A Multivariate Hurdle Count Data Model with an Endogenous Multiple Discrete-Continuous Selection System

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

Subodh K. Dubey
The University of Texas at Austin
Department of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St. Stop C1761, Austin TX 78712
Phone: 512-471-4535, Fax: 512-475-8744
E-mail: subbits@gmail.com

Raghuprasad Sidharthan
The University of Texas at Austin
Department of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St. Stop C1761, Austin TX 78712
Phone: 512-471-4535, Fax: 512-475-8744
E-mail: raghu@mail.utexas.edu

Prerna C. Bhat
Harvard University
1350 Massachusetts Avenue, Cambridge, MA 02138
Phone: 512-289-0221
E-mail: prernabhat@college.harvard.edu

*corresponding author

July 20, 2013

ABSTRACT

This paper proposes a new econometric formulation and an associated estimation method for multivariate count data that are themselves observed conditional on a participation selection system that takes a multiple discrete-continuous model structure. This leads to a joint model system of a multivariate count and a multiple discrete-continuous selection system in a hurdle-type model. The model is applied to analyze the participation and time investment of households in out-of-home activities by activity purpose, along with the frequency of participation in each selected activity. The results suggests that the number of episodes of activities as well as the time investment in those activities may be more of a lifestyle- and lifecycle-driven choice than one related to the availability of opportunities for activity participation

Keywords: multivariate count data, generalized ordered-response, multiple discrete-continuous models, hurdle model system, endogeneity.

1. INTRODUCTION

In this paper, we develop a new econometric formulation and an associated estimation method for multivariate count data that are themselves observed based on a participation selection system. The participation selection system may be potentially endogenous to the multivariate count data in a hurdle-type model, which then leads to a joint count model system and participation selection system. The important feature of our proposed model is that the participation selection system itself takes a multiple discrete-continuous formulation in which multiple discrete states (with associated continuous intensities) may be simultaneously chosen for participation. A defining feature of our model is, therefore, that decision agents jointly choose one or more discrete alternatives *and* determine a continuous outcome as well as a count outcome for each discrete alternative. Further, if the decision agent does not choose a discrete alternative, there is no continuous or count outcome observed for this discrete alternative. Many empirical contexts in different fields conform to such a decision framework and can benefit from our proposed model. For instance, consider an individual's daily engagement in non-work activities, an issue of substantial interest in the time-use and transportation fields. The individual chooses to participate in different activity types (such as shopping, visiting, and recreation), and jointly determines the amount of time to invest in each activity type and the number of episodes of each activity type to participate in. Of course, should an individual choose not to participate in a specific activity type, there is no issue of time investment and number of episodes associated with that activity type. Another example from the transportation and energy fields would be the case of a household's choice and use of motorized vehicles. Here, a household may choose to own different numbers of various body types of vehicles (such as a compact sedan and/or a pick-up truck), and put different mileages on the different vehicles. Again, the count and mileage are not relevant for body types not chosen by the household. Econometrically speaking, the potentially inter-related nature of the choices in these situations originates from common unobserved factors. For instance, underlying household factors such as environmental consciousness may make a household more likely to own multiple compact sedans and use compact sedans for much of the household's travel needs. These same unobserved factors can potentially also reduce the likelihood of the household owning one or more pick-ups and putting mileage on the pick-up(s).

Our formulation for the joint model combines a multiple discrete-continuous (MDC) model system with a multivariate count (MC) model system. The MDC system takes a MDC probit (MDCP) form in our formulation, while the MC system is quite general and takes the form of a multivariate generalized ordered-response probit (MGORP) model. In particular, we use Castro, Paleti, and Bhat's (CPB's) (2011) recasting of a univariate count model as a restricted version of a univariate GORP model. This GORP system provides flexibility to accommodate high or low probability masses for specific count outcomes without the need for cumbersome treatment (especially in multivariate settings) using zero-inflated mechanisms. The error terms in the underlying latent continuous variables of the univariate GORP-based count models for each discrete alternative also provide a convenient mechanism to tie the counts of different alternatives together in a multivariate framework. Further, these error terms form the basis for tying the MC model system with the MDCP model system using a comprehensive correlated latent variable structure. Overall, the model system extends extant models for count data with endogenous participation (for example, see Greene, 2009) that have focused on the simpler situation of a binary choice selection model and a corresponding univariate count outcome model.

The frequentist inference approach we use in the paper to estimate the joint MDCP-MC system is based on an analytic (as opposed to a simulation) approximation of the multivariate normal cumulative distribution (MVNCD) function. Bhat (2011) discusses this analytic approach, which is based on earlier works by Solow (1990) and Joe (1995). The approach involves only univariate and bivariate cumulative normal distribution function evaluations in the likelihood function (in addition to the evaluation of the closed-form multivariate normal density function).

The paper is structured as follows. The next section presents the modeling frameworks for the two individual components of the overall model system—the MDCP model and the MC model. This sets the stage for the joint model system formulated in this paper and presented in Section 3. Section 4 develops a simulation experiment design and evaluates the ability of the proposed estimation approach to recover the model parameters. Section 5 focuses on an illustrative application of the proposed model to the analysis of households' daily activity participation. Finally, Section 6 concludes the paper by summarizing the important findings and contributions of the study.

2. THE INDIVIDUAL MODEL COMPONENTS

The use of the MDCP model in the current paper, rather than the MDC extreme value (MDCEV) model (Bhat, 2005, 2008) is motivated by the need to tie the MDC model with the MC model. For the MC model, as discussed in the previous section, we use a latent variable representation with normal error terms that also facilitates the tie with the MDCP model.

2.1 The MDCP model

Without loss of generality, we assume that the number of consumer goods in the choice set is the same across all consumers. Following Bhat (2008), consider a choice scenario where a consumer maximizes his/her utility subject to a binding budget constraint (for ease of exposition, we suppress the index for consumers):

$$\max U(\mathbf{x}) = \sum_{k=1}^{K} \frac{\gamma_k}{\alpha_k} \psi_k \left(\left(\frac{x_k}{\gamma_k} + 1 \right)^{\alpha_k} - 1 \right)$$
$$s.t. \quad \sum_{k=1}^{K} p_k x_k = E \ , \tag{1}$$

where the utility function $U(\mathbf{x})$ is quasi-concave, increasing and continuously differentiable, $\mathbf{x} \geq 0$ is the consumption quantity (vector of dimension $K \times 1$ with elements x_k), and γ_k, α_k, and ψ_k are parameters associated with good k. In the linear budget constraint, E is the total expenditure (or income) of the consumer $(E > 0)$, and p_k is the unit price of good k as experienced by the consumer. The utility function form in Equation (1) assumes that there is no essential outside good, so that corner solutions (*i.e.*, zero consumptions) are allowed for all the goods k (though at least one of the goods has to be consumed, given a positive E). The assumption of the absence of an essential outside good is being made only to streamline the presentation; relaxing this assumption is straightforward and, in fact, simplifies the analysis.[1]

[1] The issue of an essential outside good is related to a complete versus incomplete demand system. In a complete demand system, the demands of all goods (that exhaust the consumption space of consumers) are modeled. However, the consideration of complete demand systems can be impractical when studying consumptions in finely defined commodity/service categories. In such situations, it is common to use an incomplete demand system in the form of a Hicksian composite commodity approach. In this approach, one replaces all the elementary alternatives within each broad group that is not of primary interest to the analyst by a single composite alternative representing the broad group (one needs to assume in this approach that the prices of elementary goods within each broad group of consumption items vary proportionally). The analysis proceeds then by considering the composite goods as

The parameter γ_k ($\gamma_k > 0$) in Equation (1) allows corner solutions for good *k*, but also serves the role of a satiation parameter. The role of α_k ($\alpha_k \leq 1$) is to capture satiation effects, with a smaller value of α_k implying higher satiation for good *k*. ψ_k ($\psi_k > 0$) represents the stochastic baseline marginal utility; that is, it is the marginal utility at the point of zero consumption (see Bhat, 2008 for a detailed discussion).

Empirically speaking, it is difficult to disentangle the effects of γ_k and α_k separately, which leads to serious empirical identification problems and estimation breakdowns when one attempts to estimate both parameters for each good. Thus, Bhat (2008) suggests estimating both a γ-profile (in which $\alpha_k \to 0$ for all goods and all consumers, and the γ_k terms are estimated) and an α-profile (in which the γ_k terms are normalized to the value of one for all goods and consumers, and the α_k terms are estimated), and choose the profile that provides a better statistical fit. However, in this section, we will retain the general utility form of Equation (1) to keep the presentation general.

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

$$\psi_k = \exp(\boldsymbol{\beta}' \mathbf{z}_k + \xi_k), \tag{2}$$

where $\mathbf{z}_k$ is a *D*-dimensional column vector of attributes that characterize good *k* (including a dummy variable for each good except one, to capture intrinsic preferences for each good except one good that forms the base), $\boldsymbol{\beta}$ is a corresponding vector of coefficients (of dimension $D \times 1$), and ξ_k captures the idiosyncratic (unobserved) characteristics that impact the baseline utility of good *k*. We assume that the error terms ξ_k are multivariate normally distributed across goods *k*: $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots \xi_K)' \sim MVN_K(\mathbf{0}_K, \boldsymbol{\Lambda})$, where $MVN_K(\mathbf{0}_K, \boldsymbol{\Lambda})$ indicates a *K*-variate normal distribution with a mean vector of zeros denoted by $\mathbf{0}_K$ and a covariance matrix $\boldsymbol{\Lambda}$.

"outside" goods and modeling consumption in these outside goods as well as in the "inside" goods representing the consumption group of main interest to the analyst. It is common in practice in this Hicksian approach to include a single outside good with the inside goods. If this composite outside good is not essential, then the utility formulation in Equation (1) applies. If this composite outside good is essential, then the formulation needs minor revision to accommodate the essential nature of the outside good, as we will discuss later (see also Bhat, 2008).

The analyst can solve for the optimal consumption allocation vector $\mathbf{x}^* = (x_1^*, x_2^*, ..., x_K^*)$ corresponding to Equation (1) by forming the Lagrangian and applying the Karush-Kuhn-Tucker (KKT) conditions. To do so, let's say that m is the consumed good with the lowest value of k for the consumer.[2] The order in which the goods are organized does not affect the model formulation or estimation, though the labeling of the goods must remain the same across consumers. Also, define $V_k = \boldsymbol{\beta}'\mathbf{z}_k + (\alpha_k - 1)\ln\left(\frac{x_k^*}{\gamma_k} + 1\right) - \ln p_k$, $U_k = V_k + \xi_k$, and $u_{km}^* = U_k - U_m$. Then, following Bhat (2008), the KKT conditions may be written as:

$$\begin{aligned} &u_{km}^* = 0, \text{ if } x_k^* > 0,\ k = 1,2,...,K,\ k \neq m \\ &u_{km}^* < 0, \text{ if } x_k^* = 0,\ k = 1,2,...,K,\ k \neq m\ . \end{aligned} \tag{3}$$

For later use, stack U_k, V_k, and ξ_k into $K\times 1$ vectors: $\mathbf{U} = (U_1, U_2, ..., U_K)'$, $\mathbf{V} = (V_1, V_2, ..., V_K)'$, and $\boldsymbol{\xi} = (\xi_1, \xi_2, ..., \xi_K)'$, respectively, and let $\mathbf{z} = (\mathbf{z_1}, \mathbf{z_2}, ..., \mathbf{z_K})'$ be a $K\times D$ matrix of variable attributes. Then, we may write, in matrix notation, $\mathbf{U} = \mathbf{V} + \boldsymbol{\xi} = \mathbf{z}\boldsymbol{\beta} + \boldsymbol{\xi}$ and $\mathbf{U} \sim MVN_{K-1}(\mathbf{V}, \boldsymbol{\Lambda})$. Also, for later use, define $\mathbf{u}_m^* = (u_{1m}, u_{2m}, \ldots u_{Km})'$ as a $(K\text{-}1)\times 1$ vector, and $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, ..., \gamma_K)'$ and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_K)'$. As already indicated, only one of the vectors $\boldsymbol{\gamma}$ or $\boldsymbol{\alpha}$ will be estimated.

Three important identification issues need to be noted here because the KKT conditions above are based on differences, as reflected in the u_{km}^* terms. First, a constant coefficient cannot be identified in the $\boldsymbol{\beta}$ 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 $\varepsilon_{k1} = \xi_k - \xi_1$, $k \neq 1$ are estimable. However, the KKT conditions take the difference against the first consumed good m for the consumer. Thus, in translating the KKT conditions in Equation (3) to the consumption probability, the covariance matrix $\boldsymbol{\Lambda}_m$ is desired. Since m will vary across

[2] The consumer has to consume at least one of the alternatives, because the alternatives are goods and $E > 0$ in Equation (1).

consumers, $\mathbf{\Lambda}_m$ will also vary across consumers. But all the $\mathbf{\Lambda}_m$ matrices must originate in the same covariance matrix $\mathbf{\Lambda}$ for the original error term vector ξ. To achieve this consistency, $\mathbf{\Lambda}$ is constructed from $\mathbf{\Lambda}_1$ 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. $\mathbf{\Lambda}_m$ may then be obtained appropriately for each consumer based on the same $\mathbf{\Lambda}$ matrix. Third, an additional scale normalization needs to be imposed on $\mathbf{\Lambda}$ if there is no price variation across goods for each consumer (*i.e.*, if $p_k = \tilde{p} \ \ \forall k$ for all consumers). For instance, one can normalize the element of $\mathbf{\Lambda}$ 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 $\mathbf{\Lambda}_1$ are estimable (see Bhat, 2008).

2.2 The MC model

Let y_k be the index for the count for discrete alternative k, and let l_k be the actual count value observed for the alternative. In this section, we develop the basics of the multivariate count model, without any hurdle based on the MDC model.

Consider the recasting of the count model for each discrete alternative using a generalized ordered-response probit (GORP) structure as follows:

$$y_k^* = \eta_k, \ y_k = l_k \ \ if \ \ \psi_{k,l_k-1} < y_k^* < \psi_{k,l_k}, \ \ l_k \in \{0,1,2,.....\}, \tag{4}$$

$$\psi_{k,l_k} = f_{k,l_k}(s) = \Phi^{-1}\left[e^{-\lambda_k} \sum_{r=0}^{l_k} \left(\frac{\lambda_k^r}{r!} \right) \right] + \varphi_{k,l_k}, \text{ where } \lambda_k = e^{\boldsymbol{\varsigma}_k' \mathbf{s}_k}.$$

In the above equation, y_k^* is a latent continuous stochastic propensity variable associated with alternative k that maps into the observed count l_k through the $\boldsymbol{\psi}_k$ vector (which is itself a vertically stacked column vector of thresholds $(\psi_{k,0}, \psi_{k,1}, \psi_{k,2}, ...,)'$). This variable, which is equated to η_k in the GORP formulation above, is a standard normal random error term.[3] $\boldsymbol{\varsigma}_k$ is a

[3] The use of the standard normal distribution rather than a non-standard normal distribution for the error term η_k is an innocuous normalization (see McKelvey and Zavoina, 1975; Greene and Hensher, 2010). Note also that any other proper continuous error distribution may be assumed for the η_k error terms, such as the logistic distribution or the extreme value distribution. However, for our purpose of tying the counts across the discrete alternatives as well as accommodating the endogeneity of the MDC model, the normal distribution is convenient.

vector of parameters (of dimension $\widetilde{C} \times 1$) corresponding to the conformable vector of observables $\mathbf{s}_k$ (including a constant).

