the residential location on AT patterns is the one that randomly locates households in residential locations. The problem then, econometrically speaking, is that the analyst has to extract out the "true" BE effect from a potentially non-randomly assigned (to residential locations) observed cross-sectional sample. If the non-random

assignment can be completely captured by observed non-travel characteristics of households and the BE (such as, say, poor households locating in areas with low housing cost), then a conventional travel model accommodating the observed non-AT characteristics of households and the BE characteristics would suffice to extract the "true" BE effect on AT patterns. However, it is quite possible (if not likely) that there are some antecedent personality, attitude, and lifestyle characteristics of households that are unobserved to the analyst and that impact both residential location choice and activity-travel behavior, as discussed earlier. Ignoring such self-selection effects in residence choices can lead to a "spurious" causal effect of neighborhood attributes on activity-travel behavior, and potentially lead to misinformed BE design policies.

Many different approaches may be used to account for residential self-selection effects, a detailed review of which is beyond the scope of this dissertation (the reader is referred to Bhat and Guo, 2007, Bhat and Eluru, 2009, Mokhtarian and Cao, 2008, and Bhat, 2015). But, within the context of cross-sectional data, one broad direction is to more explicitly capture what is traditionally "unobserved" (latent) in typical travel survey data sets, and include these as "independent" variables. It is here that our proposed GHDM model comes into play.

Another important point of departure of the current empirical study from most earlier studies in the land use-transportation domain is that we examine residential self-selection (and more generally integrated land use-transportation modeling) in the context of an activity-based modeling (ABM) paradigm (see, for example, Bhat and Koppelman, 1993). As pointed out by Pinjari *et al.* (2009) and more recently by Chen *et al.* (2014), despite the fact that the ABM paradigm is increasingly now accepted even in practice as the approach of choice for travel analysis, there has been little consideration of residential self-selection issues within the ABM modeling paradigm. The central basis of the ABM paradigm is that individuals' activity-travel patterns are a result of their time-use decisions; individuals have 24 hours in a day (or multiples of 24 hours for longer periods of time) and decide how to use that time among activities and travel (and with whom) subject to their sociodemographic, spatial, temporal, transportation system, and other contextual constraints; see Bhat *et al*. (2004) and Pinjari and Bhat (2011b). In the activity-based approach, the impact of land-use and demand management policies on time-use behavior is an important precursor step to assessing the impact of such polices on individual travel behavior. Accordingly, in this research, we jointly model residential location-related choices along with auto ownership and activity time-use in different activities.

4.3.2 Data source and sample formation

The data source used in this study is the Puget Sound household travel survey conducted by the Puget Sound Regional Council (PSRC) in the spring (April–June) of 2014 in the four county PSRC planning region (the four counties are King, Kitsap, Pierce, and Snohomish) in the State of Washington. Households were randomly sampled, with the intent of obtaining a representative sample of households from the region for analyzing activity-travel patterns. The survey was administered by recruiting households using a stratified address-based sampling method based on the US Post Office's Computerized Delivery Sequence File (CDSF) that is a compilation of all mailing addresses in the US, providing coverage for approximately 97% of all households. Households were initially contacted using a "recruit survey" through which information on household-level socio-demographics (including motorized vehicle ownership by type, and home location address, housing type, and tenure status) and person-level information (including work and student status) was obtained. Only one adult household member (age 18 or older) was asked to complete the "recruit survey", and the corresponding household respondent was designated as the household reference person. The "recruit survey" also elicited information from the household reference person on the factors that influenced the current residential choice. This included the importance of the following six factors: (1) having a walkable neighborhood and being near local activities, (2) being close to public transit, (3) being within a 30-minute commute to work, (4) quality of schools in the neighborhood, (5) having space and separation from others, and (6) being close to the highway. Another part of the survey was a "retrieval survey" that comprised a comprehensive travel diary for a pre-defined household-specific mid-weekday (Tuesday, Wednesday, or Thursday) that each individual in the household (5 years or older) was asked to fill in at a "dashboard" web site generated for the household. Following the 24-hour diary portion of the retrieval survey, respondents were asked a series of questions about their typical transportation behaviors (to provide additional information beyond a single day's travel). Additional details of the survey recruitment and administration procedures are available in RSG (2014).

The survey collected information from a total of 6,036 households, of which 4,631 households had at least one worker employed in the household and with a work location outside the residential dwelling unit. The focus of the current analysis is on these 1+-worker households, to acknowledge the rather substantial differences in household residence and activity-travel

patterns between zero-worker households (retired couples, unemployed individual households, and student households) and 1+-worker households (see, for example, Rajagopalan *et al.*, 2009). After further screening to remove households with incomplete residence, travel, attitude, or demographic information, the final sample used in the current analysis included 3,637 households. In Appendix D, we provide descriptive characteristics of the socioeconomic characteristics of the sample.

4.3.3 Dependent variable characteristics

The dependent variables in our model system include a combination of a continuous variable, multiple ordinal indicators, a count variable, a nominal variable, and an MDC variable. The construction of each of these variables is discussed in turn in the subsequent paragraphs. Table 4.1 provides descriptive statistics of the dependent variables.

Commute distance, the continuous variable, was not reported directly by members of the household; it was derived by the Puget Sound Regional Council from shortest-path distance skims based on the home and primary work locations of each individual. We then computed a household average commute distance (miles) as the average one-way distance in miles between the home and the primary workplace across those individuals working outside the home (for brevity, from here on, we will refer to this variable as household commute distance). As may be observed from Table 4.1, the minimum and maximum household commute distances in the sample are 0.05 miles and 99.95 miles, respectively. The 95th percentile value for the household commute distance is 41.7 miles. In our estimation, we used the natural logarithm of household commute distance as the continuous dependent variable.

As indicated in the previous section, the household reference person was asked a series of questions to elicit preferences regarding residential choices. The responses to these questions were all collected on a five-point ordinal Likert scale. These questions and the distribution of the corresponding responses are shown in the second panel of Table 4.1. The statistics reveal, not surprisingly, that being within 30 minutes of work and proximity/walkability to local activities are "important" or "very important" considerations to more than 75% of the respondents when making residential choices.[22]

[22] "Quality of schools" is rated quite low in the overall. To examine if there is a substantial difference between households with children and without children, we examined the ratings on this question by presence or absence of children. The percentage of households that rated this attribute as being important or very important in the segment

The number of motorized personal vehicles in the household (that is, auto ownership), as reported in the survey by the household reference person, is a count dependent variable. The distribution of this variable (see the third panel of Table 4.1) indicates that most households have one or two cars (75.0%) and the average number of autos per household is 1.69.

Each household's residential location was assigned to one of the following nominal density categories: (a) 0–749 households per square mile, (b) 750–1,999 households per square mile, (c) 2,000–2,999 households per square mile, and (d) ≥ 3,000 households per square mile. The descriptive statistics in Table 4.1 for this nominal variable indicate that half of the households in the sample are located in high density areas, while about 13.2% are located in the lowest density areas. In the estimation, the highest density category is considered the base category. The use of density, along with commute distance, to characterize residential choice makes the definition of the residential choice alternatives clear and manageable. It also provides a convenient way to capture land-use/BE effects on auto ownership levels and activity time-use patterns, particularly because of the strong association between density and other BE elements. Indeed, there is a long and strong precedent for using residential density as a proxy for land-use/BE elements in the transportation literature (see, for example, Bhat and Singh, 2000, Chen *et al*., 2008, Kim and Brownstone, 2013, Paleti *et al.*, 2013, and Cao and Fan, 2012).

of households with children was 72.1%, relative to 22.0% in the segment without children. Clearly, as expected, there is a difference in the quality of school ratings based on the presence of children. This effect is captured in our analysis, as discussed later.

Table 4.1. Sample characteristics of dependent variables

Dependent variable: Continuous variable				
Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**
Household commute distance	14.47	13.78	0.05	99.95

Indicator variable: Ordinal variables					
Attitudinal Question	**Response rate**				
	Very Unimportant 1	**2**	**3**	**4**	**Very Important 5**
How important when choosing current home:					
Having a walkable neighborhood and being near to local activities	5.5%	7.6%	11.1%	32.3%	43.5%
Being close to public transit	15.4%	12.0%	17.0%	24.8%	30.8%
Being within a 30-minute commute to work	6.6%	6.5%	10.0%	24.4%	52.5%
Quality of schools (K-12)	31.2%	7.5%	26.7%	14.6%	20.0%
Having space and separation from others	9.2%	13.7%	21.8%	34.3%	21.0%
Being close to the highway	12.7%	16.0%	21.4%	38.0%	11.9%

Dependent variable: Count variable							
Motorized Vehicle Count	**Frequency**						
	0	**1**	**2**	**3**	**4**	**5**	**>6**
Number	304	1,378	1,354	413	135	36	17
%	8.4	37.8	37.2	11.4	3.7	1.0	0.5

Dependent variable: MNP variable	
Residential Density (households per sq. mile)	**Number of observations (%)**
<750	478 (13.2)
750-2,000	866 (23.8)
2,000-3,000	525 (14.4)
>3,000	1,768 (48.6)

Dependent variable: MDC variables				
Activity	**Participation (%)**	**Mean* fraction**	**Number of households (% of total number) spent time…**	
			Only in activity type**	**In other activity types too****
In home (IH)	3,637 (100.0)	0.780	533 (14.7)	3,104 (85.3)
Personal Business	1,607 (44.2)	0.202	216 (13.4)	1,391 (86.6)
Shopping	1,664 (45.8)	0.060	355 (21.3)	1,309 (78.7)
Recreation	1,011 (27.8)	0.131	148 (14.6)	863 (85.4)
Dining Out	1,092 (30.0)	0.081	203 (18.6)	889 (81.4)
Social	659 (18.1)	0.180	82 (12.4)	557 (87.6)
Serve Passenger	751 (20.6)	0.047	26 (3.5)	725 (96.5)

*: The mean duration of activities reported in the table are for only those who participated.

**: For the IH activity, the splits refer to participation only in IH activity and participation in IH activity and at least one OH activity purpose. For each OH activity purpose, the splits refer to participation in that OH activity purpose as well as another OH activity purpose (in addition to IH activity)

The MDC alternatives include in-home (IH) activity and six purposes of out-of-home (OH) activity: personal business, shopping, eating out, social activities, recreation, and serve passenger. The discrete component corresponds to household-level participation in these different activity purposes, while the continuous component corresponds to the amount of household time invested in these activity purposes. The following two step process was used to obtain the time spent on different activities by each household: (1) The activity episodes undertaken by each individual during the survey day were collected together by each of the seven activity purposes, and the total individual daily time-investment in each activity purpose was computed across all episodes of the activity purpose, (2) The activity times by purpose were aggregated across all individuals in each household to obtain household-level participations and time investments in IH activity and the six OH activity purposes. The total household time budget in the MDC model corresponds to the sum across the seven activity purposes (that is, this corresponds to total household time, or 24 hours times the number of individuals in the household, minus the time (across all individuals) spent on work, education, and sleep). In our analysis, for convenience, we use the household-level participations and fractions of time investments in each activity purpose as the dependent variables (that is, we effectively are normalizing the household time investments in each purpose by the total household budget, so that the continuous components correspond to fractions, and the total budget is 1 for each household).[23]

The final panel of Table 4.1 provides descriptive statistics of the time-use of households in the sample. All households participate in IH activity, which constitutes the outside good in the MDC model. Among the OH activity purposes, there is a relatively high participation level in personal business activity (44.2% of households) and shopping activity (45.8% of households), suggesting relatively high intrinsic baseline preferences for these two activity purposes. The social activity purpose and the serve passenger activity purpose, on the other hand, have the least participation rates, suggesting relatively low intrinsic baseline preferences for these two activity purposes. The third column indicates the fraction of time spent on each activity purpose, as averaged across households that participate in the corresponding activity purpose. For example, the first entry for IH activity shows that, on average, 78.0% of the total household time budget is

[23] The determination of how the OH participations and times are allocated across individuals in the household can be determined in a downstream allocation model, as in Gliebe and Koppelman, 2002.

spent on IH activity, while the entry for personal business activity reveals that, on average across the 44.2% of households who actually participate in personal business activity, 20.2% of the total household budget is spent on personal business activity. The implication from this third column is that, if participated in, the shopping, dining out, and serve passenger activity purpose are the ones on which the least time is spent, suggesting high satiation rates for these activity purposes.[24] The final two columns highlight the multiple-discrete nature of activity participations. The first row for IH activity shows that 14.7% of households participate in only IH activity (and no OH activity), while 85.3% of households participate in IH activity as well as one or more OH activity purposes. The second row for personal business reveals that 13.4% of households partake in personal business as the only OH activity (in addition to IH activity, which all households participate in), while 86.6% of households pursue personal business and at least one other OH activity purpose.

The discussions above are helpful to get a general idea of the patterns of preferences and satiation. However, the final baseline preference and satiation parameters for the activity purposes in the MDC model are based on a combination of participation rates, conditional-upon-participation durations, and the split between sole participations and participations with other activity purposes.

4.3.4 Latent constructs

In developing the latent variables to characterize attitudes and lifestyles, we examined earlier studies investigating (directly or indirectly) lifestyle-related characteristics affecting residential choice decisions, auto ownership choice, and activity time-use decisions (see, for example, Schwanen and Mokhtarian, 2007, Walker and Li, 2007, Van Acker *et al.*, 2014, Bohte *et al.*, 2009, de Abreu e Silva *et al.*, 2012, and Bhat *et al.*, 2014 for reviews of this literature). Some of these studies are based on intensive qualitative focus group interviews and/or ethnographic studies that tease out underlying psycho-social factors. These earlier studies, while labeling the factors sometimes differently, converge to two basic lifestyle-related factors: (1) *Green lifestyle propensity* and (2) *luxury lifestyle propensity*. The first latent variable drives the overall attitude

[24] Note that the mean fractions in this third column sum to greater than one across all activity purposes because the means are computed for each activity purpose conditional on households participating in that activity purpose. But the reader will note that the participation-weighted fractions in this third columns sum to 1: that is, 1*0.78+0.442*0.202+0.458*0.06+0.278*0.131+0.300*0.081+0.181*0.18+0.206*0.047=1 (after accounting for rounding).

and concern toward the environment, while the second reflects a penchant for consuming more, marked by a desire for privacy, spaciousness, and exclusivity. From a residential choice standpoint, the first latent variable has sometimes been referred to as "urban living propensity", while the second has been associated with "suburban/rural living propensity" and better quality public schools. From an auto ownership/modal standpoint, the first is sometimes referred to as "pro-public transportation" attitude, while the second has been associated with "pro-driving" attitude. From an activity time-use standpoint, the first latent variable has typically been associated with active recreation and non-motorized mode use, while the second has been associated with increased time investments in shopping and dining out activity participations. While one can justifiably argue that the latent variables above specific to each of the residential choice, modal/car ownership, and activity time-use dimensions are not perfectly correlated in the way suggested above, there are clearly very strong associations to the two basic lifestyle factors of green lifestyle propensity and luxury lifestyle propensity. So, from the standpoint of parsimony, as well as from the viewpoint of mapping the six ordinal attitudinal indicators and other dependent variable outcomes (see previous section) with the latent variable constructs, we decided to work with the two factors of (1) *green lifestyle propensity* (GLP) and (2) *luxury lifestyle propensity* (LLP). The first latent variable is a measure of the overall attitude and concern toward the environment, while the second reflects a penchant for consuming more, marked by a desire for privacy, spaciousness, and exclusivity. Our expectation is that households with a GLP disposition will prefer to reside in high density neighborhoods close to their workplace, own few or no vehicles, and engage more in IH activities and OH social and active recreation activities, while those with an LLP disposition will be inclined to locate in low to medium density neighborhoods, own many vehicles, and potentially be engaged in more OH shopping and dining out activities. However, these will be tested empirically in the measurement equation model during the specification and statistical testing process, as discussed later.

The reader will note that, as discussed above, we use earlier ethnographic and qualitative studies investigating (directly or indirectly) general lifestyle-related characteristics that affect residential choice, auto ownership, and activity time-use decisions as the basis to identify our latent variables (or constructs). As stated by Golob (2003), *"Theory and good sense must guide model specification"*. The fact that we have additional ordinal indicators related to residential choice preferences helps provide stability to the model system, but does not play a central role in

identifying the latent constructs per se. This is different from studies in psychology that collect a battery of tens (and sometimes hundreds) of indicators, and use exploratory factor analysis to identify a much fewer number of factors (or latent constructs) through analytic variance minimization. In our case, we identify plausible latent constructs first based on intuition and the findings from previous studies, and then use both the ordinal indicators as well as the actual endogenous variable outcomes together to help relate observed covariates to the latent constructs in the structural equation system. Once the latent constructs are identified, the final specification in the structural equation system and the measurement equation system (for the loadings of the latent constructs, and the effects of observed covariates, on the ordinal indicators and the dependent outcomes) is based on statistical testing using nested predictive likelihood ratio tests and non-nested adjusted predictive likelihood ratio tests.[25] For additional details, please see how the structural and measurement equation systems in Equation (B.1) of the Appendix B are converted to the joint reduced form system of Equation (B.2) for estimation.

