A MULTIVARIATE MULTIPLE DISCRETE CONTINUOUS PROBIT MODEL OF TIME ALLOCATION TO COMMUTING MODES AND PHYSICAL ACTIVITY

Sebastian Astroza
The University of Texas, 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: secuastroza@gmail.com

Subodh K. Dubey
The University of Texas, 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: subbits@gmail.com

Venu M. Garikapati
Georgia Institute of Technology, School of Civil and Environmental Engineering
Mason Building, 790 Atlantic Drive, Atlanta, GA 30332-0355
Tel: 480-522-8067; Fax: 404-894-2278, Email: venu.garikapati@gatech.edu

Daehyun You
Georgia Institute of Technology, School of Civil and Environmental Engineering
Mason Building, 790 Atlantic Drive, Atlanta, GA 30332-0355
Tel: 480-626-3716; Fax: 404-894-2278, Email: daehyun.you@gatech.edu

Abdul R. Pinjari *(corresponding author)*
The University of South Florida, Department of Civil and Environmental Engineering
4202 East Fowler Avenue, Tampa, FL 33620-5350
Phone: 813-974-9671, Email: apinjari@usf.edu

Chandra R. Bhat
The University of Texas, 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

Ram M. Pendyala
Georgia Institute of Technology, School of Civil and Environmental Engineering
Mason Building, 790 Atlantic Drive, Atlanta, GA 30332-0355
Tel: 404-385-3754; Fax: 404-894-2278, Email: ram.pendyala@ce.gatech.edu

Submitted for Presentation Only

Word count: 5419 text + 4 tables/figures x 250 = 6419 words
95th Annual Meeting of the Transportation Research Board

Committee on Traveler Behavior and Values (ADB10)
August 2015

ABSTRACT

This paper presents the formulation of a multivariate multiple discrete continuous probit (MV-MDCP) choice model system. Many choice phenomena in transportation and other fields of study are multiple discrete continuous choice situations where individuals can choose multiple alternatives from a choice set. When several such dimensions of disparate types interact with one other, and are simultaneously influenced by common unobserved attributes, then a model formulation capable of jointly modeling such phenomena is needed. The MV-MDCP model system presented in this paper is capable of modeling such phenomena in a computationally tractable manner. The methodology is illustrated on a physical activity, nutrition, and health data set collected in the United Kingdom. The model estimation results demonstrate the efficacy of the model.

Keywords: discrete continuous choice model, multivariate multiple discrete continuous probit, model estimation methodology, simultaneous choice modeling, unobserved attributes

1. INTRODUCTION

This paper presents a methodological advance that allows the modeling of two multiple discrete continuous phenomena jointly. Over the past several years, there has been considerable interest in modeling multiple discrete continuous choice phenomena because many travel behavior choices are of the multiple discrete continuous nature. For example, time allocation to activities, ownership and utilization of multiple vehicle types, and monetary budget allocation to various household expense categories are all examples of multiple discrete continuous phenomena. In such choice situations, individuals are choosing multiple alternatives from among a choice set of alternatives and allocating a continuous budget across the chosen alternatives. The multiple discrete continuous extreme value (MDCEV) model, and its several variants, has made it possible to model such phenomena easily (Bhat, 2008; Bhat, 2011).

A methodological challenge that has not been adequately addressed thus far relates to the ability to model several multiple discrete continuous choice phenomena simultaneously or jointly. There may be situations where such joint modeling efforts need to be undertaken. When there are disparate multiple discrete phenomena that are based on different units of measurement, then it may be beneficial to use a multivariate MDCP modeling approach. For example, consider the case where time is allocated to various types of physically active pursuits such as sports, exercise, gardening, and recreational bicycling; and total caloric intake (nutritional diet) is allocated to various types of food groups including fruits and vegetables, meats, and grains. It is not possible to combine the two phenomena because they are measured on completely different units of measurement. And yet they are intricately related in a number of ways. Those who are health conscious individuals may allocate more time to vigorous physical activities and budget more of their caloric intake to healthy foods. Thus an unobserved attribute is affecting both phenomena, calling for the joint modeling of these behavioral dimensions. Second it is possible that one dimension affects the other; for example, a large caloric intake associated with less healthy food groups may motivate an individual to dedicate more time to exercise to compensate for such dietary intake. Alternatively, an individual who exercised for a long duration may feel that he or she is entitled to indulge in a hearty meal and desert, thus leading to relationships between the two dimensions of interest. Time spent on physical activity pursuits may influence caloric intake allocation, or caloric intake allocation may affect time allocated to physical activities. Even when the units of measurement are the same (as in the case of the example used in this paper), the use of a multivariate model may be warranted when there is a clear recursive relationship between the choice dimensions. In the event that one choice dimension has a causal and sequential relationship with another, a multivariate MDCP model would be warranted.

Despite the recognition that such joint relationships may exist, methodological limitations have made it difficult to model diverse multiple discrete continuous choice phenomena simultaneously. This paper offers a major methodological advance in the form of a multivariate multiple discrete continuous probit (MV-MDCP) model system capable of jointly modeling two or more multiple discrete continuous choice dimensions while accounting for endogeneity across the choice dimensions of interest.