The threshold terms satisfy the ordering condition (*i.e.*, $-\infty < \psi_{k,0} < \psi_{k,1} < \psi_{k2} .... < \infty$), as long as $-\infty < \varphi_{k,0} < \varphi_{k,1} < \varphi_{k,2} .... < \infty$.[4] The presence of these $\boldsymbol{\varphi}$ terms provides flexibility to accommodate high or low probability masses for specific count outcomes without the need for cumbersome treatment using zero-inflated or related mechanisms. For identification, we set $\psi_{k,-1} = -\infty \; \forall k$, and $\varphi_{k,0} = 0 \forall k$. In addition, we identify a count value e_k^* ($e_k^* \in \{0,1,2,.....\}$) above which $\varphi_{k,e}$ ($e \in \{0,1,2,.....\}$) is held fixed at φ_{k,e_k^*}; that is, $\varphi_{k,e} = \varphi_{k,e_k^*}$ if $e_k > e_k^*$, where the value of e_k^* can be based on empirical testing. With such a specification of the threshold values, the GORP model in Equation (4) is a flexible count model that can predict the probability of an arbitrary count. $\Phi^{-1}(\cdot)$ in the threshold function of Equation (4) is the inverse function of the univariate cumulative standard normal. For later use, let $\boldsymbol{\varphi}_k = (\varphi_{k,1}, \varphi_{k,2}, \ldots \varphi_{k,l_k^*})'$ ($l_k^* \times 1$ matrix), and $\boldsymbol{\varphi} = (\boldsymbol{\varphi}_1', \boldsymbol{\varphi}_2', \ldots \boldsymbol{\varphi}_K')'$ ($\left(\sum_k l_k^* \right) \times 1$ vector). [5]

The η_k terms may be correlated across different alternatives because of unobserved factors.. Formally, define $\boldsymbol{\eta} = (\eta_1, \eta_2, \eta_3, \ldots, \eta_K)'$. Then $\boldsymbol{\eta}$ is assumed to be multivariate standard normally distributed: $\boldsymbol{\eta} \sim MVN_K(\mathbf{0}_K, \boldsymbol{\Gamma})$, where $\boldsymbol{\Gamma}$ is a correlation matrix. For later use, define the following vectors and matrices. Let $\mathbf{y}^* = (y_1^*, y_2^*, ..., y_K^*)'$ ($K \times 1$ vector), $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_K)'$ ($K \times 1$ vector), and $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_K)'$ ($K \times 1$ vector). Define $\mathbf{s}$ as a $(K \times K\widetilde{C})$ block diagonal matrix, with each block-diagonal occupied by a $(1 \times \widetilde{C})$ vector $\mathbf{s}_k'$ (organized so that $\mathbf{s}_1'$ appears in the first row, $\mathbf{s}_2'$ appears in the second row, and so on). Let $\boldsymbol{\varsigma} = (\boldsymbol{\varsigma}_1', \boldsymbol{\varsigma}_2', ..., \boldsymbol{\varsigma}_K')'$ ($K\widetilde{C} \times 1$ vector). Then, $\mathbf{y}^* = \boldsymbol{\eta}$, and $\mathbf{y}^* \sim MVN_K(\mathbf{0}_K, \boldsymbol{\Gamma})$. Also, using an extension of conventional matrix

[4] The non-linear nature of the functional form for the non-φ component of the thresholds satisfies the ordering conditions by construction.

[5] The specification of the GORP-based count model in Equation (4) provides a flexible mechanism to model count data. It subsumes the traditional count models as specific and restrictive cases. In particular, if all elements of the $\boldsymbol{\varphi}_k$ vector are zero, the result is the Poisson count model (see CPB).

notation so that the exponent of a matrix returns a matrix of the same size with the exponent of each element of the original matrix, we write $\boldsymbol{\lambda} = \exp(\mathbf{s}\boldsymbol{\varsigma})$.

3. THE JOINT MODEL SYSTEM AND ESTIMATION APPROACH

An important feature of the proposed joint model system is that y_k (the count corresponding to discrete *k*) is observed only if there is some positive consumption of the alternative *k* as determined in the MDC model. That is, y_k is observed only if $x_k^* > 0$, and $y_k > 0$ in this case (y_k is not observed if $x_k^* = 0$). Thus, the proposed model resembles the typical hurdle model used in the count literature, but with three very important differences that make the proposed model much more general. *First*, the hurdle is set by an MDC model, as opposed to a simple binary model of participation (if the MDC model has only two alternatives, and individuals choose only one of the two alternatives, the satiation parameter α_k=1 for all *k* and the MDC model can be shown to collapse to a simple binary probit model). *Second*, there are multiple hurdles, each hurdle corresponding to a discrete alternative *k*. To the extent that the stochastic elements in U_k are allowed to be correlated, the hurdle conditions also get correlated. This leads to a multivariate truncation system. *Third*, we allow correlation in the counts across discrete alternatives, and also allow a fully general covariance structure between the MDC and MC models in a joint framework. As a result, the estimation approach involves the joint estimation of the MDC and MC model components.

Our joint model is based on the KKT conditions of the MDC model from Equation (3), supplemented by the following revised mechanism (from that discussed in the previous section) for observing counts for each alternative *k*:

$$y_k^* = \eta_k,\ y_k = l_k \text{ if } \psi_{k,l_k-1} < y_k^* < \psi_{k,l_k},\ y_k \text{ observed only if } x_k^* > 0 \tag{5}$$

$$\text{with } \psi_{k,l_k} = f_{k,l_k}(s) = \Phi^{-1}\left[e^{-\lambda_k}\sum_{r=0}^{l_k}\left(\frac{\lambda_k^r}{r!}\right)\right] + \varphi_{k,l_k},\ \lambda_k = e^{\boldsymbol{\varsigma}_k'\mathbf{s}_k},\ l_k \in \{1, 2, ...\}$$

Note that there is truncation present in the system above, since we are confining attention to positive values of the counts. Thus, there needs to be a scaling undertaken so that the probabilities of the positive count outcomes sum to one; this is achieved by restricting the region of y_k^* to not include the range from –inf to $\Phi^{-1}\left[e^{-\lambda_k}\right]$; that is, to not include

$y_k^* < \psi_{k,0} = \Phi^{-1}\left[e^{-\lambda_k}\right]$. Of course, to the extent that there is correlation in the y_k^* values across the discrete alternatives, this truncation itself takes a multivariate form, as considered later in the estimation section.

To proceed, define a $(2K \times 1)$-dimensional vector $\mathbf{G} = \left[\mathbf{U}', \left(\mathbf{y}^*\right)'\right]'$. Let $\mathbf{H} = (\mathbf{V}', \mathbf{0}_K')'$ and let $\boldsymbol{\Xi}$ be the covariance between the vectors $\mathbf{U}$ and $\mathbf{y}^*$. Then, $\mathbf{G} \sim MVN_{2K}(\mathbf{H}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\Sigma} = \begin{bmatrix} 0 & \mathbf{0}' & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\Lambda}_1 & \boldsymbol{\Xi}' \\ \mathbf{0} & \boldsymbol{\Xi} & \boldsymbol{\Gamma} \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\Sigma}_1 \end{bmatrix}, \text{ where } \boldsymbol{\Sigma}_1 = \begin{bmatrix} \boldsymbol{\Lambda}_1 & \boldsymbol{\Xi}' \\ \boldsymbol{\Xi} & \boldsymbol{\Gamma} \end{bmatrix}, \tag{6}$$

and $\boldsymbol{\Lambda}_1$ is as defined in Section 2.1. Next, define $\mathbf{M}$ as an identity matrix of size $2K$–1 with an extra column added at the m^{th} column of the consumer (thus, $\mathbf{M}$ is a matrix of dimension $(2K-1) \times (2K)$. This m^{th} column of $\mathbf{M}$ has the value of '-1' in the first $(K-1)$ rows and the value of zero in the remaining K rows. Then, $\widetilde{\mathbf{G}} = \left[\mathbf{u}_m^*, \left(\mathbf{y}^*\right)'\right]' \sim MVN_{2K-1}(\widetilde{\mathbf{H}}, \widetilde{\boldsymbol{\Sigma}})$, with $\mathbf{u}_m^*$ defined in Section 2.1, and $\widetilde{\mathbf{H}} = \mathbf{M}\boldsymbol{H}$ and $\widetilde{\boldsymbol{\Sigma}} = \mathbf{M}\boldsymbol{\Sigma}\mathbf{M}'$ ($\widetilde{\mathbf{G}}$ is a $(2K\text{-}1) \times 1$ vector). Next, stack the lower thresholds $\psi_{k,l_k-1}(k = 1, 2, ..., K)$ in the MC model into a $(K \times 1)$ vector $\boldsymbol{\Psi}_{low}$ and the upper thresholds $\psi_{k,l_k}(k = 1, 2, ..., K)$ into another vector $\boldsymbol{\Psi}_{up}$. If a specific discrete alternative is not consumed, place a zero value in the corresponding row of both $\boldsymbol{\Psi}_{low}$ and $\boldsymbol{\Psi}_{up}$ (technically, any value can be assigned to these non-consumption alternatives in the thresholds, since the likelihood expression derived later will not involve these entries in the thresholds). Also, stack the thresholds $\psi_{k,0}\ (k = 1, 2, ..., K)$ into a $(K \times 1)$ vector $\boldsymbol{\Psi}_0$. The vectors $\boldsymbol{\Psi}_{low}$, and $\boldsymbol{\Psi}_{up}$ are functions of the vectors $\boldsymbol{\lambda}$, $\boldsymbol{\theta}$, and $\boldsymbol{\varphi}$, while the vector $\boldsymbol{\Psi}_0$ is a function of the vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\theta}$.

Next, re-arrange the elements of the vector $\widetilde{\mathbf{G}}$ so that the elements in $\mathbf{u}_m^*$ that correspond to the consumed discrete alternatives (but not including alternative m) appear first and the elements of $\mathbf{u}_m^*$ that correspond to the non-consumed discrete alternatives appear later. Let L_C $(0 \le L_C \le K-1)$ be the number of consumed goods ($x_k^* > 0$ for these goods), but

excluding the alternative *m*). Let L_{NC} ($0 \le L_{NC} \le K-1$) correspondingly be the number of non-consumed goods ($x_k^* = 0$ for these goods) ($L_{NC} + L_C = K-1$). Also, from the $\mathbf{y}^*$ component vector of $\widetilde{\mathbf{G}}$, select out only those elements y_k^* that correspond to the consumed alternatives (including the element corresponding to alternative *m*). Both the re-arrangement of the elements of $\mathbf{u}_m^*$ as well as the selection of those elements of $\mathbf{y}^*$ corresponding to the consumed alternatives may be accomplished using a matrix $\mathbf{R}$ of dimension ($L_C + L_{NC} + L_C + 1 = L_C + K) \times (2K-1)$) so that $\breve{\mathbf{G}} = \mathbf{R}\widetilde{\mathbf{G}}$. For example, consider a consumer who chooses among five goods (*K*=5), and selects goods 2, 3, and 5 for consumption. Thus, $m = 2$, $L_C = 2$ (corresponding to the consumed goods 3 and 5, with good 2 serving as the base good needed to take utility differentials), $L_{NC} = 2$ (corresponding to the non-consumed goods 1 and 4). Then, the re-arrangement matrix $\mathbf{R}$ is:

$$\mathbf{R} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R}_C \\ \hline \mathbf{R}_{NC} \\ \hline \mathbf{R}_{y^*} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_C \\ \hline \widetilde{\mathbf{R}} \end{bmatrix}, \text{ where } \widetilde{\mathbf{R}} = \begin{bmatrix} \mathbf{R}_{NC} \\ \hline \mathbf{R}_{y^*} \end{bmatrix}. \qquad (7)$$

where the sub-matrix $\mathbf{R}_C$ corresponds to the consumed goods excluding *m* (of dimension $L_C \times (2K-1)$), the sub-matrix $\mathbf{R}_{NC}$ corresponds to the non-consumed goods (of dimension $L_{NC} \times (2K-1)$), and the sub-matrix $\mathbf{R}_{y^*}$ corresponds to the elements of the vector y^* associated with the consumed alternatives including alternative *m* (of dimension $(L_C + 1) \times (2K-1)$).

Consistent with the above re-arrangement, define $\breve{\mathbf{G}}_\mathbf{C} = \mathbf{R}_C\widetilde{\mathbf{G}}$, $\breve{\mathbf{G}}_{\mathbf{NC}} = \mathbf{R}_{NC}\widetilde{\mathbf{G}}$, $\breve{\mathbf{G}}_{\mathbf{y}^*} = \widetilde{\mathbf{R}}_{y^*}\widetilde{\mathbf{G}}$, and $\breve{\mathbf{G}}_\mathbf{2} = \widetilde{\mathbf{R}}\widetilde{\mathbf{G}} = (\breve{\mathbf{G}}'_{\mathbf{NC}}, \mathbf{G}'_{\mathbf{y}^*})'$, so that $\breve{\mathbf{G}} = (\breve{\mathbf{G}}'_\mathbf{C}, \breve{\mathbf{G}}'_{\mathbf{NC}}, \breve{\mathbf{G}}'_{\mathbf{y}^*})' = (\breve{\mathbf{G}}'_\mathbf{C}, \breve{\mathbf{G}}'_\mathbf{2})'$. In addition, let $\breve{\mathbf{H}} = \mathbf{R}\widetilde{\mathbf{H}}$, $\breve{\mathbf{H}}_\mathbf{C} = \mathbf{R}_\mathbf{C}\widetilde{\mathbf{H}}$, $\breve{\mathbf{H}}_{NC} = \mathbf{R}_{NC}\widetilde{\mathbf{H}}$, $\breve{\mathbf{H}}_{y^*} = \mathbf{R}_{y^*}\widetilde{\mathbf{H}}$,, $\breve{\mathbf{H}}_\mathbf{2} = \widetilde{\mathbf{R}}\widetilde{\mathbf{H}} = (\breve{\mathbf{H}}'_{NC}, \breve{\mathbf{H}}_{y^*})'$, and

$$\boldsymbol{\Theta} = \mathbf{R}\tilde{\boldsymbol{\Sigma}}\mathbf{R}' = \begin{bmatrix} \boldsymbol{\Theta}_C & \boldsymbol{\Theta}'_{C,NC} & \boldsymbol{\Theta}'_{C,y^*} \\ \boldsymbol{\Theta}_{C,NC} & \boldsymbol{\Theta}_{NC} & \boldsymbol{\Theta}'_{NC,y^*} \\ \boldsymbol{\Theta}_{C,y^*} & \boldsymbol{\Theta}_{NC,y^*} & \boldsymbol{\Theta}_{y^*} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Theta}_C & \boldsymbol{\Theta}'_{C2} \\ \boldsymbol{\Theta}_{C2} & \boldsymbol{\Theta}_2 \end{bmatrix}, \quad \text{where} \quad \boldsymbol{\Theta}_C = \mathbf{R}_C\tilde{\boldsymbol{\Sigma}}\mathbf{R}'_C,$$

$\boldsymbol{\Theta}_{C2} = \tilde{\mathbf{R}}\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{R}}'_C = \begin{bmatrix} \boldsymbol{\Theta}_{C,NC} \\ \boldsymbol{\Theta}_{C,y^*} \end{bmatrix}$, and $\Theta_2 = \tilde{\mathbf{R}}\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{R}}' = \begin{bmatrix} \boldsymbol{\Theta}_{NC} & \boldsymbol{\Theta}'_{NC,y^*} \\ \boldsymbol{\Theta}_{NC,y^*} & \boldsymbol{\Theta}_{y^*} \end{bmatrix}$. Also, let $\mathbf{T} = \mathbf{R}_{y^*}[., K : 2K-1]$; that is, $\mathbf{T}$ is a $(L_C+1)\times K$ sub-matrix of $\mathbf{R}_{y^*}$ with all rows of $\mathbf{R}_{y^*}$ included, but only the K^{th} through $(2K\text{-}1)^{th}$ columns of $\mathbf{R}_{y^*}$ Now, define $\tilde{\boldsymbol{\psi}}_{low} = \left[(\mathbf{T}\boldsymbol{\psi}_{low})'\right]$, where $\tilde{\boldsymbol{\psi}}_{low}$ is a $(L_C+1)\times 1-$ column vector. Similarly, define $\tilde{\boldsymbol{\psi}}_{up} = \left[(\mathbf{T}\boldsymbol{\psi}_{up})'\right]$, where $\tilde{\boldsymbol{\psi}}_{up}$ is again a $(L_C+1)\times 1-$column vector. Finally, define $\tilde{\boldsymbol{\psi}}_{0} = \left[(\mathbf{T}\boldsymbol{\psi}_{0})'\right]$.