4.3.5 Model estimation results

The final variable specification was obtained based on a systematic process of eliminating statistically insignificant variables, supplemented with a healthy dose of judgment and results from earlier studies. In the MDC activity time-use model, the τ-profile came out to be consistently superior to the α-profile for all variable specifications, and so is the one used.

4.3.5.1 Latent variable structural equation model results

The results of the structural equation model that relate the two latent psycho-social constructs of GLP and LLP as a function of demographic attributes are presented in Table 4.2.

The results suggest that lower income households have a higher GLP relative to higher income households (note that the highest income category is the base category in Table 4.2, and the coefficients for the other income categories are all positive with the magnitude being the highest for the lowest income category and decreasing thereafter). Table 4.2 also indicates that

[25] Indeed, almost all applications in the transportation literature that collect a handful of indicators use a combination of intuitiveness, judgment, and earlier studies to identify the latent constructs, rather than undertake a factor analysis of any kind to identify the latent factors (see, for example, Daly *et al.*, 2012, Bolduc *et al.*, 2005, de Abreu e Silva *et al.*, 2014, La Paix *et al.*, 2013, Temme *et al.*, 2008). But we acknowledge that there is some level of subjectivity in the number and "labels" of the latent variables, and these constructs can be questioned. But model building will always retain that element of judgment and subjectivity. The important point is that we have provided a conceptual basis for our selection of latent variables.

households with a high fraction of young adults (less than the age of 34 years) have a higher GLP relative to those with a low fraction of young adults. This latter effect is consistent with the environmental sociology literature (see, for example, Liu *et al*., 2014), which attributes this effect to young adults (especially the millennials) being increasingly exposed to environmental issues in the past decade through both school curricula and social media. Interestingly, age appears to have a U-shaped effect on GLP, with households with a high fraction of senior adults (65 years or older) having a higher GLP than households with a high fraction of middle-aged adults. Overall, households with a high fraction of adults in the 35-54 years age group seem to be the least "green". During the late 1990s, the Puget Sound Region succeeded in attracting young, well-educated workers into their region's workforce (Council, 2005). These young and highly-skilled "creative class" workers played a key role in the development of new technologies and industries, the creation of startup firms, and associated job growth during the technology boom of the late 90s. This creative class should be aged 35-54 years now and their past context of economic growth may explain their relatively low environmental consciousness (for an analysis of the inverse relationship between green life-style tendency and economic growth in the late 1990s, see Diekmann and Franzen, 1999).

The results also suggest a higher GLP associated with households with a high fraction of women (relative to a low fraction of women) and a high fraction of well-educated individuals in the household (relative to a low fraction of well-educated individuals).

The Table 4.2 results corresponding to LLP show that LLP increases with household income, the number of children in the household, and the age of household members in the household. The effect of income is very intuitive, because higher incomes provide not only the financial wherewithal to indulge, but an explicit show of indulgence may be viewed as a socio-cultural vehicle to signal wealth, power and status, and privileged access to limited resources.

The effect of children on LLP may be attributed to the desire for more privacy and separation from others to "protect" children from perceived unsafe levels of traffic and social environments (including safety from crime), a felt need to provide spacious indoor and outdoor play room for children, a desire for good quality schools (as observed in the descriptive statistics section), and an increase in motorized access to chauffeur children to activities, all of which are indicators of LLP (see next section).

Finally, the association between age and LLP may be related to the decrease in familial responsibilities with age, an increasing awareness of one's decreasing lifespan in which to expend any accumulated wealth, and a desire to experience the "unexperienced" (see Cleaver and Muller, 2001, and Twitchell, 2013). In earlier studies, age has been linked to luxury fashion consumption (see for example Li *et al.*, 2012), luxury cars purchases (Rosecky and King 1996), and luxury trips, such as cruises or exotic destinations (Hwang and Han, 2014).

The correlation coefficient between the GLP and LLP latent constructs is statistically significant at any reasonable level of significance, with a value of -0.16 and a t-statistic of -5.4. This negative correlation is reasonable, since a green lifestyle is associated with careful and conservative consumption of resources, while a luxury lifestyle correlates with extravagant living and indulgence beyond an indispensable minimum.

Table 4.2. Estimation results of structural equation

Variable	Coefficient	T-stat
Green Lifestyle Propensity (GLP)		
Household income (base: 75,000 or more)		
Less than 25,000	0.702	12.001
25,000 – 34,999	0.523	7.234
35,000 – 49,999	0.401	6.944
50,000 – 74,999	0.198	7.104
Age (base: fraction of adults in the age group 18-34)		
Fraction of adults in the age group 35-54	-0.478	-9.623
Fraction of adults in the age group 55-64	-0.331	-4.978
Fraction of adults in the age group 65 or more	-0.132	-1.941
Gender (base: fraction of female adults in the household)		
Fraction of male adults in the household	-0.029	-1.850
Education status (base: fraction of adults with less than a bachelor's degree)		
Fraction of adults with a bachelor's degree	0.160	4.101
Fraction of adults with an MS or PhD degree	0.203	2.103
Luxury Lifestyle Propensity (LLP)		
Household income (base: 75,000 or more)		
Less than 25,000	-0.201	-11.933
25,000 – 34,999	-0.322	-8.000
35,000 – 49,999	-0.431	-7.924
50,000 – 74,999	-0.472	-6.424
Number of children (less than 18 years old) in the household	0.473	11.926
Age (base: fraction of adults in the age group 18-34)		
Fraction of adults in the age group 35-54	0.130	3.553
Fraction of adults in the age group 55-64	0.412	3.567
Fraction of adults in the age group 65 or more	0.450	2.210
Correlation coefficient between 'active living/pro-environment attitude' and 'travel affinity/privacy desire' latent constructs	-0.168	-5.421

4.3.5.2 Measurement equation results for non-nominal variables

The results for the non-nominal variables are presented in Table 4.3. The dependent variables are organized column-wise and the independent variables are arranged row-wise.

The standard error corresponding to the natural logarithm of the household commute distance is 1.333 with a t-statistic of 3.28. The constants in the many equations, as well as the thresholds (note that in the model formulation, the first threshold ($\tilde{\psi}_{n,1}$) and the first flexibility parameter ($\varphi_0 = 0$) for the ordinal and count variable have been fixed to zero), do not have any substantive interpretations. For the auto ownership variable, the dispersion parameter (θ) became quite large during the estimation and was fixed at the value of 5.0 for estimation stability. The resulting specification is effectively the same as a flexible Poisson-based specification. The flexibility arises because we estimated two flexibility parameters for the auto ownership count to accommodate spikes in ownership of one car and two cars (see Table 4.2). These came out to be very statistically significant as follows: $\varphi_1 = 0.832$ (t-statistic of 10.22) and $\varphi_2 = 1.710$ (t-statistic of 11.09), and are not reported in Table 4.3.

The "number of children" effects in Table 4.3 (corresponding to elements of the coefficient vectors $\tilde{\boldsymbol{d}}$ and $\breve{\boldsymbol{d}}$ in Section 4.2 and in Appendix B) suggest that the presence of a child leads to a shorter household commute distance compared to the case without a child. Further, as the number of children increases, there is a continued linear reduction effect on household commute distance. In contrast to the negative relationship between number of children and household commute distance, there is a positive relationship between number of children and auto ownership propensity, presumably due to additional mobility needs placed upon the household to chauffeur children from one activity to another (see also Potoglou and Susilo, 2008 and Ma and Srinivasan, 2010 for a similar result).

The latent construct effects in Table 4.3 indicate, not surprisingly, that "green" households have a lower household commute distance relative to their peers, as such households are likely to consciously locate themselves closer to work locations to enable the use of non-motorized forms of transportation. The loadings of the latent constructs on the ordinal indicator variables are intuitive, and indicate that "green" households are likely to value, in terms of importance in residential choice decisions, being in a walkable neighborhood in proximal reach of activity opportunities, and being close to public transit and the work place. On the other hand,

households with a high LLP propensity value prefer neighborhoods with good quality of schools perhaps as a means to signal exclusivity as neighborhoods with good quality schools are typically synonymous with relatively wealthy neighborhoods with a good tax base (note also that the number of children does affect LLP propensity). Households with high LLP propensity also value space and privacy, have a preference to be in close proximity of highways (presumably as a means to retain the ability to reach activities quickly even while maintaining a very private, spacious, and exclusive living quarter), and have a penchant for owning more cars.

The endogenous effects in Table 4.3 are discussed together with the endogenous effects in Table 4.4 in Section 4.3.5.4.

Table 4.3. Estimation results for non-nominal variables of measurement equation

Independent variables	Continuous variable		Ordinal variables												Count variable	
	Natural logarithm of household commute distance*		Having a walkable neighborhood and being near local activities		Being close to public transit		Being within a 30-minute commute to work		Having space and separation from others		Quality of schools		Being close to the highway		Auto ownership	
	Coeff	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**
Constants	1.881	11.78	1.461	4.49	0.910	4.10	1.382	8.67	1.071	7.12	0.333	7.21	0.865	6.30	0.899	6.34
Thresholds for ordinal indicators																
Somewhat unimportant & not important			0.462	15.65	0.418	13.00	0.375	20.00	0.591	15.56	0.732	18.26	0.573	14.52		
Not important & somewhat important			0.822	15.10	0.873	20.01	0.717	14.61	1.184	18.32	3.640	19.33	1.110	15.68		
Somewhat important and very important			1.678	14.10	1.513	11.89	1.374	13.20	2.142	14.22	5.722	19.69	2.295	17.21		
Household characteristics																
Number of children in the household	-0.334	-6.23	------	------	------	------	------	------	------	------	------	------	------	------	0.070	2.190
Latent constructs																
Green Lifestyle Propensity (GLP)	-0.761	-12.12	0.203	13.72	0.262	11.81	0.297	14.71	------	------	------	------	------	------	-0.292	-11.41
Luxury Lifestyle Propensity (LLP)	------	------	------	------	------	------	------	------	0.251	4.66	3.800	17.82	0.201	5.08	0.110	2.19
Endogenous Effects																
Residential density (base: >3000 hh/sq-mile)																
Less than 750 hh/sq-mile	------	------	------	------	------	------	------	------	------	------	------	------	------	------	0.511	6.145
750-1999 hh/sq-mile	------	------	------	------	------	------	------	------	------	------	------	------	------	------	0.438	5.793
2000-3000 hh/sq-mile	------	------	------	------	------	------	------	------	------	------	------	------	------	------	0.311	5.454

------: Not significant in the case of the effect of residential density on commute distance, and not applicable in the case of the effects of residential density on the ordinal indicators (note that these ordinal indicators serve purely the purpose of better pinning down the latent constructs and the relationship between the latent constructs and exogenous covariates in the structural equation system).

*: Estimated variance of commute distance is 1.333 and the associated t-stat is 3.281.

4.3.5.3 Residential density choice model and activity time-use results

The estimation results for residential density and activity time use are presented in Table 4.4. The constant parameters do not have any substantive interpretation because of the presence of the continuous latent variables.

The effects of the family structure variables indicate that single person households are most likely to stay away from the lowest density neighborhoods, while households with children (in particular, nuclear and single parent families) are most likely to live in the lowest density neighborhoods. Earlier research (see Kim and Chung, 2011) does suggest that single person households tend to locate themselves in denser neighborhoods, enabling easy access to social and related activity opportunities. Interestingly, single person households also appear to prefer medium-high density (2000-2,999 households per square mile) neighborhoods relative to the highest density neighborhoods, perhaps as a way of balancing space/privacy with activity accessibility and social networking opportunities in the immediate vicinity. The effects of the family structure variables on activity time-use indicate that single person households have the highest preference for in-home activities, while nuclear families and single-parent families, relative to other household types, have a clear higher baseline preference for OH shopping and serve passenger activities. On the other hand, there is an indication that single parent households, relative to nuclear families, are time poor (lack of time for leisure, sports, and relaxation activities) and have the danger of social exclusion (broadly defined as the "inability to participate fully in society", one aspect of which is not being able to participate in the "normal activities of daily life"; see Farber *et al*., 2011).

Table 4.4. Estimation results for nominal variables of measurement equation

Independent Variables	Residential location (base: >3000 hh/sq-mile)						Fraction of time spent on various activities by household (base: In-home)											
	Less than 750 hh/sq-mile		750-1999 hh/sq-mile		2000-3000 hh/sq-mile		Personal Business		Shopping		Recreation		Dining Out		Social		Serve Passenger	
	Coeff	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**
Constant	-0.680	-6.73	-0.393	-4.88	-0.636	-9.53	-0.143	-3.64	-0.43	-7.70	-0.69	-12.66	-0.567	-14.05	-1.344	-17.48	-1.549	-19.99
Family structure[a]																		
Single person HH	-0.180	-3.21	------	------	0.088	2.10	------	------	-0.596	-6.94	-0.565	-5.89	-0.144	-1.86	-0.405	-3.50	-1.411	-7.87
Nuclear family	0.355	10.23	------	------	------	------	------	------	1.611	23.13	0.446	5.49	------	------	0.312	3.13	1.923	22.29
Single parent family	0.619	7.43	------	------	0.312	9.08	------	------	2.472	13.58	------	------	------	------	------	------	1.392	5.61
Fraction of adults by work status in HH[b]																		
Part-time workers	0.256	2.19	0.282	2.26	0.110	2.00	0.365	3.06	0.679	5.73	------	------	------	------	0.493	2.88	0.492	3.07
Self-employed Workers	0.320	3.04	0.284	4.82	0.132	2.16	------	------	0.274	1.98	------	------	------	------	0.394	2.06	------	------
Non-workers	0.410	2.87	0.290	3.32	0.187	3.21	0.762	5.85	1.167	8.80	0.391	2.51	------	------	0.771	4.09	1.122	6.56
Latent constructs																		
Green Lifestyle Propensity (GLP)	-0.051	-2.22	-0.152	-6.09	-0.098	-3.62	-0.720	-2.42	-0.681	-5.72	0.089	4.68	-1.030	-8.39	0.124	2.26	------	------
Luxury Lifestyle Propensity (LLP)	-0.190	-2.17	0.073	2.90	0.051	2.82	------	------	0.265	2.29	------	------	0.125	2.20	------	------	------	------
Satiation parameters							0.029	24.23	0.075	20.62	0.092	18.33	0.038	19.22	0.168	14.26	0.017	15.98
Endogenous Effects																		
Commute distance	------	------	------	------	------	------	------	------	0.203	7.67	0.152	5.42	0.268	3.28	------	------	------	------
Residential density (base: >3000 hh/sq-mile)																		
Less than 750 hh/sq-mile	------	------	------	------	------	------	------	------	-0.681	-7.90	-0.203	-2.38	-0.456	-4.15	0.269	2.33	0.971	8.81
750-1999 hh/sq-mile	------	------	------	------	------	------	0.177	2.66	-0.517	-7.16	------	------	-0.423	-4.88	------	------	0.614	6.19
2000-3000 hh/sq-mile	------	------	------	------	------	------	------	------	-0.234	-2.72	-0.245	-2.41	-0.493	-4.76	------	------	0.510	4.42

------: Not significant

[a]: base is couple family and multi-adult households

[b]: base is full-time workers

The next set of variables relate to the fraction of part-time, self-employed, and non-workers in the household, with the fraction of full-time workers in the household constituting the base category. Overall, these coefficients indicate a pattern where households with a high fraction of full-time workers have a clear preference to reside in the highest density areas, with a generally increasing tendency of households with higher fractions of part-time, self-employed, and non-workers to locate in progressively lower density areas. This result may be a reflection of the benefits of knowledge spillovers through networking opportunities in highly dense urban regions, which enable full time workers to retain (and enhance) their competitive edge in the market place (see Autant-Bernard and LeSage, 2011).

In terms of the latent constructs, "green" households tend to locate themselves in the highest density neighborhoods (>3000 households per square mile) and shy away from the medium density categories (750–1,999 or 2,000-2,999 households per square mile), while households with a high LLP tend to locate themselves in the medium density categories. The latter effect may be attributed to seeking a good balance between less dense, exclusive neighborhoods and good auto-based accessibility to OH activity opportunities. In addition, the effects of the latent constructs in the activity time-use model suggest that households with a high GLP, relative to their peers with a low GLP, spend more time at home, are less likely to pursue the more money-consuming (and potentially viewed as less "green") personal business, shopping, and dining out activities, and are more likely to seek social networking opportunities as well as pursue active recreation and other recreation activities (such as going to sports events, theaters, cinemas or art galleries). Finally, in terms of the latent construct effects, households with a high LLP spend more time than their peers with a low LLP on shopping and dining out. This is reasonable, because such individuals not only have the financial wherewithal to consume goods and services, but may also use shopping and dining activities at fancy places as a way to seek social differentiation and signal power and wealth.