In the rest of this section, we will use the following key notation: $f_E(.;\boldsymbol{\mu},\boldsymbol{\Delta})$ for the multivariate normal density function of dimension E with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Delta}$, $\boldsymbol{\omega}_\Delta$ for the diagonal matrix of standard deviations of $\boldsymbol{\Delta}$ (with its r^{th} element being $\omega_{\Delta,r}$), $\phi_E(.;\boldsymbol{\Delta}^*)$ for the multivariate standard normal density function of dimension E and correlation matrix $\boldsymbol{\Delta}^*$, such that $\boldsymbol{\Delta}^* = \boldsymbol{\omega}_\Delta^{-1}\boldsymbol{\Delta}\boldsymbol{\omega}_\Delta^{-1}$, $F_E(.;\boldsymbol{\mu},\boldsymbol{\Delta})$ for the multivariate normal cumulative distribution function of dimension E with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Delta}$, and $\Phi_E(.;\boldsymbol{\Delta}^*)$ for the multivariate standard normal cumulative distribution function of dimension E and correlation matrix $\boldsymbol{\Delta}^*$.

Defining $\overline{\boldsymbol{\omega}} = (\boldsymbol{\beta}, \boldsymbol{\gamma}\ or\ \boldsymbol{\alpha}, \boldsymbol{\varsigma}, \boldsymbol{\varphi}, \overline{\boldsymbol{\Sigma}})'$, where $\overline{\boldsymbol{\Sigma}}$ represents the vector of upper triangle elements of $\boldsymbol{\Sigma}$, the likelihood function contribution of the individual may be obtained from the KKT conditions in Equation (3) and from Equation (5) as:

$$L(\overline{\boldsymbol{\omega}}) = \det(\mathbf{J}) \times P(\breve{\mathbf{G}}_C = \mathbf{0}_{\mathbf{L_C}},\ \breve{\mathbf{G}}_{NC} < \mathbf{0}_{\mathbf{L_{NC}}}) \times \frac{P(\breve{\mathbf{G}}_C = \mathbf{0}_{\mathbf{L_C}},\ \breve{\mathbf{G}}_{NC} < \mathbf{0}_{\mathbf{L_{NC}}},\ \tilde{\boldsymbol{\psi}}_{low} < \tilde{\mathbf{G}}_{y^*} < \tilde{\boldsymbol{\psi}}_{up})}{\mathrm{P}(\breve{\mathbf{G}}_C = \mathbf{0}_{\mathbf{L_C}},\ \breve{\mathbf{G}}_{NC} < \mathbf{0}_{\mathbf{L_{NC}}}, \tilde{\mathbf{G}}_{y^*} > \tilde{\boldsymbol{\psi}}_{0})}, \quad (8)$$

and $\det(\mathbf{J})$ is the determinant of the Jacobian of the transformation from $\mathbf{u}_m^*$ to the consumption quantity vector $\mathbf{x}^* = (x_1, x_2, x_K)'$ (corresponding to the consumed alternatives; see Bhat, 2008):

$$\det(\mathbf{J}) = \left\{\prod_{k\in\breve{C}} \frac{1-\alpha_k}{x_k^*+\gamma_k}\right\}\left\{\sum_{k\in\breve{C}}\left(\frac{x_k^*+\gamma_k}{1-\alpha_k}\right)\left(\frac{p_k}{p_m}\right)\right\} \quad (9)$$

with $\mathfrak{C}$ being the set of alternatives consumed by the individual (including good m).

Using the marginal and conditional distribution properties of the multivariate normal distribution, we can write the second component in the likelihood function of Equation (8) as:

$$P(\breve{\mathbf{G}}_C = \mathbf{0}_{\mathbf{L_C}}, \breve{\mathbf{G}}_{NC} < \mathbf{0}_{\mathbf{L_{NC}}}) = f_{L_C}(\mathbf{0}_{\mathbf{L_C}}; \breve{\mathbf{H}}_C, \mathbf{\Theta}_{\mathbf{C}}) \times \int_{\breve{\mathbf{G}}_{NC}=-\infty_{[L_{NC}]}}^{\mathbf{0}_{\mathbf{L_{NC}}}} f_{NC}(\breve{\mathbf{G}}_{\mathbf{NC}}; \ddot{\mathbf{H}}_{NC}, \ddot{\mathbf{\Theta}}_{NC}) d\breve{\mathbf{G}}_{NC}, \tag{10}$$

The numerator of the third component in the likelihood can be written as follows:

$$P(\breve{\mathbf{G}}_C = \mathbf{0}_{\mathbf{L_C}}, \breve{\mathbf{G}}_{NC} < \mathbf{0}_{\mathbf{L_{NC}}}, \tilde{\mathbf{\Psi}}_{low} < \tilde{\mathbf{G}}_{y^*} < \tilde{\mathbf{\Psi}}_{up}) = f_{L_C}(\mathbf{0}_{\mathbf{L_C}}; \breve{\mathbf{H}}_C, \mathbf{\Theta}_{\mathbf{C}}) \times \int_{\breve{\mathbf{G}}_{NC}=-\infty_{[L_{NC}]}}^{\mathbf{0}_{\mathbf{L_{NC}}}} \int_{\tilde{\mathbf{G}}_{\mathbf{y}^*}=\tilde{\mathbf{\Psi}}_{low}}^{\tilde{\mathbf{\Psi}}_{up}} f_K(\breve{\mathbf{G}}_{\mathbf{NC}}, \tilde{\mathbf{G}}_{\mathbf{y}^*}; \ddot{\mathbf{H}}_2, \ddot{\mathbf{\Theta}}_2) d\tilde{\mathbf{G}}_{\mathbf{y}^*} d\breve{\mathbf{G}}_{NC} \tag{11}$$

The denominator of the third component in the likelihood can be written as follows:

$$\mathrm{P}(\breve{\mathbf{G}}_C = \mathbf{0}_{\mathbf{L_C}}, \breve{\mathbf{G}}_{NC} < \mathbf{0}_{\mathbf{L_{NC}}}, \tilde{\mathbf{G}}_{y^*} > \tilde{\mathbf{\Psi}}_{\boldsymbol{0}}) = f_{L_C}(\mathbf{0}_{\mathbf{L_C}}; \breve{\mathbf{H}}_C, \mathbf{\Theta}_{\mathbf{C}}) \times \int_{\breve{\mathbf{G}}_{NC}=-\infty_{[L_{NC}]}}^{\mathbf{0}_{\mathbf{L_{NC}}}} \int_{\tilde{\mathbf{G}}_{\mathbf{y}^*}=\tilde{\mathbf{\Psi}}_0}^{\infty_{[L_C+1]}} f_K(\breve{\mathbf{G}}_{\mathbf{NC}}, \tilde{\mathbf{G}}_{\mathbf{y}^*}; \ddot{\mathbf{H}}_2, \ddot{\mathbf{\Theta}}_2) d\tilde{\mathbf{G}}_{\mathbf{y}^*} d\breve{\mathbf{G}}_{NC} \tag{12}$$

Substituting expressions from Equations (10), (11) and (12), we can write Equation (8) as given below:

$$L(\overline{\boldsymbol{\omega}}) = \det(\mathbf{J}) \times f_{L_C}(\mathbf{0}_{\mathbf{L_C}}; \breve{\mathbf{H}}_C, \mathbf{\Theta}_{\mathbf{C}}) \times \int_{\breve{\mathbf{G}}_{NC}=-\infty_{[L_{NC}]}}^{\mathbf{0}_{\mathbf{L_{NC}}}} f_{NC}(\breve{\mathbf{G}}_{\mathbf{NC}}; \ddot{\mathbf{H}}_{NC}, \ddot{\mathbf{\Theta}}_{NC}) d\breve{\mathbf{G}}_{NC} \times \frac{\int_{\breve{\mathbf{G}}_{NC}=-\infty_{[L_{NC}]}}^{\mathbf{0}_{\mathbf{L_{NC}}}} \int_{\tilde{\mathbf{G}}_{\mathbf{y}^*}=\tilde{\mathbf{\Psi}}_{low}}^{\tilde{\mathbf{\Psi}}_{up}} f_K(\breve{\mathbf{G}}_{\mathbf{NC}}, \tilde{\mathbf{G}}_{\mathbf{y}^*}; \ddot{\mathbf{H}}_2, \ddot{\mathbf{\Theta}}_2) d\tilde{\mathbf{G}}_{\mathbf{y}^*} d\breve{\mathbf{G}}_{NC}}{\int_{\breve{\mathbf{G}}_{NC}=-\infty_{[L_{NC}]}}^{\mathbf{0}_{\mathbf{L_{NC}}}} \int_{\tilde{\mathbf{G}}_{\mathbf{y}^*}=\tilde{\mathbf{\Psi}}_0}^{\infty_{[L_C+1]}} f_K(\breve{\mathbf{G}}_{\mathbf{NC}}, \tilde{\mathbf{G}}_{\mathbf{y}^*}; \ddot{\mathbf{H}}_2, \ddot{\mathbf{\Theta}}_2) d\tilde{\mathbf{G}}_{\mathbf{y}^*} d\breve{\mathbf{G}}_{NC}}, \tag{13}$$

where $\ddot{\mathbf{H}}_{NC} = \breve{\mathbf{H}}_{NC} + \mathbf{\Theta}_{\mathbf{C,NC}}(\mathbf{\Theta}_{\mathbf{C}})^{\ 1}(-\breve{\mathbf{H}}_C)$, $\ddot{\mathbf{\Theta}}_{\mathbf{NC}} = \mathbf{\Theta}_{\mathbf{NC}} - \mathbf{\Theta}_{\mathbf{C,NC}}(\mathbf{\Theta}_{\mathbf{C}})^{-1}\mathbf{\Theta}'_{\mathbf{C,NC}}$,

$\ddot{\mathbf{H}}_2 = \breve{\mathbf{H}}_2 + \mathbf{\Theta}_{\mathbf{C2}}(\mathbf{\Theta}_{\mathbf{C}})^{\ 1}(-\breve{\mathbf{H}}_C)$, $\ddot{\mathbf{\Theta}}_{\mathbf{2}} = \mathbf{\Theta}_{\mathbf{2}} - \mathbf{\Theta}_{\mathbf{C2}}(\mathbf{\Theta}_{\mathbf{C}})^{-1}\mathbf{\Theta}'_{\mathbf{C2}}$, $-\infty_{[\mathbf{L_{NC}}]}$ is an $(L_{NC} \times 1)-$column vector of negative infinity values, and $\infty_{[L_C+1]}$ is a $(L_C+1)\times 1-$column vector of infinity values. Let h be an index that takes a value between 1 and (L_C+1). Let $\breve{\mathbf{\Psi}}_{0,h} = \lfloor \tilde{\mathbf{\Psi}}_{\boldsymbol{low}} \rfloor_h$, $\breve{\mathbf{\Psi}}_{1,h} = \lfloor \tilde{\mathbf{\Psi}}_{\boldsymbol{up}} \rfloor_h$,

$\ddot{\boldsymbol{\psi}}_{0,h} = \left\lfloor \tilde{\boldsymbol{\psi}}_{\boldsymbol{0}} \right\rfloor_h$, and $\ddot{\boldsymbol{\psi}}_{1,h} = \left\lfloor \infty_{[L_C+1]} \right\rfloor_h$. Also, let $\boldsymbol{\Theta}_C^* = \boldsymbol{\omega}_{\Theta_C}^{-1} \boldsymbol{\Theta}_C \boldsymbol{\omega}_{\Theta_C}^{-1}$, $\ddot{\boldsymbol{\Theta}}_2^* = \boldsymbol{\omega}_{\ddot{\Theta}_2}^{-1} \ddot{\boldsymbol{\Theta}}_2 \boldsymbol{\omega}_{\ddot{\Theta}_2}^{-1}$, and $\boldsymbol{\Theta}_{NC}^* = \boldsymbol{\omega}_{\ddot{\Theta}_{NC}}^{-1} \boldsymbol{\Theta}_{NC} \boldsymbol{\omega}_{\ddot{\Theta}_{NC}}^{-1}$. The three integrals in Equation (13) maybe written as:

$$L_{NUM1} = F_K\left(\mathbf{0}_{L_{NC}}; \ddot{\mathbf{H}}_{\boldsymbol{NC}}, \ddot{\boldsymbol{\Theta}}_{NC}\right) = \Phi_K\left(\boldsymbol{\omega}_{\ddot{\Theta}_{\mathbf{NC}}}^{-1}\left[-\ddot{\mathbf{H}}_{\boldsymbol{NC}}\right] \; ; \ddot{\boldsymbol{\Theta}}_{\mathbf{NC}}^*\right) \tag{14}$$

$$\begin{aligned} L_{NUM2} &= \sum_{a_1=1}^{2}\sum_{a_2=1}^{2}\cdots\sum_{a_{L_C+1}=1}^{2}(-1)^{a_1+a_2+\ldots a_{L_C+1}}\left[F_K\left((\mathbf{0}'_{L_{NC}}, \breve{\boldsymbol{\psi}}_{a_1-1,1}, \breve{\boldsymbol{\psi}}_{a_2-1,2}, \ldots \breve{\boldsymbol{\psi}}_{a_{L_C+1}-1,L_C+1})'; \ddot{\mathbf{H}}_{\boldsymbol{2}}, \ddot{\boldsymbol{\Theta}}_{\mathbf{2}}\right)\right] \\ &= \sum_{a_1=1}^{2}\sum_{a_2=1}^{2}\cdots\sum_{a_{L_C+1}=1}^{2}(-1)^{a_1+a_2+\ldots a_{L_C+1}}\left[\Phi_K\left(\boldsymbol{\omega}_{\ddot{\Theta}_2}^{-1}\left[(\mathbf{0}'_{L_{NC}}, \breve{\boldsymbol{\psi}}_{a_1-1,1}, \breve{\boldsymbol{\psi}}_{a_2-1,2}, \ldots \breve{\boldsymbol{\psi}}_{a_{L_C+1}-1,L_C+1})' - \ddot{\mathbf{H}}_{\boldsymbol{2}}\right]; \ddot{\boldsymbol{\Theta}}_{\mathbf{2}}^*\right)\right] \end{aligned} \tag{15}$$