The satiation parameters in Table 4.4, along with the baseline preference constants and baseline parameters, are estimated for each activity purpose (except the IH activity purpose) to best replicate the combination of participation rates, conditional-upon-participation durations, and the split between sole and joint participations with other activity purposes. The satiation parameters in Table 4.4 correspond to the τ-profile. Satiation increases for purpose *k* as τ_k goes closer to zero ($\tau_k \to 0$ for the IH activity in the τ-profile by construction, because the IH

activity is always participated in and has a high baseline constant that has to be compensated by the high satiation). As expected initially from the descriptive statistics, the shopping, dining out, and serve passenger activity purposes have high satiation rates (low values of τ_k) among the OH activity purposes. The social activity purpose has a low participation rate, but a high duration conditional on participation, which leads to the low satiation (high value of τ_k) for this purpose given its high negative baseline constant. For the personal business purpose, while it has both a high participation rate and a high duration conditional on participation, it has the lowest participation all by itself as an OH activity purpose excepting for the social and serve passenger purposes (see Table 4.1). The result is that the satiation parameter has to accommodate this high tendency for non-solo personal business participations, which leads to a relatively high satiation (low value of τ_k) parameter for the personal business purpose.

In each of the residential density and activity time use models, we also allowed a general error covariance matrix but we could not reject the hypothesis that the error covariance matrix was different from an independent and identically distributed error structure.

4.3.5.4 Endogenous effects

Tables 4.3 and 4.4 also present the endogenous effects. The final directions of the recursive endogenous effects were obtained in the current research after extensive testing of various model specifications, and choosing the specification that provided the best data fit in terms of the composite marginal log-likelihood value (note, however, that regardless of the presence or absence of recursive effects, the model is a joint model because of the presence of latent variables that impact the many dependent variables).

Figure 4.1 presents the overall directions of the endogenous relationships, while also including the effects of the GLP and LLP latent constructs on the endogenous outcomes, as discussed in the previous two sections. Further, the figure presents the sign of the effects of the GLP and LLP constructs on the residential density, commute distance, and auto ownership endogenous outcomes (but not on the activity time-use variable, because this is a multiple discrete variable with differing effects of the latent constructs on different activity purposes). All of the latent constructs and the endogenous outcomes in Figure 4.1 are affected by demographic factors, which we do not show in Figure 4.1 to focus on the endogenous effects. Our results (see Figure 4.1 as well as Tables 4.3 and 4.4) of the endogenous effects indicate that, after

accommodating the jointness among the dependent variables caused by the latent (and stochastic) GLP and LLP latent constructs, the choice of residential density impacts both auto ownership and activity time-use. In particular, residing in lower (higher) density neighborhoods leads to a higher (lower) auto ownership level, as has been well established in much of the earlier literature (see, for example, Bhat and Guo, 2007; Bhat *et al.*, 2009; Aditjandra *et al.*, 2012, Bhat *et al.*, 2014, and Brownstone and Fang, 2014). Also, lower (higher) density tends to result in lower (higher) baseline preferences for (*i.e.*, participations and time investments in) OH recreational activities, shopping, and dining out. These impacts may be attributed to higher densities being strongly correlated with more walk and bicycle infrastructure, better public transit services, and more opportunities for OH activities, and are consistent with earlier studies on time-use and physical activity. For example, Forsyth *et al.* (2009) and McCormack *et al.* (2014) indicate that higher density and mixed land-use increase time spent in neighborhood physical activity (primarily walking), while Wendel-Vos *et al.* (2007) and Ding *et al.* (2013) identify proximity to recreational activities (such as parks and exercise facilities) and even shopping locations as promoters of leisure time and overall physical activity. Also, Bhat *et al.* (2013b) and Born *et al.* (2014) find, consistent with our findings, that households in urban areas and high OH activity accessibility areas participate more in recreation, shopping, and dining out than peer households residing in other areas. On the other hand, the increased preference for OH social activities in the most sparsely populated neighborhoods is presumably because social activities are the easiest to pursue in locations with few to no activity centers (shopping places, restaurants, gyms, *etc*.). Further, as discussed in earlier studies (see Coleman, 2009, Romans *et al.*, 2011, and Bernardo *et al.*, 2015), this result is suggestive of a business-like culture in urban areas that is moving away from the relatively close-knit, informal, and social networks, but that still exists in non-urban areas for visiting and social get-togethers. Finally, in terms of residential location effects on time-use, time investment in serve passenger activity increases as one moves from the highest density neighborhoods to progressively lower density neighborhoods.

Interestingly, we did not find any statistically significant evidence of a direct causal relationship between residential (household) density and commute distance, or auto ownership and commute distance. The former result suggests that simply building compact cities will not necessarily translate to more sustainable travel in terms of shorter commute distance, contrary to some other studies that suggest there are commuting-based sustainability benefits of compact

cities (see, for example, Boussauw *et al.*, 2012). That is, while building compact neighborhoods may lead to shorter commutes for households who choose to reside in these compact neighborhoods, our results suggest that this is because households with a green lifestyle propensity self-select to live in such neighborhoods while those who are not green move out of such neighborhoods and have long commute distances. Thus, in the population as a whole, compact developments may not lead to shorter commute distances. The results in Figure 4.1 also indicate that auto ownership, by itself, has no impact on activity time-use. The implication, as in Bhat and Steed (2002) and Grigolon *et al.* (2013), is that lifestyles, demographics, and activity opportunities are the main drivers of activity-travel patterns.

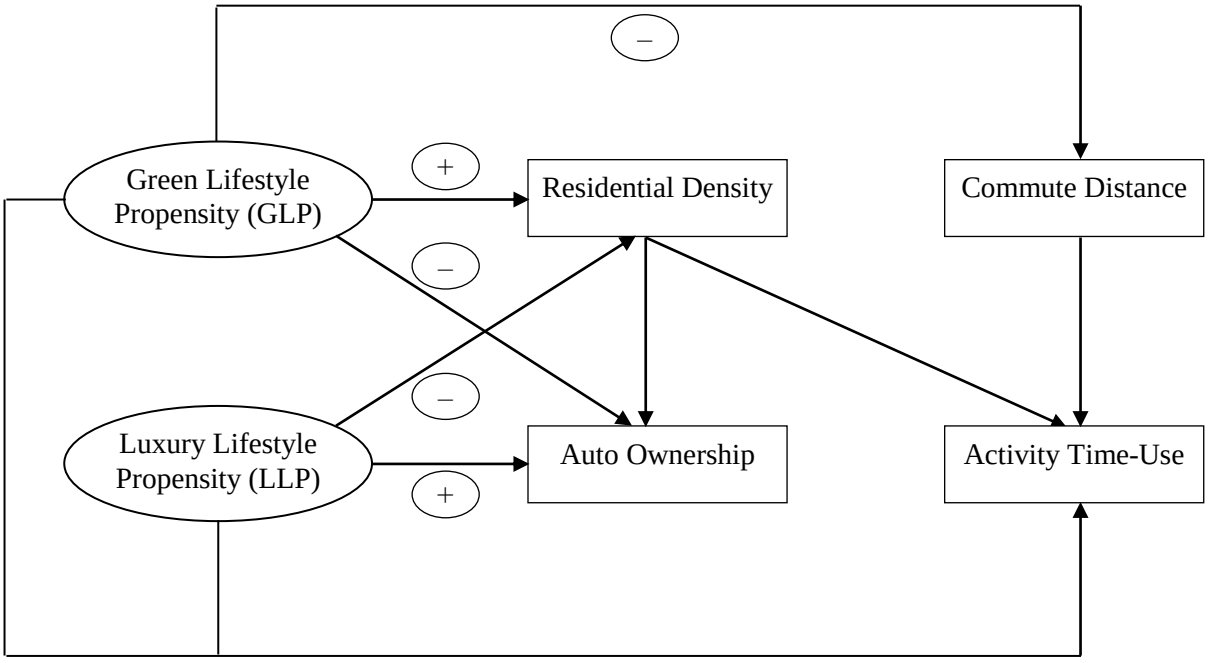


Figure 4.1. Effects of latent constructs and endogenous effects

Commute distance, causally speaking, impacts only time use (Figure 4.1 and Table 4.4); households with longer commute distances spend more time on shopping, recreation, and dining out. This may the result of two reinforcing effects. First, as household commute distance increases, the number of opportunities for shopping, recreation, and dining out increases. Second, as household commute distance increases, it puts more time pressure on the household, which may be released by shopping more for easy-to-prepare meals and dining out. Some earlier

studies, including Wang *et al.* (2013) and Castro *et al.* (2011), have suggested the reverse -- that households with shorter commute distances participate more in non-work activities because of denser non-work activity locations and less time pressure. However, these earlier studies do not consider residential self-selection effects as we do. But this subject of the relationship between commute distances and non-work activity participation certainly deserves more exploration and the disentangling of multiple push-pull effects, as also acknowledged by the earlier studies just identified.

4.3.5.5 Model data fit comparisons

To assess the importance of considering jointness across choice dimensions, we also estimated an Independent Heterogeneous Data Model (IHDM) that does not consider such jointness (that is, the covariances engendered by the stochastic latent constructs in the GHDM model are ignored). In this IHDM model, we introduce the exogenous variables (sociodemographic variables) used to explain the latent constructs as exogenous variables in the choice dimension equations. This way, the contribution to the observed part of the utility due to sociodemographic variables is still maintained (and is allowed to vary relative to the GHDM to absorb, to the extent possible, the GHDM covariances due to unobserved effects). The resulting IHDM may be compared to the GHDM using the composite likelihood information criterion (CLIC) introduced by Varin and Vidoni (2005). The CLIC takes the following form (after replacing the composite marginal likelihood (CML) with the maximum approximate CML (MACML)):

$$\log L^{*}_{MACML}(\hat{\tilde{\boldsymbol{\theta}}}) = \log L_{MACML}(\hat{\tilde{\boldsymbol{\theta}}}) - tr\left[\hat{\boldsymbol{J}}(\hat{\tilde{\boldsymbol{\theta}}})\hat{\boldsymbol{H}}(\hat{\tilde{\boldsymbol{\theta}}})^{-1}\right] \tag{4.17}$$

The model that provides a higher value of CLIC is preferred. The $\log L_{MACML}(\hat{\tilde{\boldsymbol{\theta}}})$ values for the GHDM and IHDM models were estimated to be -227,321.0 and -253,231.1, respectively, with the corresponding CLIC statistic values of -227,504.0 and -253,432.0. These CLIC statistics clearly favor the GHDM over the IHDM.

All the ordinal variables used in the measurement equation are included solely for the purpose of model identification and do not serve any purpose in predicting the choice bundle once the model is estimated. Therefore, we can also use the familiar non-nested likelihood ratio test to compare the two models. To do so, we evaluate a predictive log-likelihood value of both the GHDM and IHDM models using the parameter values at the MACML convergent values by excluding the six ordinal variables. The same is also done to obtain to the constants-only log-

likelihood value. Then, one can compute the adjusted likelihood ratio index of each model with respect to the log-likelihood with only the constants as follows:

$$\bar{\rho}^2 = 1 - \frac{\mathcal{L}(\hat{\hat{\boldsymbol{\theta}}}) - M}{\mathcal{L}(c)}, \tag{4.18}$$

where $\mathcal{L}(\hat{\hat{\boldsymbol{\theta}}})$ and $\mathcal{L}(c)$ are the predictive log-likelihood functions at convergence and at constants, respectively, and *M* is the number of parameters (not including the constant(s) for each dimension and not including the ordinal indicators) estimated in the model. This test determines if the adjusted likelihood ratio indices of two non-nested models are significantly different. In particular, if the difference in the indices is $(\bar{\rho}_2^2 - \bar{\rho}_1^2) = \tau$, then the probability that this difference could have occurred by chance is no larger than $\Phi\left\{-[-2\tau\mathcal{L}(c) + (M_2 - M_1)]^{0.5}\right\}$ in the asymptotic limit. A small value for the probability of chance occurrence indicates that the difference is statistically significant and that the model with the higher value for the adjusted likelihood ratio index is to be preferred. The $\mathcal{L}(\hat{\hat{\boldsymbol{\theta}}})$ values (number of parameters) for the GHDM and IHDM models were computed to be -21,322.1 (number of parameters = 89) and -32,028.1 (number of parameters = 152), respectively. The $\mathcal{L}(c)$ value was -44,402.1, with the corresponding predictive $\bar{\rho}^2$ values of 0.518 and 0.275 for the GHDM and IHDM models, respectively. The non-nested adjusted likelihood ratio test returns a value of $\Phi(-147)$, which is literally zero, clearly rejecting the IHDM model in favor of the GHDM model and underscoring the importance of considering the stochastic latent constructs that engender covariation among the choice dimensions.

4.3.6 Examining "true" effects of neo-urbanist densification efforts

To demonstrate the value of the proposed model, consider the GLP-caused associations among the many dimensions and, for now, ignore the LLP-caused associations. Also, we confine our attention to residential density, auto ownership, and OH recreational activity. According to our GHDM results, households with a high GLP have a generic preference (due to unobserved factors) to reside in the highest density neighborhoods, have low auto ownership levels, and are likely to pursue more OH recreational pursuits. Thus, because of GLP, households who happen to reside in the highest density neighborhoods tend to be there already because they are generically auto-disinclined and like to pursue recreational activities. But, even after capturing

these pre-dispositions (or associations) due to residential self-selection caused by unobserved factors, the GHDM indicates, through the endogenous effects, that the higher density "truly causes" households to own fewer cars and partake more in recreation pursuits. But if the residential self-selection effects were 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 auto ownership reduction (because the low auto ownership predisposition of the people living in the highest density neighborhoods would get tagged on to the "true" negative causal effect). Similarly, the positive effect of residential density on OH recreational pursuits would also be magnified (because the high OH recreational participation of the people living in the highest density neighborhood would again get tagged on to the "true" positive causal effect. In both these cases, there would be an overestimation of auto ownership reduction and OH recreational activity participation increase attributable to densification. Of course, how these impact motorized travel and traffic patterns will have to be determined through downstream models in an activity-based modeling system. The important point is that ignoring residential self-selection could lead to incorrect conclusions on the effects on auto ownership and activity time-use.

The intuitive explanation above does not consider the LLP-caused associations. Also, in the IHDM model, we allow explanatory demographic variables to impact the many choice dimensions directly. Thus, the final "net" effect of not accommodating residential self-selection cannot be gleaned as easily as described above. But to show a cumulative effect of capturing versus not capturing residential self-selection effects, we compute average treatment effects (ATEs) from the GHDM and IHDM 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 $i' \neq i$. We compute this measure for auto ownership and activity time-use as discussed in Appendix E.

The analyst can compute the ATE measures for all the pairwise combinations of residential density category relocations. Here, we focus on the case when a household in the lowest density neighborhood (<750 households per square mile) is transplanted to the highest density neighborhood (>3000 households per square mile). For ease in discussion, in the rest of this section, we will refer to the former neighborhood type as a low density neighborhood, and the latter neighborhood type as a high density neighborhood. Table 4.5 presents the estimated

ATE values (and standard errors) for auto ownership and out-of-home activities for both the GHDM and IHDM models. The first row under the "GHDM model" heading indicates that a random household that is shifted from the low density category location to the high density category location is, on an average, likely to reduce its auto ownership level by 0.143 vehicles (standard error of 0.011). Equivalently, if 100 random households are relocated from the low density neighborhood to the high density neighborhood, the point estimate indicates a reduction in auto ownership by about 14 vehicles. On the other hand, the IHDM model estimate predicts a reduction of 0.340 vehicles (standard error of 0.021). That is, if 100 random households are relocated from the low density neighborhood to the high density neighborhood, the independent model point estimate projects a reduction in motorized vehicle ownership by about 34 vehicles. The exaggeration in the reduction in auto ownership based on the IHDM model (because of the change in residence from the low density to the high density neighborhood) is readily apparent, and is a reflection of unobserved residential self-selection effects not being controlled for. The t-statistic value for the hypothesis of equality in the ATE estimates is 9.4, much higher than the table value even at the 0.005 level of significance, strongly rejecting equality between the two models.

The other rows of the table provide the ATE values with respect to each of the OH activity purposes. For example, the ATE for the GHDM corresponding to personal business indicates that a random household that is shifted from the low density category location to the high density category location is, on average, likely to reduce its participation probability in personal business activity by 0.037. Equivalently, if 100 random households are relocated from the low density neighborhood to the high density neighborhood, the point estimate indicates a reduction in personal business activity by 3.7 participations during the course of the day. Other values may be similarly interpreted. The results show that the IHDM model exaggerates the ATE for every OH purpose, whether positive or negative. The ATEs for all OH activity purposes and both models are statistically significant at least at the 0.1 level of significance, and generally at a much lower level of significance. The t-statistics for testing the differences in the ATE estimates between the two models are in the range of 1.0-2.3 for the shopping, recreation, dining out, and social activities, though there is literally no statistically significant difference the personal business and serve passenger purposes. Overall, the results show that, if self-selection effects are ignored, the result is exaggerated effects of densification.

Table 4.5. Treatment effects corresponding to transplanting a random household from a lowest density neighborhood (<750 hh/sq. mile) to highest density neighborhood (>3000 hh/sq. mile) (standard error in parenthesis)

Variable	ATE from GHDM	ATE from IHDM	% Difference Attributable to	
			"True" Effect	Self-Selection Effect
Vehicle ownership	0.143 (0.011)	0.340 (0.021)	42	58
Participation on				
Personal business	-0.037 (0.013)	-0.041 (0.013)	90	10
Shopping	0.011 (0.004)	0.019 (0.007)	65	35
Recreation	0.134 (0.021)	0.190 (0.014)	71	29
Dining out	0.094 (0.020)	0.119 (0.021)	79	21
Social	-0.056 (0.014)	-0.078 (0.017)	72	28
Serve Passenger	-0.156 (0.033)	-0.162 (0.025)	96	4

One can also quantify the magnitude of the "true" effect and the spurious residential self-selection effect because the IHDM model comingles these effects, while the joint model estimates the "true" effect. Because the IHDM model consistently exaggerates the ATE, the "true" effect may be computed as a percentage of the GHDM ATE relative to the IHDM ATE, while the self-selection effect may be computed as the difference of the ATE of the two models as a percentage of the IHDM ATE. The last two columns of Table 4.5 indicate that unobserved self-selection effects are estimated, based on the point estimates, to constitute about 58% of the difference in the number of autos between low density and high density households, while "true" built environment effects constitute the remaining 42% of the difference. Clearly, the self-selection effect is larger than the "true" effect, showing that ignoring self-selection will substantially overestimate the benefits of densification from an auto ownership reduction standpoint. Among the OH activity purposes, the self-selection effect is highest for the shopping, recreation, and social purposes, and the lowest for the serve passenger and personal business purposes. While the self-selection effect is lower than the "true density effect" for the OH activity purposes, it is still of the order of 30% for the shopping, recreation, and social purposes.

4.4 CONCLUSIONS

In this chapter, we introduce a joint mixed model that includes an MDC outcome and a nominal discrete outcome, in addition to count, binary/ordinal outcomes, and continuous outcomes. The outcomes are modeled jointly in a parsimonious fashion by specifying latent underlying unobserved individual lifestyle, personality, and attitudinal factors. Reported subjective attitudinal indicators for the latent variables help provide additional information and stability to the model system. In addition, we formulate and implement a practical estimation approach for the resulting model using Bhat's (2011) maximum approximate composite marginal likelihood (MACML) inference approach.

From an empirical standpoint, we focus on examining residential self-selection in the context of an activity-based modeling (ABM) paradigm. In the activity-based approach, the impact of land-use and demand management policies on time-use behavior is an important precursor step to assessing the impact of such polices on travel behavior. Accordingly, in this research, we jointly model residential location-related choices (density of residential location and commute distance), along with auto ownership and activity time-use, in a way that has a social-psychological underpinning through latent variables while also explicitly considering residential self-selection issues.

The empirical application uses data from the 2014 Puget Sound Household Travel Survey. Two basic lifestyle-related factors; *Green lifestyle propensity* and *luxury lifestyle propensity*; are used to explain the multiple mixed dependent variables. These two latent and stochastic psycho-social constructs impact the dependent variables and engender covariation among them. The proposed generalized heterogeneous data model (GHDM) model with an MDC variable clearly rejects a simpler independent heterogeneous data model (IHDM) that ignores the effects of the latent constructs. Effectively, this implies the presence of self-selection effects (endogeneity), and suggests that modeling the choice processes independently will not capture true relationships that exist across the choice dimensions. This is also evidenced in the ATE measures, which emphasize that accounting for residential self-selection effects are not simply esoteric econometric pursuits, but can have important implications for land-use policy measures that focus on neo-urbanist design.

To summarize, this research proposes and applies an integrated framework to model multiple types of variables, including continuous, ordinal, count, nominal, and multiple discrete-

continuous (MDC) variables. This research also contributes to disentangling residential self-selection effects from "true" density effects on activity pursuits and auto ownership. We hope that the elegant way of tying the mixed types of dependent variables, including an MDC variable, through a parsimonious latent structure approach will open new doors in the exploration of the nexus between land use and activity-based travel modeling, as well as contribute to empirical research in many other fields where MDC variables occur frequently.

Chapter 5: Conclusions and Directions for Future Research

5.1 DISSERTATION CONTRIBUTIONS

The primary objective of this dissertation is to advance the econometric modeling of MDC choice situations, with an emphasis on two aspects of this modeling. The first is to include, in a general way, heterogeneity in the sensitivity to exogenous variables. The second is to extend the joint modeling of mixed outcomes to include MDC outcomes. Specific contributions of the dissertation include the following.

5.1.1 Finite discrete mixture of normal (FDMN) version of the MDCP model

Chapter 2 has proposed a new econometric formulation and a complete blueprint of an associated estimation method for a finite discrete mixture of normals version of the multiple discrete-continuous probit (or FDMN-MDCP) model. The model allows consumers to choose multiple alternatives at the same time, along with the continuous dimension of the amount of consumption, and captures heterogeneity in the response coefficients of the baseline utility function. This is a very general way of including heterogeneity in the sensitivity to exogenous variables in the multiple discrete-continuous context, with the normally distributed random parameters approach and the latent class approach constituting special cases. The proposed approach is applied to model individuals' recreational (long distance leisure trips) choice among alternative destination locations and the number of trips to each recreational destination location, using data drawn from the 2012 New Zealand Domestic Travel Survey (DTS). The Bayesian Information Criterion indicates that the preferred specification is a three-segment solution, with one segment loading on high flying low family commitment (HFLFC) individuals, the second on low income parents (LIP), and the third on couple baby-boomers (CBB). In a comparative empirical assessment of the FDMN-MDCP with the simpler LC-MDCP and RC-MDCP models, the FDMN-MDCP came out clearly as the winner in terms of data fit. More importantly, the FDMN-MDCP formulation appears to be a valuable methodology for marketing and positioning in markets that are characterized by multiple discreteness.

5.1.2 Multivariate skew-normal (MVSN) distribution for unobserved heterogeneity in the spatial MDC model

In the third chapter, the MVSN distribution is used to include non-normality in the unobserved heterogeneity and kernel error structure. The MVSN distribution is tractable, parsimonious in parameters that regulate the distribution and its skewness, and includes the normal distribution as a special interior point case. To our knowledge, this is the first time a flexible and parametric skew-normal distribution for the kernel error term and/or random response coefficients has been used in both spatial- and aspatial-MDC models. The resulting model is estimated by using Bhat's (2011) maximum approximate composite marginal likelihood (MACML) inference approach. Simulation exercises are undertaken to examine the ability of this estimation method to recover parameters from finite samples. As an empirical demonstration, the proposed approach is applied to land-use-change decisions using the city of Austin's parcel-level land-use data. The results highlight the importance of introducing social dependence effects and non-normal kernel error terms from a policy standpoint.

5.1.3 Incorporating a MDC outcome in the estimation of joint mixed models

A joint mixed model that includes an MDC outcome and a nominal discrete outcome, in addition to count, binary/ordinal outcomes, and continuous outcomes was presented in Chapter 4. The outcomes are modeled jointly in a parsimonious fashion by specifying latent underlying unobserved individual lifestyle, personality, and attitudinal factors. Reported subjective attitudinal indicators for the latent variables help provide additional information and stability to the model system. In addition, a practical estimation approach for the resulting model using was implemented. From an empirical standpoint, residential self-selection in the context of an activity-based modeling (ABM) paradigm was examined. Residential location-related choices (density of residential location and commute distance) were modeled jointly, along with auto ownership and activity time-use, in a way that has a social-psychological underpinning through latent variables while also explicitly considering residential self-selection issues.

5.2 LIMITATIONS OF THE CURRENT RESEARCH AND DIRECTIONS FOR FUTURE WORK

This dissertation makes several contributions, as we discussed in the previous section. However, there are limitations of the current research that need to be explored in the future. Moreover, there are research areas that can expand the scope of the current study. A few of these ideas/thoughts are discussed below.

1. The FDMN framework allow us to implement an extremely flexible error structure for the MDCP model. On the other hand, the multivariate skew distribution is a more simple (fewer parameters) way to introduce non-normality in the MDCP model. A way to compare (or possibly combine) both frameworks is still in discussion.
2. The empirical application of the FDMN-MDCP model is only one limited example. Future research should focus on applying the FDMN-MDCP formulation to other multiple discrete contexts. Also, while the application to recreational destination choice demonstrates the value of the formulation, future work should consider a much richer set of destination region attributes.
3. Only one empirical application is presented for the SSN-MDC model. It is still under question the advantages of implementing such sophisticated model (skewness and spatial dependency modeled together) in other empirical contexts.
4. The empirical application of the SSN-MDC model needs to be validated with future year data of urban development and land-use changes. Also, the link between our proposed land-use change model and an economic analysis (from the different stakeholders' point of view, including environmental impacts) is still needed.
5. Alternative ways to include non-normality are available and their interaction with spatial dependency need to be explored within the MDC framework.
6. We hope that the elegant way of tying the mixed types of dependent variables, including an MDC variable, through a parsimonious latent structure approach will open new doors in the exploration of the nexus between land use and activity-based travel modeling, as well as contribute to empirical research in many other fields where MDC variables occur frequently.
7. The MDC model that was included in the mixed modeling framework is simple and customized to the particular case of time allocation and activity participation. Further studies exploring incorporating individual-specific satiation parameters may improve the model framework. In addition, a model with goods (and price variation) instead of times can be applied to many other different MDC contexts.

References

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.

Ahn, J., Jeong, G., and Kim, Y., (2008). A forecast of household ownership and use of alternative fuel vehicles: A multiple discrete-continuous choice approach. *Energy Economics* 30(5), 2091-2104.

Amador, F.J., R. Gonzales and J. Ortuzar (2005). Preference heterogeneity and willingness to pay for travel time savings. *Transportation*, 32(6), 627-647.

Allenby, G.M. (1990). Cross-validation, the Bayes theorem, and small-sample bias. *Journal of Business & Economic Statistics*, 8(2), 171-178.

Allenby, G. M., Garratt, M. J., and Rossi, P. E., 2010. A model for trade-up and change in considered brands. *Marketing Science* 29(1), 40-56.

Arellano-Valle, R.B., Azzalini, A., (2006). On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics* 33(3), 561-574.

Autant-Bernard, C., and LeSage, J.P. (2011). Quantifying knowledge spillovers using spatial econometric models. *Journal of Regional Science*, 51(3), 471-496.

Azzalini, A., (2005). The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics* 32(2), 159-188.

Azzalini, A., and Dalla Valle, A., (1996). The multivariate skew-normal distribution. *Biometrika* 83(4), 715-726.

Balcombe, K., A. Chalak and I.M. Fraser (2009). Model selection for the mixed logit with Bayesian estimation. *Journal of Environmental Economics and Management*, 57(2), 226-237.

Bastin, F., C. Cirillo and P.L. Toint (2010). Estimating non-parametric random utility models, with an application to the value of time in heterogeneous populations. *Transportation Science*, 44(4) 537-549.

Bernardo, C., Paleti, R., Hoklas, M., and Bhat, C.R. (2015). An empirical investigation into the time-use and activity patterns of dual-earner couples with and without young children. *Transportation Research Part A*, 76, 71-91.

Berry, S.T., and Haile, P.A., (2010). Nonparametric identification of multinomial choice demand models with heterogeneous consumers. Cowles Foundation Discussion Paper 1718, Cowles Foundation, Yale University. Available at: http://cowles.econ.yale.edu/P/cd/d17a/d1718.pdf.

Bhat, C.R. (1997). An endogenous segmentation mode choice model with an application to intercity travel. *Transportation Science*, 31(1), 34-48.

Bhat, C.R. (2000). A multi-level cross-classified model for discrete response variables, *Transportation Research Part B*, 34(7), 567-582.

Bhat, C. R. (2001). Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B*, 35(7), 677-693.

Bhat, C.R. (2005). A multiple discrete-continuous extreme value model: Formulation and application to discretionary time-use decisions. *Transportation Research Part B,* 39(8), 679-707.

Bhat, C.R. (2008). The multiple discrete-continuous extreme value (MDCEV) model: Role of utility function parameters, identification considerations, and model extensions. *Transportation Research Part B,* 42(3), 274-303.

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

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

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

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

Bhat, C.R., and Koppelman, F.S. (1993). A conceptual framework of individual activity program generation, *Transportation Research Part A,* 27(6), 433-446.

Bhat, C. R., and Sen, S., (2006). Household vehicle type holdings and usage: an application of the multiple discrete-continuous extreme value (MDCEV) model. *Transportation Research Part B* 40(1), 35-53.

Bhat, C.R., and Sidharthan, R., (2012). A new approach to specify and estimate non-normally mixed multinomial probit models. *Transportation Research Part B* 46(7), 817-833.

Bhat, C.R., and Singh, S.K. (2000). A comprehensive daily activity-travel generation model system for workers. *Transportation Research Part A,* 34(1), 1-22.

Bhat, C. R., and Srinivasan, S., (2005). A multidimensional mixed ordered-response model for analyzing weekend activity participation. *Transportation Research Part B* 39(3), 255-278.

Bhat, C.R., and Steed, J.L. (2002). A continuous-time model of departure time choice for urban shopping trips. *Transportation Research Part B,* 36(3), 207-224.

Bhat, C.R., and Gossen, R., (2004). A mixed multinomial logit model analysis of weekend recreational episode type choice. *Transportation Research Part B,* 38(9), 767-787.

Bhat, C.R., and R. Sidharthan (2012). A new approach to specify and estimate non-normally mixed multinomial probit models. *Transportation Research Part B,* 46, 817-833.

Bhat, C.R., A. Govindarajan, and V. Pulugurta (1998). Disaggregate attraction-end choice modeling. *Transportation Research Record*, 1645, 60-68.

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., Guo, J.Y., Srinivasan, S., and Sivakumar, A. (2004). Comprehensive econometric microsimulator for daily activity-travel patterns. *Transportation Research Record: Journal of the Transportation Research Board*, 1894, 57-66.

Bhat, C.R., Castro, M., and Khan, M., (2013a). A new estimation approach for the multiple discrete-continuous probit (MDCP) choice model. *Transportation Research Part B*, 55, 1-22.

Bhat, C. R., Dubey, S. K., Alam, M. J. B., and Khushefati, W. H., (2015). A new spatial multiple discrete-continuous modeling approach to land use change analysis. *Journal of Regional Science* 55(5), 801-841.

Bhat, C.R., Paleti, R., Pendyala, R.M., Lorenzini, K., and Konduri, K.C. (2013b). Accommodating immigration status and self-selection effects in a joint model of household auto ownership and residential location choice. *Transportation Research Record: Journal of the Transportation Research Board*, 2382(1), 142-150.

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

Bhat, C. R., Astroza, S., and Bhat, A. C., (2016a). On allowing a general form for unobserved heterogeneity in the multiple discrete–continuous probit model: Formulation and application to tourism travel. *Transportation Research Part B* 86, 223-249.

Bhat, C. R., Astroza, S., Bhat, A. C., and Nagel, K., (2016b). Incorporating a multiple discrete-continuous outcome in the generalized heterogeneous data model: Application to residential self-selection effects analysis in an activity time-use behavior model. *Transportation Research Part B* 91, 52-76.

Bhat, C.R., Pinjari, A.R., Dubey, and S.K., Hamdi, A., (2016c). On accommodating spatial interactions in a generalized heterogeneous data model (GHDM) of mixed types of dependent variables. *Transportation Research Part B* 94, 240-263.

Bhat, C.R., Astroza, S., and Hamdi, A., (2017). A spatial generalized ordered-response model with skew normal kernel error terms with an application to bicycling frequency, *Transportation Research Part B* 95, 126-148.

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

Bolduc, D., Ben-Akiva, M., Walker, J., Michaud, A. (2005). Hybrid choice models with logit kernel: applicability to large scale models. In *Integrated Land-Use and Transportation Models: Behavioral Foundations*, Lee-Gosselin, M., Doherty, S. (eds.), Elsevier, Oxford, 275-302.

Born, K., Yasmin, S., You, D., Eluru, N., Bhat, C., and Pendyala, R. (2014). Joint model of weekend discretionary activity prticipation and episode duration. *Transportation Research Record: Journal of the Transportation Research Board*, 2413, 34-44.

Boussauw, K., Neutens, T., and Witlox, F. (2012). Relationship between spatial proximity and travel-to-work distance: the effect of the compact city. *Regional Studies*, 46(6), 687-706.

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.

Bujosa, A., A. Riera and R. L. Hicks (2010). Combining discrete and continuous representations of preference heterogeneity: a latent class approach. *Environmental and Resource Economics*, 47(4), 477-493.

Caffo, B., An, M.W., and Rohde, C., (2007). Flexible random intercept models for binary outcomes using mixtures of normals. *Computational Statistics & Data Analysis* 51(11), 5220-5235.

Campbell, D., E. Doherty, S. Hynes and T. Van Rensburg (2010). Combining discrete and continuous mixing approaches to accommodate heterogeneity in price sensitivities in environmental choice analysis. *84th Agricultural Economics Society Annual Conference*, March 29-31, Edinburgh, Scotland.

Cao, X., and Chatman, D., (2016). How will smart growth land-use policies affect travel? A theoretical discussion on the importance of residential sorting. *Environment and Planning B: Planning and Design* 43(1), 58-73.