The integral in the denominator may be written as:

$$\begin{aligned} L_{DEN} &= \sum_{a_1=1}^{2}\sum_{a_2=1}^{2}\cdots\sum_{a_{L_C+1}=1}^{2}(-1)^{a_1+a_2+\ldots a_{L_C+1}}\left[F_K\left((\mathbf{0}'_{L_{NC}}, \ddot{\boldsymbol{\psi}}_{a_1-1,1}, \ddot{\boldsymbol{\psi}}_{a_2-1,2}, \ldots \ddot{\boldsymbol{\psi}}_{a_{L_C+1}-1,L_C+1})'; \ddot{\mathbf{H}}_{\boldsymbol{2}}, \ddot{\boldsymbol{\Theta}}_{\mathbf{2}}\right)\right] \\ &= \sum_{a_1=1}^{2}\sum_{a_2=1}^{2}\cdots\sum_{a_{L_C+1}=1}^{2}(-1)^{a_1+a_2+\ldots a_{L_C+1}}\left[\Phi_K\left(\boldsymbol{\omega}_{\ddot{\Theta}_2}^{-1}\left[(\mathbf{0}'_{L_{NC}}, \ddot{\boldsymbol{\psi}}_{a_1-1,1}, \ddot{\boldsymbol{\psi}}_{a_2-1,2}, \ldots \ddot{\boldsymbol{\psi}}_{a_{L_C+1}-1,L_C+1})' - \ddot{\mathbf{H}}_{\boldsymbol{2}}\right]; \ddot{\boldsymbol{\Theta}}_{\mathbf{2}}^*\right)\right] \end{aligned} \tag{16}$$

The expressions L_{NUM1}, L_{NUM2} and L_{DEN} may be computed using simulation-based methods or an analytic approximation approach to approximate the MVNCD functions. Typical simulation-based methods can get inaccurate and time-consuming as the dimensionality increases. On the other hand, the analytic approximation approach of Joe (1995) and Bhat (2011) 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. This is the approach used in the current paper. The accuracy and stability of the analytic approximation approach for the MVNCD function has already been evaluated for the multinomial probit model (Bhat and Sidharthan, 2011). These results indicate that the approximation provides parameter values very close to the "true" population parameter values in simulation experiments, with the empirical absolute percentage bias being smaller than that from simulation-based techniques to evaluate the MVNCD function. Further, the time to convergence using the analytic approximation is an order less than the time to convergence using simulation-based approaches. Recently, Bhat *et al.* (2013) have demonstrated the ability of the analytic approximation to recover parameters very accurately even for MDCP models. They also noted that, for the typical size of samples employed in discrete model estimation, the asymptotic standard errors computed using the

second derivatives of the analytic approximation-based likelihood function provides a very good estimate of the true finite sample error. This is not surprising, because the MVNCD-approximated log-likelihood function is close to the log-likelihood function for all parameters in a neighborhood of the "true" parameter values, which implies that the covariance matrix computed using the analytic approximation should be accurate for the actual covariance matrix. Here, we extend the use of the analytic approximation to estimate the joint MDC-MC model of this paper.

The likelihood contribution of the individual in Equation (8) collapses to the expression given below:

$$L(\varpi) = \det(\mathbf{J}) \times \left(\prod_{g=1}^{L_C} \omega_{\Theta_{C,g}} \right)^{-1} \left(\phi_{L_C} (\boldsymbol{\omega}_{\Theta_C}^{-1} (-\breve{\mathbf{H}}_C), \boldsymbol{\Theta}_C^*) \right) \times L_{NUM\,1} \frac{L_{NUM\,2}}{L_{DEN}} \tag{17}$$

Several constrained versions of the model just discussed may be obtained. If the error covariance matrices $\boldsymbol{\Theta}_{C,y^*}$ and $\boldsymbol{\Theta}_{NC,y^*}$ are matrices with all elements being zeros (that is, if there is no dependence between the marginal utility vector $\mathbf{U}$ in the MDCP model and the latent variable vector $\mathbf{y}^*$ underlying the count outcomes), then the likelihood function of Equation (17) can be easily shown to collapse to an independent MDCP model and an independent multivariate hurdle count model (with the hurdle for the count of alternative *k* being whether or not the consumer consumes some amount of the alternative *k*, as determined in the MDCP model). Further, if $\boldsymbol{\Theta}_{y^*}$ is a diagonal matrix, then the multivariate hurdle count model collapses to independent hurdle count models for each discrete alternative *k*. However, note that the resulting independent hurdle count model structure for each discrete alternative is still more general than the traditional Poisson hurdle count model structure. Specifically, only if all elements of the vectors $\boldsymbol{\varsigma}$ and $\boldsymbol{\varphi}$ are identically zero will the structure collapse to a traditional Poisson hurdle count model.

An estimation consideration that needs to be dealt with is that the matrix $\boldsymbol{\Theta}$ for any individual must be positive definite. The simplest way is to ensure that the matrix $\tilde{\boldsymbol{\Sigma}}$ for each individual is positive definite, which can be guaranteed by using a Cholesky-decomposition of the matrix $\boldsymbol{\Sigma}$. Note that, to obtain the Cholesky factor for $\boldsymbol{\Sigma}$, we first obtain the Cholesky factor for $\boldsymbol{\Sigma}_1$ (see Equation 6), and then add a column of zeros as the first column and a row of zeros as the first row to obtain the Cholesky factor of $\boldsymbol{\Sigma}_1$. However, the top diagonal element of $\boldsymbol{\Sigma}_1$ has

to be normalized to one if there is no price variation across goods for each consumer (as discussed earlier in Section 3). Also, the matrix $\boldsymbol{\Gamma}$, which is embedded in $\boldsymbol{\Sigma}_1$, is a correlation matrix. These restrictions need to be recognized when using the Cholesky factor of $\boldsymbol{\Sigma}_1$. To do so, consider the lower triangular Choleski matrix $\breve{\mathbf{L}}$ of the same size as $\boldsymbol{\Sigma}_1$. Whenever a diagonal element (say the aa^{th} element) of $\boldsymbol{\Sigma}_1$ is to be normalized to one, the corresponding diagonal element of $\breve{\mathbf{L}}$ is written as $\sqrt{1-\sum_{j=1}^{a-1} d_{aj}^2}$, where the d_{aj} elements are the Cholesky factors that are to be estimated. With this parameterization, $\boldsymbol{\Sigma}_1$ obtained as $\breve{\mathbf{L}}\breve{\mathbf{L}}'$ is positive definite and adheres to the scaling conditions.

Thus far, the discussion has assumed that there is no essential outside numeraire good (*i.e.*, no essential Hicksian composite good). If an outside good is present, label the outside good as the first good which now has a unit price of one (*i.e.*, $p_1 = 1$). This good, being an essential good, serves as a convenient base alternative to take utility differences off (that is, in our earlier notation, m=1 for all consumers). The utility functional form of Equation (4) now needs to be modified as follows:

$$\max U(\mathbf{x}) = \frac{1}{\alpha_1}\psi_1(x_1+\gamma_1)^{\alpha_1} + \sum_{k=2}^{K}\frac{\gamma_k}{\alpha_k}\psi_k\left(\left(\frac{x_k}{\gamma_k}+1\right)^{\alpha_k}-1\right) \tag{18}$$

In the above formula, we need $\gamma_1 \le 0$, while $\gamma_k > 0$ for $k > 1$. Also, we need $x_1 + \gamma_1 > 0$. The magnitude of γ_1 may be interpreted as the required lower bound (or a "subsistence value") for consumption of the outside good (Bhat, 2008). As in the "no-outside good" case, the analyst will generally not be able to estimate both α_k and γ_k for the outside and inside goods. For identification purposes, we assume (without loss of generality) that $\psi_1 = \exp(\xi_1)$. Corresponding to the utility function above, $V_1 = (\alpha_1 - 1)\ln\left(x_1^* + \gamma_1\right)$, $V_k = \boldsymbol{\beta}'\mathbf{z}_k + (\alpha_k - 1)\ln\left(\frac{x_k^*}{\gamma_k}+1\right) - \ln p_k$ for k>1, $U_k = V_k + \xi_k$ for all k, and $u_{km}^* = U_k - U_m$, where m=1 now. All other notations remain the same. In the case in which the outside good does not have a count associated with it (such as when the outside good is "in-home time" in a model of out-of-home time investments in different activity purposes and corresponding number of out-of-home episodes), everything

remains the same as earlier except for minor modifications to the re-arrangement matrix and related matrices so that there are no count parameters to estimate for this outside good.

4. SIMULATION EVALUATION

The simulation exercise undertaken in this section examines the ability of the analytic approximation to recover parameters from finite samples in a joint MDCP-MC model by generating simulated data sets with known underlying model parameters. In addition, the exercise examines the effects of imposing a restrictive independence assumption between the MDCP and the MC components, when the true data generating process is a joint MDCP-MC process.

4.1 Simulation Design and Evaluation

Consider a three-alternative MDCP model, and the case when all alternatives may have corner solutions (that is, the case with no essential outside good). We specify a single independent variable in the z_k vector in the baseline utility of the three alternatives. The values of this variable for each of the three alternatives are drawn from standard univariate normal distributions, and a synthetic sample of 2000 realizations of the exogenous variables is generated, corresponding to a simulated data set of Q=2000 observations. The coefficient on this variable (labeled as β) is set to the value of 1. In the simulations, we use a γ-profile, and set all the γ parameters to the value of one (so, $\boldsymbol{\gamma}=(\gamma_1,\gamma_2,\gamma_3)'=(1,1,1)'$).

The covariance matrix that generates the jointness among the baseline utilities of the MDCP alternatives as well as the error terms in the count variables is specified as follows (see Section 3):

$$Var(\mathbf{G})=\boldsymbol{\Sigma}=\begin{bmatrix}0 & \mathbf{0}' & \mathbf{0}'\\ \mathbf{0} & \boldsymbol{\Lambda}_1 & \boldsymbol{\Xi}'\\ \mathbf{0} & \boldsymbol{\Xi} & \boldsymbol{\Gamma}\end{bmatrix}=\begin{bmatrix}0 & \mathbf{0}'\\ \mathbf{0} & \boldsymbol{\Sigma}_1\end{bmatrix}, \text{ where}$$

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 1.000 & 0.600 & 0.400 & 0.000 & 0.000 \\ 0.600 & 1.360 & 0.600 & 0.000 & 0.000 \\ 0.400 & 0.600 & 1.000 & 0.400 & 0.320 \\ 0.000 & 0.000 & 0.400 & 1.000 & 0.438 \\ 0.000 & 0.000 & 0.320 & 0.438 & 1.000 \end{bmatrix}$$

$$= \breve{\mathbf{L}}_{\boldsymbol{\Sigma}_1} \breve{\mathbf{L}}'_{\boldsymbol{\Sigma}_1} = \begin{bmatrix} 1.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.600 & 1.000 & 0.000 & 0.000 & 0.000 \\ 0.400 & 0.360 & 0.843 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.475 & 0.880 & 0.000 \\ 0.000 & 0.000 & 0.380 & 0.293 & 0.878 \end{bmatrix} \begin{bmatrix} 1.000 & 0.600 & 0.400 & 0.000 & 0.000 \\ 0.000 & 1.000 & 0.360 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.843 & 0.475 & 0.380 \\ 0.000 & 0.000 & 0.000 & 0.880 & 0.293 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.878 \end{bmatrix}$$

As indicated earlier, the positive definiteness of $\boldsymbol{\Sigma}_1$ is ensured in the estimations by reparameterizing the likelihood function in terms of the lower Cholesky factor $\breve{\mathbf{L}}_{\boldsymbol{\Sigma}_1}$, and estimating the six associated Cholesky matrix parameters (note that the Cholesky parameters corresponding to fixed normalization values of 1.000 in the covariance matrix $\boldsymbol{\Sigma}_1$ are not estimated, but are obtained from the other elements in that row): $l_{\boldsymbol{\Sigma}_1,1} = 0.600$, $l_{\boldsymbol{\Sigma}_1,2} = 1.000$, $l_{\boldsymbol{\Sigma}_1,3} = 0.400$, $l_{\boldsymbol{\Sigma}_1,4} = 0.360$, $l_{\boldsymbol{\Sigma}_1,5} = 0.475$, $l_{\boldsymbol{\Sigma}_1,6} = 0.380$, and $l_{\boldsymbol{\Sigma}_1,7} = 0.293$ We will also refer to these parameters collectively as $\mathbf{l}_{\boldsymbol{\Sigma}_1}$.

For the count components, we consider a single exogenous variable in the $\mathbf{s}_k$ vector for the count model for each discrete alternative (embedded in the threshold function). This exogenous variable (for the count model corresponding to each discrete alternative) is generated from a standard univariate distribution. The corresponding coefficient vector (labeled as $\boldsymbol{\varsigma} = (\varsigma_1, \varsigma_2, \varsigma_3)'$) is set to $(0.50, 0.25, 0.50)'$. For the $\boldsymbol{\varphi}_k = (\varphi_{k,1}, \varphi_{k,2}, \ldots \varphi_{k,e_k^*})'$ vector, we set $e_k^* = 1 \; \forall k$, so that only one threshold $\varphi_{k,1}$ is to be estimated for the count model corresponding to each discrete alternative *k*. In the data generation, we set $\boldsymbol{\varphi} = (\varphi_{1,1}, \varphi_{2,1}, \varphi_{3,1})' = (1, 0.5, 0.75)'$.

Using the parameters specified above, we first compute the vector **H** (see Section 3). Then, given **H** and $\boldsymbol{\Sigma}$, we have the distribution of the vector $\mathbf{G} = \left[\mathbf{U}', \left(\mathbf{y}^* \right)' \right]'$. Then, for each of the 2000 observations, we draw a realization of **G** from its multivariate truncated normal distribution. Next, using a γ-profile and the corresponding "true" values of the $\boldsymbol{\gamma}$ vector, and the

realization of the $\mathbf{U}$ vector, we generate the consumption quantity vector $\boldsymbol{x}_q^*$ for each individual, using the forecasting algorithm proposed by Pinjari and Bhat (2011). Similarly, using the values of $\varphi_{k,1}$ (k=1,2,3), the $\boldsymbol{\varsigma}$ vector values, and the realizations of the exogenous variable in the $\mathbf{s}_k$ vector, we compute the threshold values (the ψ_{k,l_k} values in Equation 5) and translate the realization of the $\mathbf{y}^*$ vector to a multivariate count value (across alternatives). The above data generation process is undertaken 50 times with different realizations of the **G** vector to generate 50 different data sets, each with 2000 observations. The MACML estimator is applied to each data set to estimate data-specific values of the 17x1 column vector $(\beta, \gamma_1, \gamma_2, \gamma_3, \varsigma_1, \varsigma_2, \varsigma_3, \varphi_{1,1}, \varphi_{2,1}, \varphi_{3,1}, \mathbf{l}_{\boldsymbol{\Sigma}_1})$. 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 MVNCD function into a product sequence of marginal and conditional probabilities (see Section 2.1 of Bhat, 2011). The estimator is applied to each dataset 10 times with different permutations to obtain the approximation error.

The performance of the proposed inference approach in estimating the parameters of the proposed model and the corresponding standard errors is evaluated as follows:

(1) Estimate the 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) For each data set *s*, compute the mean estimate for each model parameter across the 10 random permutations used. Label this as MED, and then take the mean of the MED values across the data sets to obtain **a mean estimate**. Compute the **absolute percentage (finite sample) bias** (APB) of the estimator as:

$$APB = \left| \frac{\text{mean estimate - true value}}{\text{true value}} \right| \times 100$$ [6]

(3) Compute the standard deviation of the MED values across data sets, and label this as the **finite sample standard error or FSSE** (essentially, this is the empirical standard error).

(4) For each data set, compute the mean s.e. for each model parameter across the 10 draws. Call this MSED, and then take the mean of the MSED values across the 50 data sets and

[6] If the true parameter value is zero, the APB value is computed by dividing the mean estimate by the value of 1 in the denominator, and multiplying by 100.

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 gets large).

(5) Next, to evaluate the accuracy of the asymptotic standard error formula as computed using the inference approach for the finite sample size used, compute **a relative efficiency (RE)** value as:

$$RE = \frac{\text{ASE}}{\text{FSEE}}$$

Relative efficiency values in the range of 0.8-1.2 indicate that the ASE, as computed using the Godambe matrix in the CML method, does provide a good approximation of the FSSE. In general, the relative efficiency values should be less than 1, since we expect the asymptotic standard error to be less than the FSSE. But, because we are using only a limited number of data sets to compute the FSSE, values higher than one can also occur. The more important point is to examine the closeness between the ASE and FSEE, as captured by the 0.8-1.2 range for the relative efficiency value.

(6) Compute the standard deviation of the parameter values around the MED parameter value for each data set, and take the mean of this standard deviation value across data sets; label this as the **approximation error (APERR).**

4.2 Comparison with Restrictive Independent Model

The main purpose of the methodology proposed here is to accommodate the jointness in the MDC and the MC decisions, while ensuring that there is a positive count in a certain discrete category only if there is some positive continuous consumption in that category. To examine the implication of ignoring this jointness when it is actually present, we estimate a restrictive model on the 50 data sets generated as per the design discussed in the previous section. Then, we estimate an independent model that ignores the jointness between the MDC and MC dimensions by specifying the covariance matrix $\mathbf{\Sigma_1}$ as follows:

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 1.000 & 0.600 & 0.000 & 0.000 & 0.000 \\ 0.600 & 1.360 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 1.000 & 0.400 & 0.320 \\ 0.000 & 0.000 & 0.400 & 1.000 & 0.438 \\ 0.000 & 0.000 & 0.320 & 0.438 & 1.000 \end{bmatrix}$$

$$= \breve{\mathbf{L}}_{\boldsymbol{\Sigma}_1} \breve{\mathbf{L}}'_{\boldsymbol{\Sigma}_1} = \begin{bmatrix} 1.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.600 & 1.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.843 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.475 & 0.880 & 0.000 \\ 0.000 & 0.000 & 0.380 & 0.293 & 0.878 \end{bmatrix} \begin{bmatrix} 1.000 & 0.600 & 0.000 & 0.000 & 0.000 \\ 0.000 & 1.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.843 & 0.475 & 0.380 \\ 0.000 & 0.000 & 0.000 & 0.880 & 0.293 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.878 \end{bmatrix}$$

In the above specification, we restrict $l_{\boldsymbol{\Sigma}_1,3}$ and $l_{\boldsymbol{\Sigma}_1,4}$ to zero, and examine the APB values for the other parameters in the resulting independent model relative to the joint model. We also compare the independent and joint models based on a likelihood ratio test (LRT).

For the comparison between the independent and joint models, we use a single replication per data set (the replication is the same one for both models; that is, we use a single permutation per individual that varies across individuals but is held fixed across the two models models). We do so rather than run 10 replications for each of the models (as done for evaluating recovery of parameters in the joint model) because, as we will present in the next section, the approximation error in the parameters is negligible for any given data set. The LRT statistic needs to be computed for each data set separately, and compared with the chi-squared table value with two degrees of freedom. In this paper, we identify the number of times (corresponding to the 50 model runs, one run for each of the 50 data sets) that the LRT value rejects the independent model in favor of the joint model.

4.3 Simulation Results

4.3.1 Recoverability of Parameters in the Joint MDC-MC Model

The results of the simulation exercise to evaluate the ability of the MACML approach to recover the parameters of the joint model are presented in Table 1. The table shows that the average estimates of parameters are close to their true values used in the data generation process. The overall APB value across parameters is just 5.8% (see the last row of the table under the column "APB"); however, the APB does vary across parameters. The β parameter of the baseline utility of the MDCP component of the model is recovered quite well with an APB of only 6.1%. The

translation parameters in the $\boldsymbol{\gamma}$ vector of the MDCP component of the model has an average APB of 9.8%, but the APB of the first and third alternative is on the relatively high side with an APB value of 15.6% and 11.5%. This is not surprising, because the satiation parameters enter the utility function rather non-linearly (see Equation 1). As a consequence, it becomes difficult to pin down the $\boldsymbol{\gamma}$ parameter vector, because a range of values of the $\boldsymbol{\gamma}$ parameter vector produce a similar value for the probability of the MDC choice. The elements of the parameter vector $\boldsymbol{\varsigma}$ embedded in the thresholds of the count model is recovered very well (with an average APB value of 3.9%), as are the elements of the threshold offset parameter $\boldsymbol{\varphi}$ (with an average APB across parameters of 4.3%). Finally, the average APB for the elements of the Cholesky of the covariance matrix $\boldsymbol{\Sigma_1}$ is 5.4%, with all APB values less than 10%.

The finite sample standard errors (FSSE) are also quite small, averaging only about 9.9% of the true value of the parameters, indicating good empirical efficiency of the proposed estimator. Among the non-covariance parameters, the FSSE estimates (as a percentage of the true value) are generally higher for the $\boldsymbol{\varsigma}$ vector elements of the count model (20.3%) compared to the other parameters (5.3%). This is to be expected since the $\boldsymbol{\varsigma}$ vector affects the count thresholds in a non-linear fashion, and a whole range of values of the $\boldsymbol{\varsigma}$ vector elements around the true value lead to similar probability values for the counts. In the set of Cholesky elements, the FSSE of the MDCP-associated terms ($l_{\Sigma_1,1} = 0.600$, $l_{\Sigma_1,2} = 1.000$) are much lower than the FSSE for the other Cholesky elements. This is due to the fact that the MDCP error covariance matrix is associated with both the discrete and continuous elements of choice, and so is more easily pinned down than the count model error covariance matrix that is associated with the count element of choice.

A comparison of the finite sample standard errors and the asymptotic standard errors reveals that these error values are very close, with the relative efficiency (RE) being between 0.9-1.1 for all but four parameters. All efficiency values are within the 0.8-1.2 range. Overall, across all parameters, the average relative efficiency is 1.01, indicating that there is effectively no difference between the finite sample size standard errors and the approximation to these finite sample standard errors as computed by the asymptotic formula for the standard errors. That is, the asymptotic assumptions are working well for the dataset size used in our simulation experiment (which also is quite typical for model estimation in the transportation and other fields).

Finally, the last column of Table 1 presents the approximation error (APERR) for each of the parameters, because of the use of different permutations in the analytic approximation method in the MVNCD evaluation. These entries indicate that the APERR is of the order of 0.015 or less. More importantly, the approximation error (as a percentage of the FSEE or the ASE), averaged across all the parameters, is of the order of 9% of the sampling error. This is clear evidence that even a single permutation (per observation) of the analytic approximation provides adequate precision, in the sense that the convergent values are about the same for a given data set regardless of the permutation used for the decomposition of the multivariate probability expression.

4.3.2 Effects of Ignoring Jointness in the MDCP and MC Model Systems

This section presents the results of the estimation when the endogeneity in the participation selection system from the MDCP model in the estimation of the MC data system (in a hurdle-type model) is ignored. As discussed earlier in Section 4.2, this is tantamount to restricting $l_{\Sigma_1,3}$ and $l_{\Sigma_1,4}$ to zero. A comparison of the resulting independent model with the joint model proposed in this paper provides a sense of the biases that may accrue because of using a restrictive specification.

Table 2 presents the results of the estimations of the restrictive independent model and the proposed joint model. As expected, the results clearly show a deterioration in the APB values of the estimates in the independent model. The overall APB is 8.9% in the independent model compared to 6.1% in the joint model. However, even this is deceiving because it considers both the parameters of the MDCP and the MC models. The MDCP model parameters are likely to be less affected by ignoring jointness, as also indicated by the relatively similar APB values for the $\beta, \gamma_1, \gamma_2, \gamma_3, l_{\Sigma_1,1}, and\ l_{\Sigma_1,2}$ parameters (all these parameters are exclusive to the MDCP model; the average APB for these parameters in the joint model is 7.4% relative to 7.5% in the independent model). The real difference shows up in the parameters associated with the MC model. Indeed, the average APB for the nine parameters in the MC model ($\varsigma_1, \varsigma_2, \varsigma_3, \varphi_{1,1}, \varphi_{2,1}, \varphi_{3,1}$, $l_{\Sigma_1,5}$, $l_{\Sigma_1,6}$, and $l_{\Sigma_1,7}$) for the joint model is 5.1% compared to 9.8% in the independent model. The APB of the $l_{\Sigma_1,5}$, $l_{\Sigma_1,6}$, and $l_{\Sigma_1,7}$ parameters, in particular, shoot up to over 15% in the independent model. The superiority of the joint model is further reinforced by the LRT with two

degrees of freedom. The table chi-squared value with two degrees of freedom is 5.99 at the 95% confidence level, and the LRT value between the joint and independent models exceeds this value for each of the 50 data sets used in our simulation. In fact, the LRT rejects the independent model in favor of the joint model at even the 99% confidence level for each of the 50 data sets (the mean value of the test statistic is 137).

Overall, the simulation results show that the estimator recovers the parameters of the proposed joint model well. The estimator also seems to be quite efficient based on the low FSEE estimates. Further, the asymptotic standard error formula estimates the FSEE quite well, and the approximation error due to the use of the analytic approximation is very small. Additionally, the results clearly highlight the bias in estimates if the endogeneity of the MDC model is ignored.

5. ILLUSTRATIVE APPLICATION TO HOUSEHOLD ACTIVITY PARTICIPATION, TIME USE, AND NUMBER OF EPISODES

5.1 Background

The multivariate hurdle count data model with an endogenous MDC selection system proposed in this paper can be applied to several empirical problems. In this section, we demonstrate the application of the proposed model to analyze the participation of household members in each of several activity purposes during the day, along with the amount of time invested in each activity purpose and the number of distinct episodes of each activity purpose.

In our empirical demonstration, we use the household as the unit of analysis rather than an individual. This is because, as argued by Bhat *et al*. (2013), household members are likely to act as a collective decision-making unit in activity time-use choices and be influenced by the preferences of other individuals in the household (even if they participate individually in specific activity purposes).

5.2 Data Source and Sample Formation

The data used in the analysis is drawn from the 2000 Post-Census household travel survey, conducted by the Southern California Association of Governments. The survey obtained information from about 17,000 households, and recorded all travel and out-of-home activity episodes undertaken by each household member for a pre-specified survey day. In addition, the survey also collected detailed socio-demographic and employment-related characteristics. The survey area comprised the six-county Los Angeles region of California.

The sample formation included the following steps. The activity diaries for weekends, Mondays, and Fridays were excluded, leaving only the midweek days (Tuesday, Wednesday, and Thursday). The work and work-related episodes of individuals were then removed, because work and work-related decisions (employment decisions, number of hours of work, and work timings) usually tend to be made on a relatively longer term basis compared to the day-to-day planning and scheduling of non-work activity episodes (Rajagopalan *et al.,* 2009, Horner and O'Kelly, 2007, and Saleh and Farrell, 2005). Next, we collapsed the remaining 23 category non-work-related activity purpose taxonomy into four activity purposes: (1) shopping (including grocery shopping, clothes shopping, window shopping, purchasing gas, quick stop for coffee/newspaper maintenance), (2) social activities (including dining out, visiting friends and family, community meetings, political/civic event, public hearing, occasional volunteer work, church, temple and religious meeting), (3) recreation (including watching sports or attending a sports event, going to the movies/opera, going dancing, visiting a bar, going to the gym, playing sports, bicycling, walking, and camping), and (4) personal activities (including ATM and other banking, visiting post office, banking, paying bills, and medical/doctor visits).[7] The activity episodes of each household member were then assigned to one of the four activity purposes identified above. The durations of episodes were aggregated by purpose to obtain the total weekday duration in each activity purpose for each household member. At the same time, a count of the number of episodes of each activity purpose was also obtained at the individual level. Next, the individual-level durations and episode counts by activity purpose were aggregated across all individuals in the household to obtain household-level durations and episode counts by activity purpose, which formed the dependent variables in the study. Finally, a random sample of 2,110 households was selected.

5.3 Construction of Accessibility Measures

In addition to the 2000 SCAG survey data set, several other secondary data sets were used to obtain residential neighborhood accessibility measures that may influence household-level activity participation behavior. The secondary data sources included geo-coded block group and

[7] There is obviously some subjectivity in the activity purpose classification adopted here, though the overall consideration was to accommodate differences between the activity purposes along such contextual dimensions as location of participation, physical intensity level, duration of participation, amount of structure in activity planning, and company type of participation (see Srinivasan and Bhat, 2005).

block data within the SCAG region obtained from the Census website, roadway network skims from SCAG, the employment data from the Census Transportation Planning Package (CTPP) and Dun & Bradstreet (D&B), the 2000 Public-Use Microdata Samples (PUMS) from Census 2000, and the marginal distributions (population and household summary tables) from SCAG.