Cao, X., and Fan, Y. (2012). Exploring the influences of density on travel behavior using propensity score matching. *Environment and Planning B*, 39(3), 459-470.

Castro, M., Eluru, N., Bhat, C., and Pendyala, R. (2011). Joint model of participation in nonwork activities and time-of-day choice set formation for workers. *Transportation Research Record: Journal of the Transportation Research Board*, 2254, 140-150.

Castro, M., C.R. Bhat, R.M. Pendyala and S.R. Jara-Diaz (2012a). Accommodating multiple constraints in the multiple discrete-continuous extreme value (MDCEV) choice model. *Transportation Research Part B*, 46(6), 729-743.

Castro, M., Paleti, R., and Bhat, C.R. (2012b). 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.

Chen, C., Gong, H., and Paaswell, R. (2008). Role of the built environment on mode choice decisions: additional evidence on the impact of density. *Transportation*, 35(3), 285-299.

Chen, C., Mei, Y., and Liu, Y. (2014). Does distance still matter in facilitating social ties? The roles of mobility patterns and the built environment. Presented at the 93rd Annual Meeting Transportation Research Board, Washington, D.C, January.

Cherchi, E., C. Cirillo and J. Polak (2009). User benefit assessment in presence of random taste heterogeneity: comparison between parametric and nonparametric models. *Transportation Research Record*, 2132, 78-86.

Cleaver, M., and Muller, T.E. (2001). I want to pretend I'm eleven years younger: Subjective age and seniors' motives for vacation travel. *Social Indicators Research*, 60 (1/3), 227-241.

Coleman, L. (2009). Being alone together: From solidarity to solitude in urban anthropology. *Anthropological Quarterly*, 82(3), 755-777.

Council, P. S. R. (2005). Vision 2020+ 20 Update: Issue Paper on Environmental Planning. The Council.

Cox, D.R., and Reid, N., (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika* 91(3), 729-737.

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

Daziano, R.A., and Bolduc, D. (2013). Incorporating pro-environmental preferences towards green automobile technologies through a Bayesian hybrid choice model. *Transportmetrica A: Transport Science*, 9(1), 74-106.

de Abreu e Silva, J., Goulias, K., and Dalal, P. (2012). Structural equations model of land use patterns, location choice, and travel behavior in Southern California. *Transportation Research Record: Journal of the Transportation Research Board*, 2323, 35-45.

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

De Leon, A.R., and Chough, K.C. (Eds.). (2013). *Analysis of Mixed Data: Methods & Applications*. CRC Press.

Deng, Y., and Srinivasan, S., (2016). Urban land use change and regional access: A case study in Beijing, China. *Habitat International* 51, 103-113.

Diekmann, A., and Franzen, A. (1999). The wealth of nations and environmental concern. *Environment and Behavior*, 31(4), 540-549.

Ding, D., Adams, M.A., Sallis, J.F., Norman, G.J., Hovell, M.F., Chambers, C.D., ... and Gomez, L.F. (2013). Perceived neighborhood environment and physical activity in 11 countries: Do associations differ by country. *International Journal of Behavioral Nutrition and Physical Activity*, 10(1), 57.

Eluru, N., Bhat, C. R., Pendyala, R. M., and Konduri, K. C., (2010). A joint flexible econometric model system of household residential location and vehicle fleet composition/usage choices. *Transportation* 37(4), 603-626.

Farber, S., Páez, A., Mercado, R. G., Roorda, M., and Morency, C. (2011). A time-use investigation of shopping participation in three Canadian cities: Is there evidence of social exclusion? *Transportation*, 38(1), 17-44.

Ferdous, N., R.M. Pendyala, C.R. Bhat, and K.C. Konduri (2011) Modeling the Influence of Family, Social Context, and Spatial Proximity on Use of Nonmotorized Transport Mode. *Transportation Research Record*, 2230, 111-120.

Fonseca, J. R. (2010). On the performance of information criteria in latent segment models. *World Academy of Science, Engineering and Technology*, 63.

Forsyth, A., Oakes, J.M., Lee, B., and Schmitz, K.H. (2009). The built environment, walking, and physical activity: Is the environment more important to some people than others?. *Transportation Research Part D*, 14(1), 42-49.

Fosgerau, M. (2005). Unit income elasticity of the value of travel time savings. Presented at 8th NECTAR Conference, Las Palmas G.C., June 2-4.

Fotheringham, A.S. (1983). Some theoretical aspects of destination choice and their relevance to production-constrained gravity models. *Environment and Planning A*, 15(8), 1121-1132.

Frühwirth-Schnatter, S. (2011). Label switching under model uncertainty. *Mixtures: Estimation and Application*, 213-239.

Geweke, J. and M. Keane (1999) Mixture of normals probit models. In C. Hsiao, M.H. Pesaran, K.L. Lahiri, and L.F. Lee (eds.) Analysis of Panel and Limited Dependent Variables, 49-78, Cambridge University Press, Cambridge.

Gliebe, J.P., and Koppelman, F.S. (2002). A model of joint activity participation between household members. *Transportation*, 29(1), 49-72.

Godambe, V.P. (1960). An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics* 31(4), 1208-1211.

Golob, T.F. (2003). Structural equation modeling for travel behavior research. *Transportation Research Part B*, 37(1), 1-25.

Greene, W.H. and D.A. Hensher (2003). A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B*, 37(8), 681-698.

Greene, W.H., and Hensher, D.A., (2010). *Modeling Ordered Choices: A Primer*. Cambridge University Press, Cambridge.

Greene, W.H. and D.A. Hensher (2013). Revealing additional dimensions of preference heterogeneity in a latent class mixed multinomial logit model. *Applied Economics*, 45(14), 1897-1902.

Greene W.H., D.A. Hensher and J.M. Rose (2006). Accounting for heterogeneity in the variance of the unobserved effects in mixed logit models (NW transport study data). *Transportation Research Part B*, 40(1), 75-92.

Grigolon, A., Kemperman, A., and Timmermans, H. (2013). Mixed multinomial logit model for out-of-home leisure activity choice. *Transportation Research Record: Journal of the Transportation Research Board*, 2343, 10-16.

Habib, K. M., and Miller, E. J., (2008). Modelling daily activity program generation considering within-day and day-to-day dynamics in activity-travel behaviour. *Transportation* 35(4), 467.

Hanemann, W.M. (1978). A methodological and empirical study of the recreation benefits from water quality improvement. Ph.D. dissertation, Department of Economics, Harvard University.

Hanemann, W.M., L. Pendleton, C. Mohn, J. Hilger, K. Kuriyama, D. Layton, C. Busch, and F. Vasquez (2004). Using revealed preference models to estimate the effect of coastal water quality on beach choice in Southern California. University of California at Berkeley, report to the U.S. National Oceanic and Atmospheric Administration.

Heagerty, P.J., and Lumley, T., (2000). Window subsampling of estimating functions with application to regression models. *Journal of the American Statistical Association* 95(449), 197-211.

Hensher, D.A., J.M. Rose and W.H. Greene (2005). *Applied Choice Analysis: A Primer*. Cambridge University Press, Cambridge, U.K.

Hwang, J., and Han, H. (2014). Examining strategies for maximizing and utilizing brand prestige in the luxury cruise industry. *Tourism Management*, 40, 244-259.

Irwin, E.G., and Geoghegan, J., (2001). Theory, data, methods: Developing spatially explicit economic models of land use change. *Agriculture, Ecosystems & Environment* 85(1-3), 7–24.

Iso-Ahola, S.E. (1983). Towards a social psychology of recreational travel. *Leisure Studies*, 2(1), 45-56.

Katsikatsou, M., Moustaki, I., Yang-Wallentin, F., and Jöreskog, K.G. (2012). Pairwise likelihood estimation for factor analysis models with ordinal data. *Computational Statistics & Data Analysis*, 56(12), 4243-4258.

Kaza, N., Towe, C., and Ye., X., (2012). A hybrid land conversion model incorporating multiple end uses. *Agricultural and Resource Economics Review* 40(3), 341–359.

Keane, M.P. (1992). A note on identification in the multinomial probit model. *Journal of Business & Economic Statistics*, 10(2), 193-200.

Kim, J., G.M. Allenby and P.E. Rossi (2002). Modeling consumer demand for variety. *Marketing Science*, 21(3), 229-250.

Kim, H.Y., and Chung, J.E. (2011). Consumer purchase intention for organic personal care products. *Journal of Consumer Marketing*, 28(1), 40-47.

Kim, J., and Brownstone, D. (2013). The impact of residential density on vehicle usage and fuel consumption: Evidence from national samples. *Energy Economics*, 40, 196-206.

Kuriyama, K., W.M. Hanemann and J.R. Hilger (2010). A latent segmentation approach to a Kuhn-Tucker model: an application to recreation demand. *Environmental Economics and Management*, 60(3), 209-220.

Kuriyama, K., Y. Shoji, and T. Tsuge (2011). Estimating value of mortality risk reduction using the Kuhn-Tucker model: an application to recreation demand. Working Paper, Graduate School of Agriculture, Kyoto University.

LaMondia, J.J., and C.R. Bhat (2012). A conceptual and methodological framework of leisure activity loyalty accommodating the travel context. *Transportation*, 39(2), 321-349.

LaMondia, J.J., C.R. Bhat, and D.A. Hensher (2008). An annual time use model for domestic vacation travel. *Journal of Choice Modelling*, 1(1), 70-97.

LaMondia, J.J., T. Snell, and C.R. Bhat (2010). Traveler behavior and values analysis in the context of vacation destination and travel mode choices: European Union case study. *Transportation Research Record*, 2156, 140-149.

La Paix, L., Bierlaire, M., Cherchi, E., and Monzón, A. (2013). How urban environment affects travel behaviour: integrated choice and latent variable model for travel schedules. In *Choice Modelling: The State of the Art and the State of Practice*, chapter 10, 211-228, Hess, S., Daly, A. (Eds.), Edward Elgar Publishing Ltd.

Lee, S.X., and McLachlan, G.J., (2013). On mixtures of skew normal and skew t-distributions. *Advances in Data Analysis and Classification* 7(3), 241-266.

Li, G., Li, G., and Kambele, Z. (2012). Luxury fashion brand consumers in China: Perceived value, fashion lifestyle, and willingness to pay. *Journal of Business Research*, 65(10), 1516-1522.

Lindsay, B.G., (1988). Composite likelihood methods. *Contemporary Mathematics* 80, 221-239.

Liu, Y., Hu, Y., Long, S., Liu, L., and Liu, X., (2017). Analysis of the effectiveness of urban land-use-change models based on the measurement of spatio-temporal, dynamic urban growth: A cellular automata case study. *Sustainability* 9(5), 796.

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

Ma, L., and Srinivasan, S., (2010). Impact of individuals' immigrant status on household auto ownership. *Transportation Research Record: Journal of the Transportation Research Board*, 2156, 36-46.

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

McCormack, G.R., Shiell, A., Doyle-Baker, P.K., Friedenreich, C.M., and Sandalack, B.A. (2014). Subpopulation differences in the association between neighborhood urban form and neighborhood-based physical activity. *Health & Place*, 28, 109-115.

McMillen, D.P., (2010). Issues in spatial analysis. *Journal of Regional Science* 50(1), 119–141.

McMillen, D.P., (2012). Perspectives on spatial econometrics: Linear smoothing with structured models. *Journal of Regional Science* 52(2), 192-209.

Ministry of Business, Innovation and Employment (2008). Tourism sector profile. Available at: http://www.med.govt.nz/sectors-industries/tourism/tourism-research-data/other-research-and-reports/sector-profiles

Ministry of Business, Innovation and Employment (2013). Statement of Intent. May 2013. Available at: http://www.mbie.govt.nz/about-us/publications/soi/Ministry-of-Business-Innovation-and-Employment-SOI-2013.pdf

Mittelhammer, R. C., and G. Judge (2011). A family of empirical likelihood functions and estimators for the binary response model. *Journal of Econometrics*, 164(2), 207-217.

Mokhtarian, P.L., and Cao, X. (2008). Examining the impacts of residential self-selection on travel behavior: A focus on methodologies. *Transportation Research Part B*, 42(3), 204-228.

Molenberghs, G., and Verbeke, G., (2005). *Models for Discrete Longitudinal Data*. Springer.

Munkin, M.K., and Trivedi, P.K. (2008). Bayesian analysis of the ordered probit model with endogenous selection. *Journal of Econometrics*, 143(2), 334-348.

Neath, A. A. and Cavanaugh, J. E. (2012). The Bayesian information criterion: Background, derivation, and applications. *WIREs Computational Statistics* 4, 199-203.

New Zealand Tourism Strategy (2015). Available at: http://www.tianz.org.nz/content/library/FINAL_NZTS2015_HR.pdf

Paleti, R., Bhat, C.R., and Pendyala, R.M. (2013). 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.

Paleti, R., Copperman, R. B., and Bhat, C. R., (2011). An empirical analysis of children's after school out-of-home activity-location engagement patterns and time allocation. *Transportation* 38(2), 273-303.

Peters, M., and B.R. Clarkson (2010). Wetland restoration: a handbook for New Zealand freshwater system. Lincoln, N.Z: Manaaki Whenua Press. ISBN 978-0-478-34706-7.

Pinjari, A.R. (2011). Generalized extreme value (GEV)-based error structures for multiple discrete-continuous choice models. *Transportation Research Part B*, 45(3), 474-489.

Pinjari, A.R., and C.R. Bhat (2011a). An efficient forecasting procedure for Kuhn-Tucker consumer demand model systems: application to residential energy consumption analysis. Technical paper, Department of Civil and Environmental Engineering, University of South Florida.

Pinjari, A.R., and Bhat, C.R. (2011b). Activity based travel demand analysis. In *A Handbook of Transport Economics*, Chapter 10, 213-248, de Palma, A., Lindsey, R., Quinet, E., and Vickerman, R. (Eds.), Edward Elgar Publishing Ltd.

Pinjari, A.R., and Bhat, C.R. (2014). Computationally efficient forecasting procedures for Kuhn-Tucker consumer demand model systems: application to residential energy consumption analysis. Technical paper, Department of Civil and Environmental Engineering. University of South Florida. Avaliable at: http://www.caee.utexas.edu/prof/bhat/ABSTRACTS/Pinjari_Bhat_MDCEV_Forecasting_July21_2011.pdf.

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

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.

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.

Pinkse, J., and Slade, M., (2010). The future of spatial econometrics. *Journal of Regional Science* 50(1), 102–118.

Potoglou, D., and Susilo, Y.O., (2008). Comparison of vehicle-ownership models. *Transportation Research Record: Journal of the Transportation Research Board*, 2076, 97-105.

Rajagopalan, B.S., Pinjari, A.R., and Bhat, C.R. (2009). Comprehensive model of worker nonwork-activity time use and timing behavior.*Transportation Research Record: Journal of the Transportation Research Board*, 2134, 51-62.

Reilly, T., and O'Brien, R.M. (1996). Identification of confirmatory factor analysis models of arbitrary complexity the side-by-side rule. *Sociological Methods & Research*, 24(4), 473-491.

Romans, S., Cohen, M., and Forte, T. (2011). Rates of depression and anxiety in urban and rural Canada. *Social Psychiatry and Psychiatric Epidemiology*, 46(7), 567-575.

Rosecky, R.B., and King, A.B. (1996). Perceptual differences among owners of luxury cars: strategic marketing implications. *The Mid-Atlantic Journal of Business*, 32(3), 221.

RSG (2014). Puget Sound Regional Travel Study. Available at: http://www.psrc.org/assets/12060/2014-Household-Survey-Tech-Memo.pdf.

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

Sobhani, A., N. Eluru and A. Faghih-Imani (2013). A latent segmentation based multiple discrete continuous extreme value model. *Transportation Research Part B*, 58, 154-169.

Stapleton, D.C. (1978). Analyzing political participation data with a MIMIC model. *Sociological Methodology*, 9, 52-74.

Statistics New Zealand (2013). Tourism Satellite Account. Tatauranga Aotearoa, Wellington, New Zealand.

Teixeira-Pinto, A., and Harezlak, J. (2013). Factorization and latent variable models for joint analysis of binary and continuous outcomes. In *Analysis of Mixed Data: Methods & Applications*, De Leon, A.R., and Chough, K.C. (Eds.), Chapter 6, 81-92, CRC Press, Taylor & Francis Group, Boca Raton, FL.

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

Torres, C., N. Hanley and A. Riera (2011). How wrong can you be? Implications of incorrect utility function specification for welfare measurement in choice experiments. *Journal of Environmental Economics and Management*, 62(1), 111-121.

Tourism Industry Association (2012). New Zealand Tourism Industry Association Annual Report. Available at: https://www.tianz.org.nz/content/library/TIA_Annual_Report_1112v11_ONLINE.pdf

Train, K.E. (1998). Recreation demand models with taste variation. *Land Economics*, 74, 230-239.

Train, K.E. (2008). EM algorithms for nonparametric estimation of mixing distributions. *Journal of Choice Modelling*, 1(1), 40-69.

Train, K.E. and G. Sonnier (2005). Mixed logit with bounded distributions of correlated partworths. In *Applications of Simulation Methods in Environmental and Resource Economics*, Scarpa, R., Alberini, A., (eds.), Ch. 7, pp. 117-134, Springer, Dordrecht, The Netherlands.

Twitchell, J.B. (2013). *Living It Up: Our love affair with luxury*. Columbia University Press.

Van Acker, V., Mokhtarian, P.L., and Witlox, F. (2014). Car availability explained by the structural relationships between lifestyles, residential location, and underlying residential and travel attitudes. *Transport Policy*, 35, 88-99.

Van Nostrand, C., V. Sivaraman, and A. R. Pinjari (2013). Analysis of long-distance vacation travel demand in the United States: a multiple discrete-continuous choice framework. *Transportation*, 40(1), 151-171.

van Vliet, J., Bregt, A. K., Brown, D. G., van Delden, H., Heckbert, S., and Verburg, P. H., (2016). A review of current calibration and validation practices in land-change modeling. *Environmental Modelling & Software* 82, 174-182.

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

Varin, C., and Czado, C., (2010). A mixed autoregressive probit model for ordinal longitudinal data. *Biostatistics* 11(1), 127-138.

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

von Haefen, R.H. (2007). Empirical strategies for incorporating weak complementarity into consumer demand models. *Journal of Environmental Economics and Management*, 54(1), 15-31.

von Haefen, R.H., and D.J. Phaneuf (2003). Estimating preferences for outdoor recreation: a comparison of continuous and count data demand system frameworks. *Journal of Environmental Economics and Management*, 45, 612-630.

Wafa, Z., C.R. Bhat, R.M. Pendyala, and V.M. Garikapati (2015). Latent-segmentation-based approach to investigating spatial transferability of activity-travel models. *Transportation Research Record*, 2493, 136-144.

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

Wang, X., Grengs, J., and Kostyniuk, L. (2013). Visualizing travel patterns with a GPS dataset: How commuting routes influence non-work travel behavior. *Journal of Urban Technology*, 20(3), 105-125.

Wendel-Vos, W., Droomers, M., Kremers, S., Brug, J., and Van Lenthe, F. (2007). Potential environmental determinants of physical activity in adults: a systematic review. *Obesity Reviews*, 8(5), 425-440.

White, R. (2011). Is the staycation trend a real phenomenon? White paper, White Hutchinson Leisure & Learning Group, Kansas City, MO, January. Available at: http://www.whitehutchinson.com/leisure/articles/Staycation.shtml.

Whitehead, J.C., D.J. Phaneuf, C.F. Dumas, J. Herstine, J. Hill, and B. Buerger (2010). Convergent validity of revealed and stated recreation behavior with quality change: a comparison of multiple and single site demands. *Environmental and Resource Economics*, 45(1), 91-112.

Xiong, Y. and F.L. Mannering (2013). The heterogeneous effects of guardian supervision on adolescent driver-injury severities: A finite-mixture random-parameters approach. *Transportation Research Part B*, 49, 39-54.

Xu, X., and Reid, N., (2011). On the robustness of maximum composite likelihood estimate. *Journal of Statistical Planning and Inference* 141(9), 3047-3054.

Yi, G.Y., Zeng, L., and Cook, R.J., (2011). A robust pairwise likelihood method for incomplete longitudinal binary data arising in clusters. *Canadian Journal of Statistics* 39(1), 34-51.

Zhao, Y., and Joe, H. (2005). Composite likelihood estimation in multivariate data analysis. *Canadian Journal of Statistics*, 33(3), 335-356.

Zhuge, C., Shao, C., Gao, J., Dong, C., and Zhang, H., (2016). Agent-based joint model of residential location choice and real estate price for land use and transport model. *Computers, Environment and Urban Systems* 57, 93-105.

Appendixes

APPENDIX A: DEVELOPMENT OF THE LIKELIHOOD FUNCTION FOR THE FDMN MDCP MODEL

To develop the likelihood function, define $\mathbf{M}_q$ as an identity matrix of size *K*–1 with an extra column of "–1" values added at the m_q^{th} column. Also, stack y_{qgk} and V_{qgk} into *K*×1 vectors $\boldsymbol{y}_{qg} = (y_{qg1}, y_{qg2},..., y_{qgK})'$ and $\boldsymbol{V}_{qg} = (V_{qg1}, V_{qg2},..., V_{qgK})'$ respectively, and let $\mathbf{z}_q = (\boldsymbol{z}_{q1}, \boldsymbol{z}_{q2},..., \boldsymbol{z}_{qK})'$ be a *K*×*D* matrix of variable attributes. Then, we may write, in matrix notation, $\boldsymbol{y}_{qg} = \boldsymbol{V}_{qg} + \mathbf{z}_q\tilde{\boldsymbol{\beta}}_{qg}$ and $\boldsymbol{y}^*_{qg} = \mathbf{M}_q\boldsymbol{y}_{qg} \sim MVN_{K-1}(\mathbf{H}_{qg,}\boldsymbol{\Psi}_{qg})$, where $\mathbf{H}_{qg} = \mathbf{M}_q\boldsymbol{V}_{qg}$ and $\boldsymbol{\Psi}_{qg} = \mathbf{M}_q\,\mathbf{z}_q\boldsymbol{\Omega}_g\,\mathbf{z}'_q\,\mathbf{M}'_q$. Next, partition the vector $\boldsymbol{y}^*_{qg}$ into a sub-vector $\tilde{\boldsymbol{y}}^*_{qg,NC}$ of length $L_{q,NC}\times 1$ $(0 \le L_{q,NC} \le K-1)$ for the non-consumed goods, and another sub-vector $\tilde{\boldsymbol{y}}^*_{qg,C}$ of length $L_{q,C}\times 1$ $(0 \le L_{q,C} \le K-1)$ for the consumed goods $(L_{q,NC} + L_{q,C} = K-1)$. Let $\tilde{\boldsymbol{y}}^*_{qg} = \left(\left[\tilde{\boldsymbol{y}}^*_{qg,NC}\right]', \left[\tilde{\boldsymbol{y}}^*_{qg,C}\right]'\right)'$, which may be obtained from $\boldsymbol{y}^*_{qg}$ as $\tilde{\boldsymbol{y}}^*_{qg} = \mathbf{R}_q\boldsymbol{y}^*_{qg}$, where $\mathbf{R}_q$ is a re-arrangement matrix of dimension (*K*–1)×(*K*–1) with zeros and ones. For example, consider a consumer *q* who chooses among five goods (*K*=5), and selects goods 2, 3, and 5 for consumption. Thus, $m_q = 2$, $L_{q,NC} = 2$ (corresponding to the non-consumed goods 1 and 4), and $L_{q,C} = 2$ (corresponding to the consumed goods 3 and 5, with good 2 serving as the base good needed to take utility differentials). Then, the re-arrangement matrix $\mathbf{R}_q$ (for goods 1, 3, 4, and 5) is provided in Equation (A.1):

$$\mathbf{R}_q = \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right] = \left[\begin{array}{c} \mathbf{R}_{q,NC} \\ \hline \mathbf{R}_{q,C} \end{array}\right], \tag{A.1}$$

where the upper sub-matrix $\mathbf{R}_{q,NC}$ corresponds to the non-consumed goods (of dimension $L_{q,NC}\times(K-1)$) and the lower sub-matrix $\mathbf{R}_{q,C}$ corresponds to the consumed goods (of dimension $L_{q,C}\times(K-1)$). Note also that $\tilde{\boldsymbol{y}}^*_{qg,NC} = \mathbf{R}_{q,NC}\boldsymbol{y}^*_{qg}$ and $\tilde{\boldsymbol{y}}^*_{qg,C} = \mathbf{R}_{q,C}\boldsymbol{y}^*_{qg}$. $\mathbf{R}_{q,NC}$ has as many rows as the number of non-consumed alternatives and as many columns as the number of alternatives minus one (each column corresponds to an alternative, except the m_q^{th} alternative).

Then, for each row, $\mathbf{R}_{q,NC}$ has a value of "1" in one of the columns corresponding to an alternative that is not consumed, and the value of "0" everywhere else. A similar construction is involved in creating the $\mathbf{R}_{q,C}$ matrix.

Consistent with the above re-arrangement, define $\tilde{\mathbf{H}}_{qg} = \mathbf{R}_q \mathbf{H}_{qg}$, $\tilde{\mathbf{H}}_{qg,NC} = \mathbf{R}_{q,NC} \mathbf{H}_{qg}$, $\tilde{\mathbf{H}}_{qg,C} = \mathbf{R}_{q,C} \mathbf{H}_{qg}$, and $\tilde{\boldsymbol{\Psi}}_{qg} = \mathbf{R}_q \boldsymbol{\Psi}_{qg} \mathbf{R}_q' = \begin{bmatrix} \tilde{\boldsymbol{\Psi}}_{qg,NC} & \tilde{\boldsymbol{\Psi}}'_{qg,NC,C} \\ \tilde{\boldsymbol{\Psi}}_{qg,NC,C} & \tilde{\boldsymbol{\Psi}}_{qg,C} \end{bmatrix}$, where $\tilde{\boldsymbol{\Psi}}_{qg,NC} = \mathbf{R}_{q,NC} \boldsymbol{\Psi}_{qg} \mathbf{R}'_{q,NC}$, $\tilde{\boldsymbol{\Psi}}_{qg,C} = \mathbf{R}_{q,C} \boldsymbol{\Psi}_{qg} \mathbf{R}'_{q,C}$, and $\tilde{\boldsymbol{\Psi}}_{qg,NC,C} = \mathbf{R}_{q,NC} \boldsymbol{\Psi}_{qg} \mathbf{R}'_{q,C}$. Then, the likelihood function corresponding to the consumption quantity vector $\boldsymbol{x}_q^*$ for consumer *q* may be obtained from the KKT conditions in Equation (2.6), provided as Equation (A.2):

$$L_{qg} = \det(\mathbf{J}_{qg}) \int_{\mathbf{h}_{q,NC}=-\infty}^{\mathbf{0}} f_{K-1}\left(\boldsymbol{h}_{q,NC}, \mathbf{0}_{L_{q,C}} \mid \tilde{\mathbf{H}}_{qg}, \tilde{\boldsymbol{\Psi}}_{qg}\right) d\boldsymbol{h}_{q,NC}, \tag{A.2}$$

where $\det(\mathbf{J}_{qg})$ is the determinant of the Jacobian of the transformation from $\boldsymbol{y}_{qg}^*$ to the consumption quantities $\boldsymbol{x}_q^*$ (see Bhat, 2008), as Equation (A.3) indicates:

$$\det(\mathbf{J}_{qg}) = \left\{ \prod_{k \in \mathfrak{C}_q} \frac{1-\alpha_{qgk}}{x_{qk}^* + \gamma_{qgk}} \right\} \left\{ \sum_{k \in \mathfrak{C}_q} \left(\frac{x_{qk}^* + \gamma_{qgk}}{1-\alpha_{qgk}} \right) \left(\frac{p_{qk}}{p_{qm_q}} \right) \right\}, \tag{A.3}$$

where $\mathfrak{C}_q$ is the set of goods consumed by consumer *q* (including good m_q).

APPENDIX B: GHDM ESTIMATION INCLUDING MDC VARIABLES

Let $E = (H + N + 1)$. Define $\ddot{\boldsymbol{y}} = \left(\boldsymbol{y}', \left[\widetilde{\boldsymbol{y}}^*\right]', \left[\breve{\boldsymbol{y}}^*\right]' \right)'$ [$E \times 1$ vector],

$\ddot{\boldsymbol{\gamma}} = (\boldsymbol{\gamma}', \widetilde{\boldsymbol{\gamma}}', \mathbf{0}_{\mathbf{A}})'$ [$E \times A$ matrix], $\ddot{\boldsymbol{d}} = (\boldsymbol{d}', \widetilde{\boldsymbol{d}}', \breve{\boldsymbol{d}}')'$ [$E \times L$ matrix], and

$\ddot{\boldsymbol{\varepsilon}} = (\boldsymbol{\varepsilon}', \widetilde{\boldsymbol{\varepsilon}}', \breve{\varepsilon}')'$ ($E \times 1$ vector), where $\mathbf{0}_{\mathbf{A}}$ is a vector of zeros of dimension $A \times 1$. We will assume that the error vectors $\boldsymbol{\eta}$, $\ddot{\boldsymbol{\varepsilon}}$, $\boldsymbol{\xi}$, and $\boldsymbol{\varsigma}$ are independent of each other. While not strictly necessary (and can be relaxed in a very straightforward manner within the estimation framework of our model system as long as the resulting model is identified), the assumption aids in developing general sufficiency conditions for identification of parameters in a mixed model when the latent variable vector $\boldsymbol{z}^*$ already provides a mechanism to generate covariance among the mixed outcomes. Further, define the following: $\boldsymbol{yu} = (\ddot{\boldsymbol{y}}', \boldsymbol{u}', \widetilde{\boldsymbol{u}}')'$, $\widetilde{\boldsymbol{V}} = [(\ddot{\boldsymbol{\gamma}}\boldsymbol{x})', (\boldsymbol{b}\boldsymbol{x})', \boldsymbol{V}']'$, $\boldsymbol{\pi} = (\ddot{\boldsymbol{d}}', \boldsymbol{\varpi}', \boldsymbol{\mu}')'$, and $\kappa = (\ddot{\boldsymbol{\varepsilon}}', \boldsymbol{\varsigma}', \boldsymbol{\xi}')'$. Then, we may write the continuous (observed or latent) components of the structural and the measurement equations of the model system compactly as:

$$\boldsymbol{z}^* = \boldsymbol{\alpha}\boldsymbol{w} + \boldsymbol{\eta} \tag{B.1}$$

$$\boldsymbol{yu} = \widetilde{\boldsymbol{V}} + \boldsymbol{\pi}\boldsymbol{z}^* + \boldsymbol{\kappa},$$

$$\text{with } \text{Var}(\boldsymbol{\kappa}) = \ddot{\boldsymbol{\Sigma}} = \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{IDEN}_{\mathbf{N}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{M}\boldsymbol{\Lambda}\mathbf{M}' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Omega} \end{bmatrix} (E + I + K - 2) \times (E + I + K - 2) \text{ matrix}$$

To develop the reduced form equations, replace the right side for $\boldsymbol{z}^*$ in the second part of Equation (B.1) to obtain the following system:

$$\boldsymbol{yu} = \widetilde{\boldsymbol{V}} + \boldsymbol{\pi}\boldsymbol{z}^* + \boldsymbol{\kappa} = \widetilde{\boldsymbol{V}} + \boldsymbol{\pi}(\boldsymbol{\alpha}\boldsymbol{w} + \boldsymbol{\eta}) + \boldsymbol{\kappa} = \widetilde{\boldsymbol{V}} + \boldsymbol{\pi}\boldsymbol{\alpha}\boldsymbol{w} + \boldsymbol{\pi}\boldsymbol{\eta} + \boldsymbol{\kappa}. \tag{B.2}$$

Then $\boldsymbol{yu} \sim \boldsymbol{MVN}_{E+I+K-2}(\boldsymbol{B}, \boldsymbol{\Theta})$, where $\boldsymbol{B} = \widetilde{\boldsymbol{V}} + \boldsymbol{\pi}\boldsymbol{\alpha}\boldsymbol{w}$, and $\boldsymbol{\Theta} = \boldsymbol{\pi}\boldsymbol{\Gamma}\boldsymbol{\pi}' + \ddot{\boldsymbol{\Sigma}}$.

The question of identification relates to whether all the elements in the model system are estimable from the elements of $\boldsymbol{B}$ and $\boldsymbol{\Theta}$. One may analyze this by starting from Stapleton's (1978) sufficiency conditions for multiple-indicator multiple-cause (MIMIC) models. Conforming with the set-up of Stapleton and earlier MIMIC models, we will assume that the

number of measurement equations without the nominal and non-MDC variables exceeds the number of latent factors. Then, sufficiency conditions may be developed for the GHDM-MDC model following the same line of argument as in Bhat *et al*. (2014) for the GHDM. In particular, all parameters are estimable under the following conditions: (1) diagonality is maintained across the elements of the error term vector $\ddot{\varepsilon}$ (that is, $\ddot{\boldsymbol{\Sigma}}$ is diagonal), (2) $\boldsymbol{\Gamma}$ in the structural equation is specified to be a correlation matrix, (3) for each latent variable, there is at least one outcome variable that loads only on that latent variable and no other latent variable (that is, there is at least one factor complexity one outcome variable for each latent variable) (see also Reilly and O'Brien, 1996), (4) the element corresponding to the effect of each variable is zero in either the $\ddot{\boldsymbol{\gamma}}$ vector or the $\boldsymbol{\alpha}$ vector or both vectors, (5) if an element of $\tilde{\boldsymbol{b}}_i$ corresponding to a specific variable in the vector $\boldsymbol{x}$ is non-zero, a sufficient condition for identification is that the utility of alternative i in the nominal variable model not depend on any latent variable that contains that specific variable as a covariate in the structural equation system, (6) endogenous variable effects can be specified only in a single direction and when a continuous observed endogenous variable appears as a right side variable in the regression for another continuous observed endogenous variable, or as a right side variable in the latent regression underlying another count or ordinal endogenous variable, each latent variable appearing in the regression/latent regression for the other endogenous continuous/count/ordinal variable (say variable A) should have two factor complexity one outcome variables after excluding the equation for variable A, and (7) If an element of $\boldsymbol{\delta}_k$ corresponding to a specific variable in the vector $\boldsymbol{x}$ is non-zero, a sufficient condition for identification is that the utility of alternative *k* in the MDC model not depend on any latent variable that contains that specific variable as a covariate in the structural equation system. Of course, there may be much less restrictive situations under which the parameters are all still identified, but the number of such specific situations is too numerous to list here.