Two types of accessibility measures were constructed in the current analysis. The first set of accessibility measures represents opportunity-based indicators that measure the number of activity opportunities by twelve different industry types that can be reached within 10 minutes (on the highway network) from the centroid of the home block during the morning peak period (6am to 9am). The reader is referred to Chen *et al*. (2011) for details. These may be viewed as local accessibility measures. In addition to these activity opportunity local accessibility measures, we also computed a travel opportunity local accessibility measure as the length of freeways (in thousands of kilometers) accessible within 10 minutes from the centroid of the home block during the morning period. The second set of accessibility indicators corresponds to Hansen type zonal-level regional accessibility measures (Bhat and Guo, 2007), which take the following form:

$$Acc_{i,\tilde{t}} = \frac{1}{N}\sum_{j=1}^{N}\left(\frac{\text{Size Measure}_{\text{j}}}{\text{Impedance}_{\text{ij},\tilde{\text{t}}}}\right),$$

where i is the index for zone, $\tilde{t}$ is the index for the time period, and N is the total number of zones in the study region (four time periods were used in our analysis: AM peak (6:30 am-9 am), midday (9 am-4 pm), PM peak (4 pm-6:30 pm), and evening (6:30 pm-6:30am)). $\text{Impedance}_{\text{ij},\tilde{\text{t}}}$ is the composite impedance measure of travel between zones i and j at time period $\tilde{t}$ and is obtained as: $\text{Impedance}_{\text{ij},\tilde{\text{t}}} = IVTT_{ij,\tilde{t}} + \lambda Cost_{ij,\tilde{t}}$, where $IVTT_{ij,\tilde{t}}$ and $Cost_{ij,\tilde{t}}$ are the auto travel time (in minutes) and auto travel cost (in cents), respectively, between zones i and j in time period $\tilde{t}$, and λ is the inverse of the money value of travel time. We used $\lambda = 0.0992$ in the current study, which corresponds to about \$6 per hour of implied money value of travel time. For the zonal size measure in the accessibility formulation, we considered four variables -- retail employment, retail and service employment, total employment, and population. Finally, the time period-specific accessibility measures computed as discussed above were weighted by the durations of each time period, and a composite daily accessibility measure (for each size

measure) was computed for each traffic analysis zone, and appended to sample households based on the residence TAZs of households.[8]

5.4 Sample Description

Table 3 presents a descriptive summary of the demographics of the sample. About 28% of the sample has a single person, which is slightly higher than the 22% of single person households reported in the 2000 Census for the Los Angeles/Riverside/Orange County (LRO) metropolitan statistical area (MSA). Similarly, the percentage of households that are couple households (without children) is about 29% in the sample, compared to 24% in the 2000 Census data (in the rest of this paper, a child will be defined as an individual of age 15 years or younger, who is a son or daughter of an adult in the household). On the other hand, a little over 3% of the sample corresponds to single-parent households, which is an underestimate relative to the percentage of single-parent households as reported in the 2000 Census. The remaining households are categorized as "other" households and mainly correspond to nuclear family households (representing a heterosexual union with one or more children 15 years or younger in the household). Overall, however, the sample is not unreasonable in terms of representing the population household structure in the LRO MSA. The table also shows the distribution of household income in the sample. Nearly 50% of the households in the sample has an income lower than $50,000, which is close to the percentage of households in that income range in the NHTS 2001 data for LRO MSA. The mean household income in the sample is $62,000.

The descriptive statistics of other demographics, including household race and ethnicity, housing type and tenure, bicycle ownership, household size-related attributes (number of children, number of adults, and number of workers), and other household attributes (number of motorized vehicles and accessibility measures) are also provided in Table 3, and indicate the diverse and high vehicle-owning nature of households in the LRO MSA.

The bottom panel of Table 3 provides the descriptive statistics of household-level activity participation decisions (the dependent variables) in the final estimation dataset, including the (1) number and percentage of households who participate in each activity purpose during the survey weekday, (2) the mean duration of daily time investment among households who participate in

[8] Future studies would benefit from exploring alternate forms of accessibility as well as the consideration of transit and non-motorized mode network skims (in addition to the highway network skims used here). The transit and non-motorized mode skims were not considered in our study due to data-related quality limitations.

each activity purpose, (3) the mean number of daily episodes of participation in each activity purpose, conditional on participation in each activity purpose, and (4) the percentage of households participating in each activity purpose who solely participate in that activity and who also participate in other activity purposes (the last two columns; the sum of these last two columns is 100% for each row).

The descriptive statistics in the first numeric column in the bottom panel of Table 3 reveal that households (*i.e.*, across all individuals in the household) are most unlikely during the weekday to participate in recreational activities (such as entertainment and sports). However, more than half of all households participate in shopping, social, and personal business activities. The "mean duration of daily time investment among households who participate" column shows the high overall daily time investments of participating households in social activities (over four hours) and recreational activities (over six hours). These may seem quite high, but it should be noted that these time investments are across all individuals in a household. That is, these time durations refer to individual minutes of participation across all individuals in a household.[9]

An interesting observation from the duration statistics in Table 3 is that, while recreation activity is the least participated in, on average, it receives the highest time investment from participating households relative to other activity purposes. This suggests that there is much less satiation (or drop off in marginal valuation) in recreation activity than in other activity purposes, which is not surprising given the nature of recreation and other activity purposes. The purpose with the least time investment is the shopping purpose, with a mean duration of about 100 minutes. Also interesting to note is the lower mean number of recreation episodes relative to other types of episodes. Overall, households participate the least in recreational activity, and even if they participate in recreational activity, do so in very few episodes. However, once a participation decision has been made in recreational activity, the time duration is high. On the other hand, while daily participation in shopping and personal activity is quite high (and at about the same level as social activities), the time duration in these two activities among participating households is much lower (and the satiation is much higher) than in the more discretionary

[9] Note also that joint activities increase the time duration, since two individuals participating in shopping together for 20 minutes would imply 40 minutes of individual minutes in shopping activity. Thus, when allocating time and episodes across individuals in a household in a downstream model, one has to ensure that joint activities are assigned the same number of minutes of each individual participating in the joint activity. Gliebe and Koppelman, 2005 develop such an allocation model that can be used after the generation of (total individual) activity times and episodes at the household level.

asocial and recreation activities. At the same time, once a participation decision has been made, households make more episodes of personal business than shopping.

The final two columns in Table 3 indicate the split between single activity purpose participation (*i.e.*, household participation in only one activity purpose category) and multiple activity purpose participation (*i.e.*, household participation in multiple activity purpose categories) for each activity purpose. Thus, for instance, 20.4% of households who participate in shopping activity during the course of the day participate only in this activity during the weekday, while 79.6% of households who participate in shopping activity also participate in other activity purposes (note that these participations refer to the participations across all individuals in the household). In general, about four-fifths of households who participate in any activity purpose also participate in at least one additional activity purpose during the course of the day. Clearly, this indicates the variety of activity purposes in which individuals in a household participate over the course of a weekday, and reinforces the use of the multiple discrete-continuous model for modeling household-level activity participation.

5.5 Estimation Results

5.5.1 Variable Specification and Effects Interpretation

The selection of variables included in the final estimation results is based on previous research, intuitiveness, and parsimony considerations. For continuous variables (such as household income) and ordinal variables (such as number of workers), several different functional forms such as a linear specification form and a dummy variable characterization were considered. Each variable was considered in both the MDCP utility specification and in the count model threshold specification. If the coefficients of a variable in the baseline utilities of two different MDCP alternatives were not significantly different, they were combined. Also, we tested for different numbers of flexibility terms in the MC model to accommodate high or low probability masses (that cannot be explained by the underlying parameterized Poisson probabilities) for specific count outcomes. But the only such flexibility terms that turned out to be significant were for the shopping and personal business purposes, and only for the count of one. That is, since the counts are observed only conditional on positive time investment in the MDCP model, there was a need only for “one-inflation” for the shopping and personal business.

In this paper, we provide the aggregate elasticity effects of variables on the overall duration of time investment in each activity purpose as well as the number of daily episodes of each activity purpose. These two dimensions include the participation component, since, by definition, non-participation implies zero durations and zero number of episodes. We focus on aggregate elasticity effects rather than the parameter estimation results because the sign and magnitude of coefficients do not directly provide any indication of the sign and magnitude of the effects of variables on the durations and episodes. This is because of two reasons. First, the MDCP model is a non-linear utility model with satiation effects, because of which a negative sign for a variable on the baseline utility for an activity purpose (compared to a base activity activity purpose) can still result in a positive effect on duration of time investment in that activity purpose (due to an increase in the variable) if (a) the coefficient on the variable in the baseline utility of some other activity purpose is more negative and that other activity purpose has a satiation effect that is at least as high as the activity purpose under consideration and/or (b) if the coefficient on the variable in the baseline utility of some other activity purpose is less negative but that activity purpose is associated with higher satiation effects. Second, we specify a general matrix for $\mathbf{\Lambda}_1$, which is the covariance matrix of the differences in the error terms in the baseline preferences of each alternative in the MDCP model from the error term of the first alternative (but the first diagonal element of this matrix is normalized to one for identification, as discussed in Section 2.1). Such a specification generalizes other more restrictive structural specifications on the covariance matrix $\mathbf{\Lambda}$ of the original error terms of the baseline utilities. Unfortunately, though, such a general specification also implies that the estimated covariance elements of $\mathbf{\Lambda}_1$ do not provide any substantive insights (see Train, 2009; page 113 for a similar discussion in the case of traditional multinomial probit models).[10] Further, the general specification also renders the interpretation of the covariance matrix $\mathbf{\Xi}$ in the matrix $\mathbf{\Sigma_1}$ of Equation (6) difficult. The elements of $\mathbf{\Xi}$, however, influence the effects of variables on the time durations and number of episodes because they are the ones that are responsible for generating the jointness between the MDCP and MC elements in the paper. The net result is that the overall effect of a variable on

[10] We are able to use a general covariance specification because we have only four alternatives in the MDCP model. As the number of alternatives increase, there will be a need to impose *a priori* restrictive structures that seem appropriate to the application context to keep the number of parameters to be estimated in the covariance matrix to a reasonable number. However, in our estimations, *a priori* error-component type structures (for example, an error component for the more discretionary purposes of social and recreational purposes) provided statistically poorer fits, and so were discarded.

time durations and number of episodes is a complex interplay of the effects on the baseline utility of each alternative, the satiation effects associated with each alternative, as well as the estimated elements of the covariance matrix $\boldsymbol{\Sigma}_1$. Thus, there is little value in trying to interpret the model coefficients directly.[11] Indeed, the overall effects of variables are also a function of the value of the exogenous variables for each household because of the non-linear translation from the utility function to the probability expression in the MDCP model and the non-linear manner in which the variables appear in the thresholds in the MC model, which means that these effects are household-specific.

To present the effects of variables in a compact fashion, we compute aggregate elasticity effects as follows. To compute the aggregate-level "elasticity" effect of a dummy exogenous variable (such as whether the household owns a bicycle or not), we change the value of the variable to one for the subsample of observations for which the variable takes a value of zero and to zero for the subsample of observations for which the variable takes a value of one. We then sum the shifts in the expected aggregate amount of time investment (across households) in each activity purpose in the two subsamples after reversing the sign of the shifts in the second subsample, and compute the effective percentage change in the expected amount of time investment in each activity purpose due to change in the dummy variable from 0 to 1. We use the same approach to compute the effective percentage change in the expected number of episodes of each activity purpose.[12] To compute the aggregate level "elasticity" effect of a multinomial

[11] The actual parameter estimates of the MDCP and MC models, as well as the covariance matrix estimates, are available from the authors. Note that the elements of the covariance matrix $\boldsymbol{\Gamma}$ of the count error terms, however, are easily interpretable as the correlation in unobserved factors across the latent propensities y_k^* to participate in episodes of different activity types. In our estimation, the $\boldsymbol{\Gamma}$ matrix elements showed strong and statistically significant positive correlations in unobserved factors influencing the latent propensities in social and recreational activities (correlation of 0.389), and in shopping and personal activities (correlation of 0.388). However, there also were strong and positive correlations in shopping and social latent propensities (correlation of 0.325), and social and personal business latent propensities (correlation of 0.339). Less strong and less significant positive correlations were present between shopping and recreation (correlation of 0.149) and between recreation and personal business latent propensities (correlation of 0.195). Overall, these correlations highlight the need to accommodate the multivariate nature of the counts.

[12] Note that the amount of time investment and number of episodes by activity purpose for each household, needed to compute the aggregate effects just discussed, is obtained in the same way as the simulation exercise in Section 5.1. Specifically, we draw realizations of the $\mathbf{G}$ vector 200 times for each household, aggregate the predicted time investments and number of episodes across households for each of the 200 realizations, and obtain the expected value of the aggregate time investments and number of episodes as the mean across the 200 realizations. This provides the effective percentage change in the expected overall time investments and number of episodes. The standard deviation of these changes (across the 200 realizations) provides the standard errors of the percentage change estimates.

exogenous variable (such as household structure or race/ethnicity), we take the base category sub-sample and change the value of the variable from zero to one (for each specific non-base category) for all individuals in the base sub-sample. Subsequently, we compute the percentage change in the expected aggregate amount of time investment (and expected number of episodes) in each activity purpose across all households in the base sub-sample. For the aggregate level "elasticity" effect of an ordinal variable (such as number of children or number of motorized vehicles), we increase the value of the variable by 1 and compute the percentage change in the expected aggregate amount of time investment (and expected number of episodes) in each activity purpose across all households. Finally, to compute the aggregate level "elasticity" effect of a continuous variable, we increase the value of the continuous variable by 10%.[13]

5.5.2 Results and Elasticity Effects

In the empirical context studied in this paper, we estimated the MDCP-MC model for both a γ-profile and an α-profile. The γ-profile gave a better data fit than the α-profile for many different variable and error structure specifications, and therefore the γ-profile results are presented here. The translation parameter $\boldsymbol{\gamma}$ functions as both a translation parameter (allowing for zero time investments in activity purposes for some households) as well as a satiation parameter since we have fixed the value of $\boldsymbol{\alpha}$ (higher values of the $\boldsymbol{\gamma}$ parameter imply lower satiation, while lower value of the $\boldsymbol{\gamma}$ parameter imply higher satiation; see Bhat, 2008). The estimated values for the $\boldsymbol{\gamma}$ parameter values (and standard errors) are as follows: Shopping - 83.2 (4.8), Social - 644.8 (101.6), Recreation - 1000 (fixed), and Personal - 21.1 (2.6). These results indicate, consistent with the descriptive statistics in Table 3, that the lowest satiation is for the recreational activity purpose, while the highest satiation effects are for the shopping and personal activity purposes (the satiation parameter for recreation is fixed at 1000, because the parameter estimate was approaching quite large values even though the effect of the large values was rather small beyond a value of 1000; thus, for estimation stability, we fixed the parameter at the value of 1000).

[13] Technically speaking, the effect of each variable can be computed on combinations of time investments and combinations of episodes in the many activity purposes. But such combinations are too many, and so we provide information only on the marginal effects on each activity purpose individually.

In the rest of this section, we focus on the elasticity estimates associated with the variables that appeared in the final model specification. These are presented in Table 4. For instance, the entry in the first numeric row of the table under the column entitled "shopping" indicates that, on average, the daily shopping activity duration among single-person households is likely to be 4.9% less (with a standard error of 1.8%) than the shopping activity duration investment of other (primarily nuclear family) households. Other entries may be similarly interpreted.

5.5.2.1 Household Structure

Household structure effects are introduced in the specification through a series of dummy variables with "other" household structure (primarily nuclear family households) as the base category. For ease in interpretation, and because the "other" household is dominated by nuclear family households, we will assume that the "other" household structure is the nuclear family household structure in the following discussion.