To estimate the model, one can use a maximum simulated likelihood approach by writing the multivariate normal density function for the vector $\boldsymbol{yu}$ as the product of the marginal distribution of the continuous components in $\boldsymbol{yu}$ (corresponding to the *H* continuous outcomes and the consumed alternatives from among the *K*–1 MDC inside alternatives) and the conditional distribution of the remaining components in $\boldsymbol{yu}$ given the continuous components. Then, the conditional density function can be integrated appropriately. Specifically, define a $\tilde{E} \times \tilde{E}$ matrix

$\tilde{\mathbf{M}}$ $[\tilde{E} = (E + I + K - 2)]$, and fill it with all zeros. Then, place an identity matrix of size H in the first H rows and first H columns. Then, in the next $\tilde{F}_C$ rows, place an element of '1' in the $(H+1)^{th}$ row and the $(E + I - 1 + F_C[1])^{th}$ column, an element of '1' in the $(H+2)^{th}$ row and the $(E + I - 1 + F_C[2])^{th}$ column, and so on until an element of '1' in the $(H + \tilde{F}_C)^{th}$ row and the $(E + I - 1 + F_C[\tilde{F}_C])^{th}$ column. Also, in the $(H + \tilde{F}_C + 1)^{th}$ row through the $(H + \tilde{F}_C + E - 1)^{th}$ row, place an identity matrix of size $E-1$, starting from the $(H+1)^{th}$ column and ending at the $(E + I - 1)^{th}$ column. Finally, in the last $\tilde{F}_{NC}$ rows, place an element of '1' in the $(E + \tilde{F}_C + I)^{th}$ row and the $(E + I - 1 + F_{NC}[1])^{th}$ column, an element of '1' in the $(E + \tilde{F}_C + I + 1)^{th}$ row and the $(E + I - 1 + F_{NC}[2])^{th}$ column, and so on until an element of '1' in the $(E + I + K - 2)^{th}$ row and the $(E + I - 1 + F_C[\tilde{F}_C]^{th}$ column. Define $\boldsymbol{y\tilde{u}} = \tilde{\mathbf{M}}(\boldsymbol{yu})$, $\tilde{\boldsymbol{B}} = \tilde{\mathbf{M}}\boldsymbol{B}$, and $\tilde{\boldsymbol{\Theta}} = \tilde{\mathbf{M}}\boldsymbol{\Theta}\tilde{\mathbf{M}}'$. Next, partition the vector $\boldsymbol{y\tilde{u}}$ into two components: $\boldsymbol{y\tilde{u}}_1 = \boldsymbol{y\tilde{u}}[1 : H + \tilde{F}_C]$ and $\boldsymbol{y\tilde{u}}_2 = \boldsymbol{y\tilde{u}}[H + \tilde{F}_C + 1 : \tilde{E}]$, where $\boldsymbol{y\tilde{u}}[1 : H + \tilde{F}_C]$ is the sub-vector of $\boldsymbol{y\tilde{u}}$ corresponding to the first through the $(H + \tilde{F}_C)^{th}$ element, and $\boldsymbol{y\tilde{u}}[H + \tilde{F}_C + 1 : \tilde{E}]$ is the sub-vector of $\boldsymbol{y\tilde{u}}$ corresponding to the $(H + \tilde{F}_C + 1)^{th}$ element through the last element $\tilde{E}$. Next, partition the vector $\tilde{\boldsymbol{B}}$ into two components: $\tilde{\boldsymbol{B}}_1 = \tilde{\boldsymbol{B}}[1 : H + \tilde{F}_C]$ and $\tilde{\boldsymbol{B}}_2 = \tilde{\boldsymbol{B}}[H + \tilde{F}_C + 1 : \tilde{E}]$. Correspondingly partition $\tilde{\boldsymbol{\Theta}}$: $\tilde{\boldsymbol{\Theta}}_1 = \tilde{\boldsymbol{\Theta}}[1 : H + \tilde{F}_C, 1 : H + \tilde{F}_C]$, $\tilde{\boldsymbol{\Theta}}_2 = \tilde{\boldsymbol{\Theta}}[H + \tilde{F}_C + 1 : \tilde{E}, H + \tilde{F}_C + 1 : \tilde{E}]$, and $\tilde{\boldsymbol{\Theta}}_{12} = \tilde{\boldsymbol{\Theta}}[H + \tilde{F}_C + 1 : \tilde{E}, 1 : H + \tilde{F}_C]$.

Then, we may write:

$$\boldsymbol{y\tilde{u}} = \begin{bmatrix} \boldsymbol{y\tilde{u}}_1 \\ \boldsymbol{y\tilde{u}}_2 \end{bmatrix}, \quad \tilde{\boldsymbol{B}} = \begin{bmatrix} \tilde{\boldsymbol{B}}_1 \\ \tilde{\boldsymbol{B}}_2 \end{bmatrix}, \text{ and } \quad \tilde{\boldsymbol{\Theta}} = \begin{bmatrix} \tilde{\boldsymbol{\Theta}}_1 & \tilde{\boldsymbol{\Theta}}_{12} \\ \tilde{\boldsymbol{\Theta}}_{12} & \tilde{\boldsymbol{\Theta}}_2 \end{bmatrix} \text{ vector,} \tag{B.3}$$

Further, define $\vec{\boldsymbol{B}}_2 = \tilde{\boldsymbol{B}}_2 + \tilde{\boldsymbol{\Theta}}_{12}\tilde{\boldsymbol{\Theta}}_1\left(\boldsymbol{y\tilde{u}}_1 - \tilde{\boldsymbol{B}}_1\right)$, $\vec{\boldsymbol{\Theta}}_2 = \tilde{\boldsymbol{\Theta}}_2 - \tilde{\boldsymbol{\Theta}}_{12}\tilde{\boldsymbol{\Theta}}_1\tilde{\boldsymbol{\Theta}}'_{12}$.

$\vec{\boldsymbol{\psi}}_{low} = \left[\tilde{\boldsymbol{\psi}}'_{low}, \breve{\boldsymbol{\psi}}'_{low}, \left(-\infty_{I-1+\tilde{F}_{NC}}\right)'\right]'$ ($[(N + I + \tilde{F}_{NC}) \times 1]$ vector) and

$\vec{\boldsymbol{\psi}}_{up} = \left[\tilde{\boldsymbol{\psi}}'_{up}, \breve{\boldsymbol{\psi}}'_{up}, \left(\mathbf{0}_{I-1+\tilde{F}_{NC}}\right)'\right]'$ ($[(N + I + \tilde{F}_{NC}) \times 1]$ vector), where $-\infty_{I-1+\tilde{F}_{NC}}$ is a $(I - 1 + \tilde{F}_{NC}) \times 1$-

column vector of negative infinities, and $\mathbf{0}_{I-1+\tilde{F}_{NC}}$ is another $(I-1+\tilde{F}_{NC})\times 1$-column vector of zeros. Then, the likelihood function may be written as:

$$L(\vec{\boldsymbol{\theta}}) = \det(\mathbf{J}) \times f_{H+\tilde{F}_C}((\boldsymbol{y}, \mathbf{0}_{\tilde{F}_C}) \mid \tilde{\boldsymbol{B}}_1, \tilde{\boldsymbol{\Theta}}_1) \times \Pr\left[\vec{\boldsymbol{\psi}}_{low} \le \boldsymbol{y}\tilde{\boldsymbol{u}}_2 \le \vec{\boldsymbol{\psi}}_{up}\right], \quad \text{(B.4)}$$

$$= \det(\mathbf{J}) \times f_{H+\tilde{F}_C}((\boldsymbol{y}, \mathbf{0}_{\tilde{F}_C}) \mid \tilde{\boldsymbol{B}}_1, \tilde{\boldsymbol{\Theta}}_1) \times \int_{D_r} f_{N+I+\tilde{F}_{NC}}(\boldsymbol{r} \mid \vec{\boldsymbol{B}}_2, \vec{\boldsymbol{\Omega}}_2)\, dr$$

where $\det(\mathbf{J})$ is the determinant of the Jacobian given by $\det(\mathbf{J}) = \left\{\prod_{k\in\mathfrak{C}} \frac{1-\alpha_k}{t_k^* + \gamma_k}\right\}\left\{\sum_{k\in\mathfrak{C}}\left(\frac{t_k^* + \gamma_k}{1-\alpha_k}\right)\right\}$, $\mathfrak{C}$ is the set of activity purposes invested in by the individual (including activity purpose *K*), and the integration domain $D_r = \{\boldsymbol{r} : \vec{\boldsymbol{\psi}}_{low} \le \boldsymbol{r} \le \vec{\boldsymbol{\psi}}_{up}\}$ is simply the multivariate region of the elements of the $\boldsymbol{y}\tilde{\boldsymbol{u}}_2$ vector. $f_{H+\tilde{F}_C}(\boldsymbol{y}, \mathbf{0}_{\tilde{F}_C}) \mid \tilde{\boldsymbol{B}}_1, \tilde{\boldsymbol{\Theta}}_1)$ is the multivariate normal density function of dimension $H + \tilde{F}_C$ with a mean of $\tilde{\boldsymbol{B}}_1$ and a covariance of $\tilde{\boldsymbol{\Theta}}_1$, and evaluated at $(\boldsymbol{y}, \mathbf{0}_{\tilde{F}_C})$. The likelihood function for a sample of *Q* decision-makers is obtained as the product of the individual-level likelihood functions.

The likelihood function in Equation (B.4) involves the evaluation of an $(N + I + \tilde{F}_{NC})$-dimensional rectangular integral for each decision-maker, which can be computationally expensive. So, the Maximum Approximate Composite Marginal Likelihood (MACML) approach of Bhat (2011) is used.

The joint mixed model system and the MACML estimation approach

Consider the following (pairwise) composite marginal likelihood function formed by taking the products (across the *N* ordinal variables, the count variable, and the nominal variable) of the joint pairwise probability of the chosen alternatives for a decision-maker, and computed using the analytic approximation of the multivariate normal cumulative distribution (MVNCD) function.

$$
\begin{aligned}
L_{CML}(\vec{\theta}) = f_H(\boldsymbol{y} \mid \widetilde{\boldsymbol{B}}_y, \widetilde{\boldsymbol{\Omega}}_y) \times \left(\prod_{n=1}^{N-1} \prod_{n'=n+1}^{N} \Pr(j_n = a_n, j_{n'} = a'_n) \right) \times \\
\times \left(\prod_{n=1}^{N} \Pr(j_n = a_n, g = r) \right) \times \left(\prod_{n=1}^{N} \Pr(j_n = a_n, i = m) \right) \times \left(\Pr(g = r, i = m) \right. \\
\left. \times \left(\prod_{n=1}^{N} \Pr\left(\boldsymbol{t}^*; j_n = a_n\right) \right) \times \Pr\left(\boldsymbol{t}^*; g = r\right) \times \Pr\left(\boldsymbol{t}^*; i = m\right) \right.
\end{aligned}
\tag{B.5}
$$

In the above CML approach, the multivariate normal cumulative distribution (MVNCD) function appearing in the CML function is of dimension equal to (1) two for the second component (corresponding to the probability of each pair of observed ordinal outcomes), (2) two for the third component (corresponding to the probability of each pair of an observed ordinal outcome and the observed count outcome), (3) I for the fourth component (corresponding to the probability of each combination of the observed nominal outcome with an observed ordinal outcome), (5) I for the fifth component (corresponding to the probability of the observed nominal outcome and the observed count outcome), (6) $\widetilde{F}_{NC} + 1$ for the sixth component (corresponding to a the probability of each combination of the observed MDC outcome of the observed time investment vector $\boldsymbol{t}^*$ and an observed ordinal outcome), and (7) $\widetilde{F}_{NC} + 1$ for the seventh component (corresponding to the combination of the MDC outcome and the count outcome), and (8) $\widetilde{F}_{NC} + I - 1$ for the eighth component (corresponding to the probability of the observed MDC and observed nominal outcome).

To explicitly write out the CML function, define $\boldsymbol{\omega}_\Delta$ as the diagonal matrix of standard deviations of matrix $\boldsymbol{\Delta}$, $\omega_{\Delta h}$ as the h^{th} diagonal element of $\boldsymbol{\omega}_\Delta$, $\phi_R(.;\boldsymbol{\Delta}^{**})$ for the multivariate standard normal density function of dimension R and correlation matrix $\boldsymbol{\Delta}^*$ ($\boldsymbol{\Delta}^* = \boldsymbol{\omega}_\Delta^{-1}\boldsymbol{\Delta}\boldsymbol{\omega}_\Delta^{-1}$), and $\Phi_R(.;\boldsymbol{\Delta}^*)$ for the multivariate standard normal cumulative distribution function of dimension R and correlation matrix $\boldsymbol{\Delta}^*$. Define two selection matrices as follows: (1) $\boldsymbol{D}_v$ is an $I \times (\widetilde{E} - H - \widetilde{F}_C)$ selection matrix with an entry of '1' in the first row and the v^{th} column, and an identity matrix of size $I - 1$ occupying the last $I - 1$ rows and the $(N+2)^{th}$ through $[N+I]^{th}$ columns, and entries of '0' everywhere else, (3) $\boldsymbol{A}_v$ is a $(\widetilde{F}_{NC} + 1) \times (\widetilde{E} - H - F_C)$

selection matrix, with an entry of '1' in the first row and the v^{th} column; in the next $\tilde{\tilde{F}}_{NC}$ rows, place an identity matrix of size $\tilde{\tilde{F}}_{NC}$ occupying columns $(N+1+I)^{th}$ through $(N+I+\tilde{\tilde{F}}_{NC})^{th}$ column; all other elements of $\boldsymbol{A}_v$ take a value of zero, and (4) $\boldsymbol{C}$ is a $(\tilde{\tilde{F}}_{NC}+I-1)\times(\tilde{\tilde{E}}-H-F_C)$ selection matrix as follows: Position an identity matrix of size $(I-1)$ occupying the first $(I-1)$ rows and the $(N+2)^{th}$ through $(N+I)^{th}$ columns, and another identity matrix of size $\tilde{\tilde{F}}_{NC}$ occupying columns $(N+1+I)^{th}$ through $(N+I+\tilde{\tilde{F}}_{NC})^{th}$ column; all other elements of $\boldsymbol{C}$ take a value of zero.