As the left half of Table 4 shows, single person households, relative to single parent and nuclear family households, invest less time, in general, in shopping and social activities. Couple households, again relative to single parent and nuclear family households, have a low propensity to invest time in social activities. Both couple and single person households participate much more in recreational activities. These results are not surprising, since individuals in single-person and couple households do not have as much shopping activity responsibility as households with children. Further, individuals in single-person and couple households are also more independent and have fewer household responsibilities, leading to a higher desire and ability to participate in recreational activities (see Yamamoto and Kitamura, 1999, Pinjari *et al*., 2009, and Rajagopalan *et al.*, 2009 for similar results). The results also indicate low time investments in personal activity among single-person households, the reasons for which are not obvious. Single parent households invest less time in shopping (possibly because of tight time constraints), as well as slightly more time in social activity (perhaps a reflection of the need to be with other adults and other families with children). Indeed, several earlier studies have suggested that single parents search for outlets to socialize as a way of compensating for the unavailability of an adult partner at home (see Carpenter and DeLamater, 2012).

The effects of household structure on the number of episodes (the right half of Table 4) show that, not only are single-person and couple households less likely (than single parent and other households) to expend time in social activities and more time in recreational activities, but these tendencies also get manifested in the lower number of social activity episodes and higher number of recreational activity episodes made by these households. Interestingly, though, while couple households are likely to be spend slightly less overall time in shopping compared to nuclear family households, they participate in significantly more shopping episodes. This again may be a reflection of the need for less planning and more time flexibility among couple families, that gets manifested in a higher number of shopping episodes. The important point is that the proposed model is able to provide the differential effects of variables on overall time-use and on the number of episodes of each activity purpose, which can provide important daily pattern information for the downstream scheduling of episodes within activity-based model systems. Finally, single-parent households, on average, engage in more episodes for their personal activities, perhaps a reflection of their less flexible schedule arising from childcare duties, resulting in a squeeze of their personal activities into many separate personal care episodes.

5.5.2.2 Annual Household Income

The effect of household income reveals that low income households expend less time in shopping and personal business activities, as well as make fewer episodes for shopping and personal business activities, compared to high income households. This is consistent with the higher consumption potential of goods and services in higher income earning households (see O'Neill *et al.*, 2012 and Dai *et al.*, 2012). However, different from some earlier studies (for example, Sener and Bhat, 2012 and Pinjari *et al*., 2009), the results reveal a higher time investment in recreational activity as well as more episodes of recreational activity among low income households relative to high income households. This is interesting, and may be a result of combining active and inactive recreation pursuits under a single aggregate "recreation" category (some earlier studies such as Ferdous *et al.,* 2010 suggest that high income individuals participate more in active recreation, but less in inactive recreation). Finally, the finding that low income households pursue more social episodes is well established in the time-use literature (see Kapur and Bhat, 2007 and Parizat and Shachar, 2010), indicative of higher out-of-home

participation and variety-seeking in activities that do not necessarily impact the pocket (in terms of costs).

5.5.2.3 Household Race and Ethnicity

There is a clear pattern in time investment and number of epsiodes among Hispanic and (non-Hispanic) African American (AA) households relative to (non-Hispanic) Caucasians and other races (primarily Asian, but also Pacific islanders, mixed race, and indeterminate race). Overall, AA households invest less time in shopping and personal business activity, but pursue more episodes for these activity purposes. In terms of social activities, Hispanic and AA households spend more time in these activities, but make fewer episodes for these activities. These are again important findings, and caution against assuming that time investment decisions and episode-making decisions are always positively correlated. The higher time investment in social activities among Hispanic and AA households is consistent with similar findings from the literature (see Parks *et al*., 2003). Also, the negative coefficients on the Hispanic and African American households associated with recreational activity (for both time investments and number of episodes) reinforce the findings from earlier studies that Caucasians have higher levels of participation in recreational pursuits (see Mallett and McGuckin, 2000, Bhat and Gossen, 2004, and Humphreys and Ruseski, 2007).

5.5.2.4 Housing Type and Tenure

Households living in unattached single family homes are less inclined (relative to those living in other housing types such as condominiums, apartment complexes, and duplexes) to invest time in, and pursue episodes for, social and recreational pursuits, and more likely to invest time in shopping and personal activities. These households in single family homes also pursue more shopping episodes than those in other housing arrangements. It is quite likely that the effects above are capturing the availability of activity opportunities (in ways that are not being able to be captured through the activity accessibility measures discussed in Section 5.3); that is, single family households are more likely to be in suburban and rural areas, where there may be fewer social activity opportunities (such as restaurants) and recreational activity opportunities (such as bicycle paths, movie theatres, and workout gyms). Chen and McKnight (2007) reported a related

finding that homemakers in suburbs spend less time on discretionary activities and more time on maintenance activities.

In terms of housing tenure, households that live in rented homes (as opposed to owned homes) invest significantly less time in social activities and significantly more time in recreational activities. It is possible that recreational opportunities, such as a gym or a pedestrian pathway, or a swimming pool, are more accessible in rental communities, resulting in the higher time investment in recreational pursuits. Interestingly, however, households in rented homes also partake in significantly fewer recreational episodes, a finding that needs additional exploration in future studies.

5.5.2.5 Household Size-Related Attributes

In this group of variables, the effect of the "number of children" variable pertains to the effect of an additional child in the household beyond one (note that the presence of children effect is captured in the household structure variables). The results indicate that, as the number of children increases beyond one, households have a higher predisposition to participate in social and recreation activities rather than in shopping and personal business activities. This has also been found in Farber *et al.* (2011) and Candelaria (2010), who attribute these effects to a higher inclination to participate with young children in joint social and recreation outdoor pursuits as the number of children increase. Interestingly, and unlike some earlier studies (see, for example, Sener and Bhat, 2012 and Meloni *et al*., 2009), we did not find statistically differential effects of the number of children by age category on either time investments or the number of episodes.

As the number of workers in a household increases, so do the time investments and number of episodes in social and recreational pursuits (with decreasing time investments and number of episodes in shopping and personal business pursuits). Households with many workers are likely to be time-poor during the weekdays, and may relegate shopping and personal business to the weekend days, and channel their time mainly toward the more discretionary social and recreational pursuits during the weekdays. Lee *et al.* (2009) also observed that households with multiple workers in the household spend less weekday time on maintenance activities and more weekday time on discretionary activities.

5.5.2.6 Bicycle Ownership and Number of Motorized Vehicles

At the outset, we should acknowledge that the bicycle ownership and motorized vehicle ownership effects in the model should be viewed with some caution because we have not considered potential self-selection effects. That is, it is possible that households who want to pursue active recreation will own more bicycles, and households who would like to be mobile and pursue many episodes will own many motorized vehicles. The reader is referred to Bhat and Guo (2007), Pinjari *et al*. (2008), and De Vos *et al.* (2012) for methodologies to accommodate such self-selection effects. However, for this first demonstration application of the proposed MDCP-MC model, we ignore self-selection considerations because accommodating these will add a layer of additional econometrics over what has been proposed for the first time in this paper. So, the use of self-selection methodologies with the MDCP-MC model is left for future research.

The elasticity results of Table 4 are consistent with the notion that households that own bicycles are strongly pre-disposed to expending time in recreation pursuits and also participating in a higher number of recreation episodes, relative to households that do not own bicycles. Households who own more bicycles may be more outdoor-oriented by nature, and owning bicycles also provides an additional means to participate in outdoor recreation (Bhat, 2005, Ogilvie *et al.*, 2008). The results also indicate that the number of motorized vehicles in a household does not have a statistically significant effect on time investments, but has a clear positive and statistically significant impact on the number of episodes for social and personal activities. Overall, the positive effect of the number of vehicles on number of episodes forms the basis for using this variable as a determinant of episode generation and trip generation, but our results indicate that this effect is purpose-specific.

5.5.2.7 Accessibility Measures

The travel opportunity local accessibility measure of the length of freeways (in thousands of kilometers) accessible within 10 minutes from the residence has small, but statistically significant, positive impacts on the time investment in social and recreation activities, and weak negative impacts on the time investment in shopping and personal activities. This is perhaps because travel times and distances for social and recreational episodes are generally much longer than for other types of episodes (see Lockwood *et al.*, 2005, Carlson *et al.*, 2012), and thus the

accessibility to freeways is particularly important for social and recreation activity participation and time investments. However, this variable plays little role in the number of episodes pursued for all activity purposes, except for a small (but statistically significant) negative impact on shopping episodes.

Among the Hansen-type accessibility measures, the only one that turned out to be of importance in the final model specification was the retail and service employment accessibility. An increase in the accessibility to retail and service employment increases the time investment and the number of episodes in recreational activities, and decreases the time investment and number of episodes in other activity purposes.

Overall, though, the effects of the accessibility measures are very inelastic (note that the results in Table 4 correspond to a 10% increase in the accessibility measures). This, combined with the fact that only these two variables turned out to be statistically significant from among the many other accessibility variables considered (while several demographic variables did turn out to be important determinants) suggests that, in general, time investment in activities and the number of episodes of activities may be more of a lifestyle- and lifecycle-driven choice than related to the availability of opportunities for activity participation.[14]

5.5.3 Comparison with Independent Model

The results of the proposed joint model may be compared with the independent model that ignores the correlation between the MDCP and MC components of the model. To do so, we computed the aggregate elasticity effects as implied by the independent model. To conserve on space, we do not present an equivalent of Table 4 for the independent model, but discuss a sampling of elasticity value comparisons (the full elasticity table for the independent model is available from the authors). Note also that, since we are taking the marginals and reporting elasticity effects associated with each activity purpose, we are losing out on the richness provided by the joint model in terms of predictions of the combinations of time investments and number of episodes across all activity purposes simultaneously (for example, the number of households who participate in shopping and social, but not recreation and personal, and who

[14] However, the result that many accessibility variables are not statistically significant may also be a manifestation of the use of the TAZ as the spatial unit of resolution for computing transportation system/built environment variables. Future studies should consider more micro-scale measures to represent neighborhood physical environment variable effects, which would require some kind of geo-coded information on household residences.

make two episodes for shopping and three episodes for social activities). But, as indicated earlier, there are too many such combinations to present, and so we only present elasticity effects associated with the marginal of time investment in each activity purpose and number of episodes in each activity purpose. In such a marginal elasticity comparison exercise, the difference between the joint and independent models is due to the mis-estimated coefficients in the independent model.

According to the independent model, single person households make 0.9% fewer episodes for recreation compared to a nuclear family household, while the joint model indicates that single person households make 4.3% more recreational episodes relative to a nuclear family household. Similarly, the independent model predicts an increase of 4.8% in recreational episodes between a low income household and an observationally equivalent high income household, while the corresponding figure from the joint model is 11.9%. In terms of time investments, the independent model predicts no difference in time investment in social activities between Caucasian and AA households, while the joint model predicts an increase of 6.2% in social activity time investment between a Caucasian and an AA household. All of these indicate the differences in elasticity effects from the independent and joint models.

The substantive differences between the independent and joint models imply a need to examine the data fit of the two models. This is best done using the log-likelihood values at convergence of the two models, which are -18821.4 (for the independent model) and -18717.9 (for the joint model). The likelihood ratio test value is 207, which far exceeds the table chi-squared value with six degrees of freedom at any reasonable level of significance. The six degrees of freedom correspond to the six statistically significant covariance parameters of the 12 possible total parameters representing the covariance between the three error differentials (with respect to the shopping error term) in the MDCP model and the four purpose-specific error terms in the count model. In fact, even if one were conservative and tested the likelihood ratio test value with 12 degrees of freedom, the joint model would still resoundingly come out the winner based on the likelihood ratio test.

As a base model, we also computed the log-likelihood for the model with only the constants in the baseline preference and the satiation parameters in the MDCP model, and only the constants embedded in the $\mathbf{s}_k$ vector in the thresholds and the flexibility terms in the thresholds of the count model. This model corresponds to an independent and identically

distributed (IID) MDCP model for participation and time investment, and univariate flexible count models. The log-likelihood for this base model is -19148.2. The likelihood ratio test for testing the presence of exogenous variable effects on the baseline preference in the MDCP model, the presence of exogenous variable effects in the MC model, the presence of error covariances in the MDCP and the MC models, and the presence of error covariance between the MDCP and MC models is 860.6, which is substantially larger than the critical chi-square value with 54 degrees of freedom (corresponding to 36 non-constant parameters in the MDCP and MC models, five error covariance elements in the MDCP model, seven error covariance elements in the MC model, and six error covariance elements between the MDCP and MC models) at any reasonable level of significance. Overall, the results indicate the value of the model estimated in this paper to predict household-level activity participation, time investment, and number of episodes, based on household demographics and accessibility variables.

6. CONCLUSIONS

This paper has proposed a new econometric formulation to specify and estimate a model for multivariate count (MC) data that are themselves observed conditional on a multiple discrete-continuous (MDC) selection system. The MDC and MC systems are modeled jointly to account for any potential endogenous effects that the participation system may have on the multivariate count data in a hurdle-type model. A defining feature of the model is that decision agents jointly choose one or more discrete alternatives and determine a continuous outcome, as well as a count outcome for each chosen alternative.

A simulation exercise is undertaken to evaluate the ability of the proposed approach to recover parameters from simulated datasets generated using the proposed econometric formulation. A total of seventeen parameters, including seven error matrix components, are estimated in the simulation setup. The results from the experiments show that the proposed inference approach does well in recovering the true parameters used in the data generation. In addition, the asymptotic standard errors approximate the finite sample standard errors quite well for the typical sample sizes used in the transportation and economic literature.

This paper demonstrates the application of the proposed formulation through the study of households' decisions to participate in weekday activities, including the associated time investment as well as the frequency of episodes of each activity purpose. The data collected by

the Southern California Association of Governments for the Greater Los Angeles Area was used in the analysis. The results provide insights into the demographic and other factors that influence households' preferences for different activities, and show the importance of recognizing, from both a substantive perspective as well as a data fit perspective, the joint nature of participation, time investment, and episode frequency decisions. It is hoped that the proposed formulation will open the door for examining multivariate systems of discrete, continuous, and count data in other empirical contexts.

ACKNOWLEDGEMENTS

The authors are grateful to Lisa Macias for her help in typesetting and formatting this document.

REFERENCES

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(7), 923-939.

Bhat, C.R., 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., 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., Sidharthan, R., (2011). A simulation evaluation of the maximum approximate composite marginal likelihood (MACML) estimator for mixed multinomial probit models. *Transportation Research Part B*, 45(7), 940-953.

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

Candelaria, J.I. (2010). Physical activity of adults in households with and without children. PhD Dissertation, University of California, San Diego and San Diego State University.

Carlson, J.A., Sallis, J.F., Conway, T.L., Saelens, B.E., Frank, L.D., Kerr, J., Cain, K.L., and King, A.C. (2012). Interactions between psychosocial and built environment factors in explaining older adults' physical activity. *Preventive Medicine*, 54(1), 68-73.

Carpenter, L.M., DeLamater, J.D. (Eds) (2012). *Sex for Life: From Virginity to Viagra, How Sexuality Changes Throughout Our Lives*. New York University Press, New York.

Castro, M., Paleti, R., Bhat, C.R. (2011). 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., McKnight, C.E. (2007). Does the built environment make a difference? Additional evidence from the daily activity and travel behavior of homemakers living in New York City and suburbs. *Journal of Transport Geography*, 15(5), 380-395.

Chen, Y., Ravulaparthy, S., Deutsch, K., Dalal, P., Yoon, S.Y., Lei, T., Goulias, K.G., Pendyala, R.M., Bhat, C.R., Hu, H-H. (2011). Development of indicators of opportunity-based accessibility. *Transportation Research Record*, 2255, 58-68.

Dai, H., Masui, T., Matsuoka, Y., Fujimori, S. (2012). The impacts of China's household consumption expenditure patterns on energy demand and carbon emissions towards 2050. *Energy Policy*, 50, 736-750.

De Vos, J., Derudder, B., Van Acker, V., Witlox, F. (2012). Reducing car use: changing attitudes or relocating? The influence of residential dissonance on travel behavior. *Journal of Transport Geography*, 22, 1-9.

Farber, S., Paez, A., Mercado, R.G., Roorda, M., 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., Eluru, N., Bhat, C.R., Meloni, I. (2010). A multivariate ordered response model system for adults' weekday activity episode generation by activity purpose and social context. *Transportation Research Part B*, 44(8-9), 922-943.

Gliebe, J.P., Koppelman, F.S. (2005). Modeling household activity–travel interactions as parallel constrained choices. *Transportation*, 32(5), 449-471.

Greene, W.H. (2009). Models for count data with endogenous participation, *Empirical Economics*, 36, 133-173.

Greene, W.H., Hensher, D.A. (2010). Does scale heterogeneity across individuals matter? An empirical assessment of alternative logit models. *Transportation*, 37(3), 413-428.

Horner, M.W., O'Kelly, M.E. (2007). Is non-work travel excessive*? Journal of Transport Geography*, 15(6), 411-416.

Humphreys, B.R., Ruseski, J.E. (2007). Participation in physical activity and government spending on parks and recreation. *Contemporary Economic Policy*, 25(4), 538-552.

Joe, H. (1995). Approximations to multivariate normal rectangle probabilities based on conditional expectations. *Journal of the American Statistical Association*, 90(431), 957-964.

Kapur, A., Bhat, C.R. (2007). Modeling adults' weekend day-time use by activity purpose and accompaniment arrangement. *Transportation Research Record: Journal of the Transportation Research Board*, 2021, 18-27.

Lee, Y., Washington, S., Frank, L.D. (2009). Examination of relationships between urban form, household activities, and time allocation in the Atlanta Metropolitan Region. *Transportation Research Part A*, 43(4), 360-373.

Lockwood A., Srinivasan S., Bhat C.R. (2005). An exploratory analysis of weekend activity patterns in the San Francisco Bay area. *Transportation Research Record: Journal of the Transportation Research Board*, 1926, 70-78.

Mallett, W.J., McGuckin, N. (2000). Driving to distractions: recreational trips in private vehicles. *Transportation Research Record: Journal of the Transportation Research Board*, 1719, 267-272.

McKelvey, R.D., Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *The Journal of Mathematical Sociology* 4(1), 103–120.

Meloni, I., Portoghese, A., Bez, M., Spissu, E. (2009). Effects of physical activity on propensity for sustainable trips. *Transportation Research Record: Journal of the Transportation Research Board*, 2134, 43-50.

Ogilvie, D., Mitchell, R., Mutrie, N., Petticrew, M., Platt, S. (2008). Personal and environmental correlates of active travel and physical activity in a deprived urban population. *International Journal of Behavioral Nutrition and Physical Activity*, 5(1), 43.

O'Neill, B.C., Ren, X., Jiang, L., Dalton, M. (2012). The effect of urbanization on energy use in India and China in the iPETS model. *Energy Economics*, 34, S339-S345.

Parks, S.E., Housemann, R.A., Brownson, R.C. (2003). Differential correlates of physical activity in urban and rural adults of various socioeconomic backgrounds in the United States. *Journal of Epidemiology and Community Health*, 57(1), 29-35.

Parizat, S., Shachar, R. (2010). When Pavarotti meets Harry Potter at the Super Bowl. Available at SSRN 1711183. Chicago

Pinjari, A.R., Bhat, C.R., 2011. Computationally efficient forecasting procedures for Kuhn-Tucker consumer demand model system: Application to residential energy consumption analysis. Technical paper, Department of Civil and Environmental Engineering, The University of South Florida.

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

Rajagopalan, B.S., Pinjari, A.R., 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.

Saleh, W., Farrell, S. (2005). Implications of congestion charging for departure time choice: work and non-work schedule flexibility. *Transportation Research Part A*, 39(7), 773-791.

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

Srinivasan, S., Bhat C.R. (2005). Modeling household interactions in daily in-home and out-of-home maintenance activity participation. *Transportation,* 32(5), 523-544.

Solow, A.R. (1990). A method for approximating multivariate normal orthant probabilities. *Journal of Statistical Computation and Simulation*, 37 (3-4), 225-229.

Train, K. (2009). *Discrete Choice Methods with Simulation, Second Edition*. Cambridge University Press, Cambridge.

Yamamoto, T., Kitamura, R. (1999). An analysis of time allocation to in-home and out-of-home discretionary activities across working days and non-working days. *Transportation*, 26(2), 231-250.

LIST OF TABLES

Table 1. Simulation Results

Parameter	Parameter Estimates			Standard Error Estimates			
	True	Mean Estimate	APB	FSSE	ASE	RE	APERR
β	1.000	1.061	6.1%	0.023	0.021	0.92	0.001
γ1	1.000	0.844	15.6%	0.039	0.041	1.04	0.001
γ2	1.000	1.024	2.4%	0.053	0.051	0.97	0.001
γ3	1.000	1.115	11.5%	0.053	0.062	1.18	0.002
ς1	0.500	0.513	2.5%	0.078	0.078	0.99	0.009
ς2	0.250	0.264	5.6%	0.072	0.074	1.03	0.006
ς3	0.500	0.482	3.5%	0.082	0.067	0.82	0.006
□1	1.000	0.945	5.5%	0.064	0.064	1.00	0.008
□2	0.500	0.489	2.2%	0.042	0.040	0.95	0.003
□3	0.750	0.712	5.0%	0.043	0.041	0.94	0.003
$l_{\Sigma_1,1}$	0.600	0.552	8.1%	0.028	0.027	0.98	0.001
$l_{\Sigma_1,2}$	1.000	1.011	1.1%	0.024	0.024	1.00	0.001
$l_{\Sigma_1,3}$	0.400	0.362	9.5%	0.036	0.041	1.14	0.007
$l_{\Sigma_1,4}$	0.360	0.362	0.6%	0.038	0.039	1.01	0.007
$l_{\Sigma_1,5}$	0.475	0.449	5.3%	0.049	0.055	1.12	0.014
$l_{\Sigma_1,6}$	0.380	0.344	9.3%	0.059	0.058	0.98	0.015
$l_{\Sigma_1,7}$	0.293	0.305	4.1%	0.052	0.055	1.06	0.012
Average across all Parameters			**5.8%**	**0.049**	**0.049**	**1.01**	**0.006**

Table 2. Effects of Ignoring the Presence of the Endogenous Selection Effect

Parameter	True	Joint model		Independent Model	
		Mean Estimate	APB	Mean Estimate	APB
β	1.000	1.062	6.2%	1.060	6.0%
γ1	1.000	0.844	15.6%	0.843	15.7%
γ2	1.000	1.021	2.1%	1.025	2.5%
γ3	1.000	1.114	11.4%	1.119	11.9%
ς1	0.500	0.512	2.4%	0.498	0.5%
ς2	0.250	0.263	5.1%	0.265	6.1%
ς3	0.500	0.482	3.6%	0.484	3.2%
□1	1.000	0.945	5.5%	1.139	13.9%
□2	0.500	0.489	2.2%	0.473	5.5%
□3	0.750	0.713	4.9%	0.701	6.5%
$l_{\Sigma_1,1}$	0.600	0.552	8.1%	0.551	8.1%
$l_{\Sigma_1,2}$	1.000	1.011	1.1%	1.010	1.0%
$l_{\Sigma_1,5}$	0.475	0.446	6.0%	0.395	16.8%
$l_{\Sigma_1,6}$	0.380	0.337	11.1%	0.300	21.0%
$l_{\Sigma_1,7}$	0.293	0.309	5.5%	0.337	15.1%
Overall mean value across parameters			**6.1%**		**8.9%**
Mean log-likelihood at convergence		-10121.18		-10189.87	
Number of times the likelihood ratio test (LRT) statistic favors the Joint model		All fifty times when compared with $\chi^2_{2,0.95} = 5.99$ value (mean LRT statistic is 137)			

Table 3. Sample Characteristics

Variable	Share [%]	Variable	Share [%]
Household structure		*Housing type*	
Single-Person Household	28.2	Unattached single family home	66.1
Couple Household	29.4	Other homes (duplexes, apartment complexes, condominiums, *etc*.)	33.9
Single-Parent Household	3.1		
Other Household (primarily nuclear family households)	39.3	*Housing tenure*	
Annual Household Income		Renting	33.1
Low Income (< 50,000)	49.1	Not-renting	66.9
High Income (>50,000)	50.9	*Bicycle ownership*	
Race and Ethnicity		Own one or more bicycles	46.4
Non-Hispanic Caucasian	63.9	Not owning bicycles	53.6
Hispanic	18.1		
Non-Hispanic African-American	6.0		
Other (primarily Asian, but also including mixed race, Pacific Islander, and unidentified race)	12.0		

Descriptive Statistics

Variable	Mean	Std. Dev.	Min.	Max.
Household size-related attributes				
Number of Children (aged 15 years or younger)	0.498	0.935	0.000	6.000
Number of Adults (16 years or older)	1.931	0.862	1.000	6.000
Number of Workers	1.171	0.918	0.000	6.000
Other Household attributes				
Number of Motorized Vehicles	1.884	0.996	0.000	8.000
Length of freeways (in 1000 kms) accessible in 10 min	0.061	0.049	0.000	0.438
Retail and Service Emp. Accessibility (in 100s)	0.217	0.097	0.040	0.560

Dependent variables: Mean daily activity participation duration and mean number of daily episodes

Activity Category	Total number (%) of households participating	Mean duration of daily time investment among households who participate participation (mins)	Mean number of daily episodes among households who participate	Number of households (% of total number participating) who participate….	
				Only in activity type	In the activity type and other activity types
Shopping	1123 (53.2%)	100.0	1.34	229 (20.4%)	894 (79.6%)
Social	1175 (55.7%)	253.5	1.47	242 (20.6%)	933 (79.4%)
Recreation and Entertainment	546 (25.9%)	371.3	1.30	106 (19.4%)	440 (80.6%)
Personal	1203 (57.0%)	165.7	1.45	225 (18.7%)	978 (81.3%)

Table 4. Aggregate Elasticity Effects (and Standard Errors) of Variables

Variable	Activity duration for the activities				Mean number of episodes for the activities			
	Shopping	Social	Recreational	Personal	Shopping	Social	Recreational	Personal
Household Structure (base is other household, mainly comprised of nuclear family households)								
Single-Person Household	-4.9% (1.8%)	-3.0% (4.1%)	19.5% (8.4%)	-3.7% (2.9%)	-10.0% (2.0%)	-16.3% (3.7%)	4.3% (8.5%)	-17.6% (5.3%)
Couple Household	-0.7% (1.1%)	-5.7% (3.0%)	12.9% (4.6%)	0.4% (1.3%)	4.1% (1.7%)	-16.7% (3.4%)	2.7% (3.1%)	0.3% (1.3%)
Single Parent Household	-1.0% (0.6%)	0.9% (0.6%)	0.9% (0.8%)	-0.8% (1.1%)	-0.1% (0.3%)	0.3% (0.4%)	0.9% (0.8%)	13.3% (8.1%)
Annual Household Income (high income or income >50,000 is the base category)								
Low-Income Household	-6.0% (2.6%)	0.4% (2.9%)	18.2% (6.9%)	-5.6% (2.2%)	-4.9% (1.7%)	4.8% (3.3%)	11.9% (6.6%)	-1.7% (1.0%)
Household Race and Ethnicity (Non-Hisp. Caucasian and Other (primarily Asian, but also incl. mixed race, Pacific Islander, and unident. race) are the base)								
Hispanic Household	1.3% (1.5%)	2.3% (1.7%)	-7.1% (2.0%)	0.2% (1.2%)	-8.8% (2.8%)	-17.7% (2.9%)	-3.9% (1.6%)	-4.6% (4.2%)
African American Household	-2.4% (1.0%)	6.2% (2.3%)	-4.9% (1.8%)	-2.6% (1.5%)	0.4% (0.5%)	-13.3% (3.1%)	-1.5% (0.9%)	1.7% (1.0%)
Housing Type and Tenure [Other homes (duplexes, apartment complexes, condominiums, *etc*.) and non-renting constitute the base categories]								
Unattached Single Family House	4.2% (2.0%)	-1.6% (2.7%)	-8.2% (5.6%)	4.7% (2.0%)	6.9% (2.3%)	-2.8% (2.7%)	-9.1% (5.8%)	-1.2% (1.9%)
Renting Home	-0.6% (0.7%)	-2.8% (1.2%)	7.7% (2.8%)	-0.8% (0.8%)	-0.2% (0.6%)	1.3% (0.9%)	-4.6% (1.9%)	0.9% (1.1%)
Household Size-Related Attributes								
Number of Children	-1.0% (0.9%)	0.4% (1.0%)	1.5% (0.9%)	-0.4% (1.0%)	0.5% (0.4%)	4.1% (1.8%)	1.7% (0.9%)	-2.3% (1.2%)
Number of Workers	-4.3% (1.5%)	1.8% (1.7%)	8.0% (3.1%)	-3.3% (1.4%)	-1.4% (1.4%)	3.0% (1.9%)	9.8% (3.6%)	-4.1% (1.4%)
Bicycle Ownership and Number of Motorized Vehicles								
Owns Bicycle	-0.4% (1.1%)	-4.2% (1.7%)	9.9% (5.7%)	-0.2% (1.1%)	-1.2% (0.7%)	2.2% (3.2%)	9.5% (5.9%)	-4.3% (3.2%)
Number of Motorized Vehicles	0.6% (0.9%)	-1.0% (0.9%)	0.4% (1.0%)	0.4% (1.0%)	1.4% (1.4%)	3.1% (1.5%)	-0.5% (0.8%)	5.1% (2.5%)
Accessibility Measures								
Length of freeways (in thousands of kms) accessible in 10 min	-0.2% (0.2%)	0.2% (0.1%)	0.2% (0.1%)	-0.2% (0.1%)	-0.2% (0.1%)	0.2% (0.2%)	0.2% (0.2%)	-0.1% (0.1%)
Retail and Service Emp. Accessibility (in 100s)	-0.2% (0.1%)	-0.5% (0.2%)	1.5% (0.5%)	-0.2% (0.1%)	-0.2% (0.1%)	-0.4% (0.1%)	1.7% (0.6%)	-0.1% (0.0%)