Let $\hat{\boldsymbol{B}}_v = \boldsymbol{D}_v\ddot{\boldsymbol{B}}_2$, $\hat{\boldsymbol{\Theta}}_v = \boldsymbol{D}_v\ddot{\boldsymbol{\Theta}}_2\boldsymbol{D}_v'$, $\ddot{\boldsymbol{B}}_v = \boldsymbol{A}_v\ddot{\boldsymbol{B}}_2$, $\ddot{\boldsymbol{\Theta}}_v = \boldsymbol{A}_v\ddot{\boldsymbol{\Theta}}_2\boldsymbol{A}_v'$, $\breve{\boldsymbol{B}} = \boldsymbol{C}\ddot{\boldsymbol{B}}_2$, $\breve{\boldsymbol{\Theta}} = \boldsymbol{C}\ddot{\boldsymbol{\Theta}}_2\boldsymbol{C}'$,

$$\hat{\boldsymbol{\psi}}_{v,low} = \boldsymbol{D}_v\left(\left[\tilde{\tilde{\boldsymbol{\psi}}}_{low}', \breve{\boldsymbol{\psi}}_{low}', \left(\boldsymbol{0}_{I-1+\tilde{\tilde{F}}_{NC}}\right)'\right]'\right), \quad \hat{\boldsymbol{\psi}}_{v,up} = \boldsymbol{D}_v\ddot{\boldsymbol{\psi}}_{up}, \quad \ddot{\boldsymbol{\psi}}_{v,low} = \boldsymbol{A}_v\left(\left[\tilde{\tilde{\boldsymbol{\psi}}}_{low}', \breve{\boldsymbol{\psi}}_{low}', \left(\boldsymbol{0}_{I-1+\tilde{\tilde{F}}_{NC}}\right)'\right]'\right),$$

$$\ddot{\boldsymbol{\psi}}_{v,up} = \boldsymbol{A}_v\ddot{\boldsymbol{\psi}}_{up}, \quad \breve{\boldsymbol{\psi}}_{up} = \boldsymbol{C}\ddot{\boldsymbol{\psi}}_{up}, \quad \mu_{v,up} = \frac{\left[\ddot{\boldsymbol{\psi}}_{up}\right]_v - \left[\ddot{\boldsymbol{B}}_2\right]_v}{\sqrt{\left[\ddot{\boldsymbol{\Theta}}_2\right]_{vv}}}, \mu_{v,low} = \frac{\left[\ddot{\boldsymbol{\psi}}_{low}\right]_v - \left[\ddot{\boldsymbol{B}}_2\right]_v}{\sqrt{\left[\ddot{\boldsymbol{\Theta}}_2\right]_{vv}}}, \rho_{vv'} = \frac{\left[\ddot{\boldsymbol{\Theta}}_2\right]_{vv'}}{\sqrt{\left[\ddot{\boldsymbol{\Theta}}_2\right]_{vv}\left[\ddot{\boldsymbol{\Theta}}_2\right]_{v'v'}}},$$

where $\left[\ddot{\boldsymbol{\psi}}_{up}\right]_v$ represents the v^{th} element of $\ddot{\boldsymbol{\psi}}_{up}$ (and similarly for other vectors), and $\left[\ddot{\boldsymbol{\Theta}}_2\right]_{vv'}$ represents the vv'^{th} element of the matrix $\ddot{\boldsymbol{\Theta}}_2$.

$$\begin{aligned} L_{CML}(\ddot{\boldsymbol{\theta}}) = \det(\mathbf{J})\times & \left(\prod_{h=1}^{H+F_C}\omega_{\tilde{\tilde{\boldsymbol{\Theta}}}_1 h}\right)^{-1}\phi_H\left(\left[\boldsymbol{\omega}_{\tilde{\tilde{\boldsymbol{\Theta}}}_1}\right]^{-1}\left[(\boldsymbol{y},\boldsymbol{0}_{\tilde{\tilde{F}}_C}) - \tilde{\tilde{\boldsymbol{B}}}_1\right]; \tilde{\tilde{\boldsymbol{\Theta}}}_1^*\right)\times \\ & \left(\prod_{v=1}^{N}\prod_{v'=v+1}^{N+1}\begin{bmatrix}\Phi_2(\mu_{v,up},\mu_{v',up},\rho_{vv'}) & -\Phi_2(\mu_{v,up},\mu_{v',low},\rho_{vv'}) \\ -\Phi_2(\mu_{v,low},\mu_{v',up},\rho_{vv'}) & +\Phi_2(\mu_{v,low},\mu_{v',low},\rho_{vv'})\end{bmatrix}\right)\times \\ & \left(\prod_{v=1}^{N+1}\left[\Phi_I\left[\boldsymbol{\omega}_{\hat{\boldsymbol{\Theta}}_v}^{-1}\left\{\hat{\boldsymbol{\psi}}_{v,up} - \hat{\boldsymbol{B}}_v\right\}; \hat{\boldsymbol{\Theta}}_v^*\right] - \Phi_I\left[\boldsymbol{\omega}_{\hat{\boldsymbol{\Theta}}_v}^{-1}\left\{\hat{\boldsymbol{\psi}}_{v,low} - \hat{\boldsymbol{B}}_v\right\}; \hat{\boldsymbol{\Theta}}_v^*\right]\right]\right)\times \\ & \left(\prod_{v=1}^{N+1}\left[\Phi_{\tilde{\tilde{F}}_{NC}+1}\left[\boldsymbol{\omega}_{\ddot{\boldsymbol{\Theta}}_v}^{-1}\left\{\ddot{\boldsymbol{\psi}}_{v,up} - \ddot{\boldsymbol{B}}_v\right\}; \ddot{\boldsymbol{\Theta}}_v^*\right] - \Phi_{\tilde{\tilde{F}}_{NC}+1}\left[\boldsymbol{\omega}_{\ddot{\boldsymbol{\Theta}}_v}^{-1}\left\{\ddot{\boldsymbol{\psi}}_{v,low} - \ddot{\boldsymbol{B}}_v\right\}; \ddot{\boldsymbol{\Theta}}_v^*\right]\right]\right)\times \\ & \left(\left[\Phi_{\tilde{\tilde{F}}_{NC}+I-1}\left[\boldsymbol{\omega}_{\ddot{\boldsymbol{\Theta}}_v}^{-1}\left\{\breve{\boldsymbol{\psi}}_{up} - \breve{\boldsymbol{B}}\right\}; \breve{\boldsymbol{\Theta}}^*\right]\right]\right) \end{aligned} \tag{B.6}$$

In the MACML approach, all MVNVD function evaluation greater than two dimensions are evaluated using an *analytic approximation* method rather than a simulation method. This combination of the CML with an analytic approximation for the MVNCD function is effective because the analytic approximation involves only univariate and bivariate cumulative normal

distribution function evaluations. The MVNCD analytic approximation method used here is based on linearization with binary variables (see Bhat, 2011). Write the resulting equivalent of Equation (B.6) computed using the analytic approximation for the MVNCD function as $L_{MACML,q}(\vec{\boldsymbol{\theta}})$, after introducing the index q for individuals. The MACML estimator is then obtained by maximizing the following function:

$$\log L_{MACML}(\vec{\boldsymbol{\theta}}) = \sum_{q=1}^{Q} \log L_{MACML,q}(\vec{\boldsymbol{\theta}}). \tag{B.7}$$

The covariance matrix of the parameters $\vec{\boldsymbol{\theta}}$ may be estimated by the inverse of Godambe's (1960) sandwich information matrix (see Zhao and Joe, 2005, and Bhat, 2015).

$$V_{MACML}(\vec{\boldsymbol{\theta}}) = \frac{\left[\hat{\boldsymbol{G}}(\vec{\boldsymbol{\theta}})\right]^{-1}}{Q} = \frac{\left[\hat{\boldsymbol{H}}^{-1}\right]\left[\hat{\boldsymbol{J}}\right]\left[\hat{\boldsymbol{H}}^{-1}\right]'}{Q}, \tag{B.8}$$

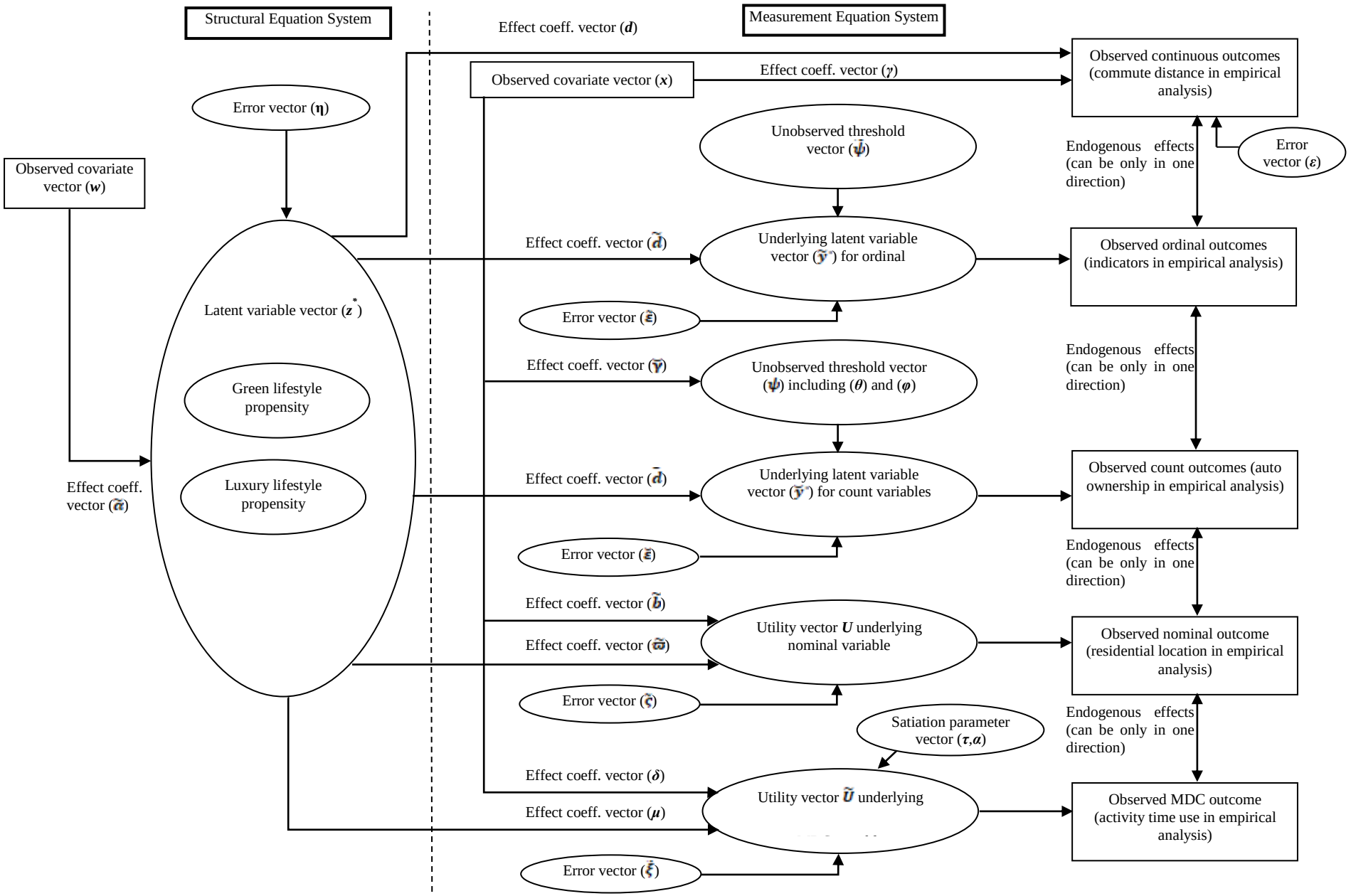
$$\text{with } \hat{\boldsymbol{H}} = -\frac{1}{Q}\left[\sum_{q=1}^{Q} \frac{\partial^2 \log L_{MACML,q}(\vec{\boldsymbol{\theta}})}{\partial\vec{\boldsymbol{\theta}}\partial\vec{\boldsymbol{\theta}}'}\right]_{\hat{\vec{\boldsymbol{\theta}}}_{MACML}}$$

$$\hat{\boldsymbol{J}} = \frac{1}{Q}\sum_{q=1}^{Q}\left[\left(\frac{\partial \log L_{MACML,q}(\vec{\boldsymbol{\theta}})}{\partial\vec{\boldsymbol{\theta}}}\right)\left(\frac{\partial \log L_{MACML,q}(\vec{\boldsymbol{\theta}})}{\partial\vec{\boldsymbol{\theta}}'}\right)\right]_{\hat{\vec{\boldsymbol{\theta}}}_{MACML}} \tag{B.9}$$

Positive definiteness

The $(L\times L)$ correlation matrix $\boldsymbol{\Gamma}$, the $\left[(I-1)\times(I-1)\right]$ covariance matrix, and the $(K\times K)$ covariance matrix have to be all positive definite. An easy way to ensure the positive-definiteness of these matrices is to use a Cholesky-decomposition and parameterize the CML function in terms of the Cholesky parameters. Further, because the matrix $\boldsymbol{\Gamma}$ is a correlation matrix, we write each diagonal element (say the aa^{th} element) of the lower triangular Cholesky matrix of $\boldsymbol{\Gamma}$ as $\sqrt{1-\sum_{j=1}^{a-1} p_{aj}^2}$, where the p_{aj} elements are the Cholesky factors that are to be estimated.

APPENDIX C: DIAGRAMMATIC REPRESENTATION OF THE MDCP GHDM SYSTEM



APPENDIX D: DESCRIPTIVE CHARACTERISTICS OF THE PSRC SAMPLE

Table D.1 provides descriptive statistics of the socioeconomic characteristics of the sample and that of the PSRC four-county region population as a whole from the 2010 Census. Of course, the comparison is not really appropriate (because we are unable to obtain, from the Census data, statistics solely on 1+-worker households with at least one person employed outside the home). But we provide the population statistics just for informational purposes.

According to the statistics provided in Table D.1, a majority of the households are couple families (34.2%) or single person (27.8%), though there are also a sizeable number of nuclear families (20.2) and multi-adult households (the term "multi-adult is used here to represent more than two adults in the household; this category includes extended families and room-mates). The percentage of single parent families in our sample of 1+worker households is very low relative to the general population. The fractions of male and female adults within the household, when averaged across all sample households, are close to the 50% split observed in the population. Not surprisingly, the sample households in general are much more educated than the households in the population. This is also reflected in the high percentage of households with an annual income of over 75,000, though we do not have the income information for the PSRC region from the 2010 Census data. The distribution of the number of children shows a high percentage of childless households, consistent with the high percentage of single person, couple, and multi-adult households (though multi-adult households contribute to 3.3% of the 74.9% of childless households). The percentage of childless households is of the same order in the sample and the Census data, though the Census does not provide the breakdown by number of children for households with children. The fraction of adults by age in the sample, when averaged across all sample household, is highly loaded on the 35-54 year category relative to the entire population, with much smaller representation of individuals in the 65 years and beyond category. This is again not surprising given the focus on 1+-worker households in our sample. The work status distribution is not available from the Census, but the sample statistics on the fraction of adults in the household in each of four categories; full-time workers, part-time workers, self-employed workers, and non-workers; clearly indicates a high fraction of full-time workers, with the fraction of adults in the other three categories being of the same order and range from 0.066 to 0.128.

Table D.1. Sample characteristics of independent variables

Socio demographic	Categories	Sample Distribution	Census Distribution
Family structure	Single person household	27.8%	31.0%
	Single parent family	1.6%	9.0%
	Couple family	34.2%	29.7%
	Nuclear family	20.2%	23.0%
	Multi-person household	16.2%	7.3%
Gender	Fraction of male adults in household (mean)	0.468	0.50
	Fraction of female adults in household (mean)	0.532	0.50
Educational attainment	Fraction of adults with High school or less in household (mean)	0.097	0.303
	Fraction of adults with Some college in household (mean)	0.233	0.327
	Fraction of adults with a Bachelor's degree in household (mean)	0.382	0.239
	Fraction of adults with Graduate degree in household (mean)	0.288	0.131
Household income	$0 to below $25,000	5.9%	NA
	$25,000 to below $35,000	7.0%	NA
	$35,000 to below $50,000	10.7%	NA
	$50,000 to below $75,000	18.3%	NA
	$75,000 and above	558.1%	NA
Number of children	No kids	74.9%	70.1%
	One kid	12.6%	NA
	Two kids	10.0%	NA
	Three or more kids	2.5%	NA
Age	Fraction of adults aged 18 to 34 in household (mean)	0.341	0.349
	Fraction of adults aged 35 to 54 in household (mean)	0.421	0.369
	Fraction of adults aged 55 to 64 in household (mean)	0.185	0.147
	Fraction of adults 65 years old or older in household (mean)	0.053	0.135
Adult work status	Fraction of full-time working adults in the household (mean)	0.700	NA
	Fraction of part-time working adults in the household (mean)	0.106	NA
	Fraction of self-employed working adults in the household (mean)	0.066	NA
	Fraction of non-working adults in the household (mean)	0.128	NA

*NA: Not available

APPENDIX E: COMPUTATION OF THE AVERAGE TREATMENT EFFECTS

For auto ownership, the measure is estimated as follows for each model:

$$\hat{\text{ATE}}_{ii'}^{auto} = \frac{1}{Q}\sum_{q=1}^{Q}\left(\sum_{g=0}^{\infty} g \cdot \left[P(\text{auto ownership} = g \mid a_{qi} = 1) - P(\text{auto ownership} = g \mid a_{qi'} = 1)\right]\right) \quad \text{(E.1)}$$

where a_{qi} is the dummy variable for the density category i for the household q. Although the summation in the equation above extends until infinity, we consider counts only up to g=10, which is the maximum vehicle ownership level observed in the data set. The standard error of the measure is computed using bootstraps from the sampling distributions of the estimated parameters.

For the activity time use variables (MDC variable), we focus only on the participation dimension here and compute the ATE measure for the out-of-home activity k (k=1,2,…,K–1) as follows:

$$\hat{\text{ATE}}_{ii'}^{OH\ act.k} = \frac{1}{Q}\sum_{q=1}^{Q}\left(\left[P(t_{qk} > 0 \mid a_{qi} = 1) - P(t_{qk} > 0 \mid a_{qi'} = 1)\right]\right) \quad \text{(E.2)}$$

where t_{qk} is the time spent by individual q on the out-of-home activity k. To compute the probability that $t_{qk} > 0$, we drew, for each individual, 100 sets of 1000 realizations from a multivariate normal sampling distribution of estimated parameters and the distribution of the error terms involved. For each individual, each set, and each realization, we used the forecasting algorithm of Pinjari and Bhat (2014) to predict time allocations and, then, for each individual and each set, evaluated the share of the 1000 realizations that predicted $t_{qk} > 0$ for each of the two density categories involved (that is, i and i'). The treatment effect is then computed as in Equation (E.1) for each set, and the mean across all the 100 sets was computed as the final ATE effect and the standard deviation across the 100 sets was computed as the standard error estimate.