The method is applied in this paper to a physical activity and nutrition data set collected in the United Kingdom. The data set includes detailed data on nutritional intake, physical activity, commuting distance and durations (by mode), socio-economic and demographic characteristics, and health variables for a sample of individuals. In studies of physical activity participation, datasets used for analysis often do not include contextual variables such as built environment variables associated with the location of residence (and/or work) of the individual respondent (e.g., the National Health and Nutrition Examination Survey in the United States). Consider time allocation to physically active pursuits such as sports and exercise of various types. At the same time, consider time allocation to commuting by different modes of transport including the more physically demanding bicycle and walk modes. These are two different multiple discrete continuous phenomena where an individual can allocate time to multiple options within each choice dimension. If no built environment variables are present, then such variables constitute unobserved factors that affect the time allocation patterns in each choice dimension. A dense mixed use environment may enhance time allocated to physically active pursuits; such an environment may also enhance the time allocated to commuting by bicycle and walk. Also, an individual who is health conscious (an unobserved personal trait) may commute by active modes more and pursue sports and exercise more as well. Thus, unobserved attributes positively contribute to both time allocation phenomena of interest. In addition, a recursive causal relationship may exist between these two choice dimensions. If an individual spends more time commuting because he or she is walking or bicycling, then he or she may feel that the act of bicycling or walking to and from work provided the exercise needed. As a result, a bike or walk commuter will spend less time for physically active recreational pursuits. These intricate relationships, and the presence of common unobserved attributes that affect disparate multiple discrete continuous phenomena, can be taken into account through the use of a multivariate multiple discrete continuous probit (MV-MDCP) model system formulated and presented in this paper.

The remainder of this paper is organized as follows. The next section presents the data used in the study. The modeling methodology is presented in Section 3 while the model estimation results are presented in Section 4. Concluding remarks appear in Section 5.

2. DATA

Data used for this modeling effort is from the United Kingdom (UK) National Diet and Nutrition Survey (NDNS), which collects nutritional data, health and energy expenditure information from a sample of individuals across the UK. The survey is carried out in all four countries of UK and a random sample of households is drawn from the Postcode Address File. In each household, one individual is selected for the data collection effort. Data is collected for four (or three, depending on the survey year) days from each selected individual. The NDNS collects information regarding diet, nutrient intake and nutritional status of the general population 1.5 years or older. Data is available for this study for four survey snapshots: 2008, 2009, 2011 and 2012. Each year, data is collected from about 1,000 people, with an equal split between adults (≥ 19 years old) and children. NDNS database is used to compute national level statistics on food consumption, additives and other food chemicals. For the four years combined, detailed data is available for about 4,100 individuals. The NDNS database contains information aggregated into the following files:

- Household Core: This file contains information regarding household composition, sex, age and marital status of all individuals in co-operating households.
- Individual Core: This file contains comprehensive data regarding the health measures of the individual, daily activity schedule of the individual and energy expenditure data.
- Day Level Dietary Data Core: This file contains information regarding the daily intake of a person's macronutrients, micronutrients and disaggregated food categories.

- Person Level Dietary Data Core: Mean intakes of food (derived from the day level dietary data) are furnished in this file.
- Food Data: A couple of additional files are available at the level of each and every food consumed by the individuals over the course of the survey period. These files are not utilized for the current analysis.

In the context of the current study, adult (≥ 19 years old) workers are selected from the individual core file for analysis and various attributes of interest are appended from the different files in the NDNS data base. Among workers, individuals who reported missing work duration/frequency were eliminated from the data set. Pertaining to the two dimensions of interest, physical activity information was readily available in the dataset. Respondents were asked to report the average duration for which they participated in a myriad of activities and the frequency of performing the activities during a four week period. Using these variables, the total duration of participation was computed and the activities were categorized into Passive, Moderate and Vigorous activities. The activity classification by level of intensity was adopted based on recommendations provided by the Center for Disease Control and Prevention (CDC, 2015). Activities under each of these categories are listed in Figure 1.

The commute duration by mode information, however, was not readily available in the NDNS data and had to be imputed from information available. The individual level data file has commute distance variable coupled with information regarding use of different modes (car, transit, walk and bike) for which the respondents answered on a Likert scale (1: "always", 2: "usually", 3: "Occssionally", 4: "Rarely"). These two pieces of information were utilized to compute commute duration by mode. First, the commute distance for the four week period for each respondent is computed as '2 x (commute distance) x (number of work trips per week) x (4 weeks)'. The Likert scale questions are then converted by assigning magnitudes or weights to each of level of the response as follows:

- Always – 50
- Usually – 35
- Occasionally – 15
- Never or Rarely – 0

If an individual responded "Usually" and "Occasionally" for "Car" and "Walk" modes respectively to commute to work in the past four weeks, the magnitudes for "Car" and "Walk" are assigned to be 35 and 15, respectively. The magnitudes for "Public Transit" and "Bike" are set to zero. Using only these magnitudes to compute allocation probabilities would lead to erroneous apportionments as there is a speed differential between the different modes. To account for this, the average travel speed by mode (Car: 29 mph, Public Trans: 24 mph, Bike: 12 mph, and Walk: 3 mph) was used to adjust the magnitudes and proportion of commute durations for each mode computed for a person. The following steps illustrate a sample computation for commute time allocation by mode:

1) Select the mode with the highest speed in the choice set of the individual. From the example, 29 is selected as the speed for the individual as the average speed of "Car" is 29 mph and that of "Walk" is 3 mph.
2) All of the speeds in the choice set are divided by the max speed to compute the "speed weight". For example, Car = 29/29 = 1 and Walk = 3/29 = 0.1034

3) Magnitudes assigned by the Likert scale questions are multiplied by "speed weight" to get the "adjusted magnitude". In the current example, adjusted weight for car is 35x1 = 35 and walk = 15x0.1034=1.551
4) The commute proportions for each individual (among the selected modes chosen) are computed using "adjusted magnitude". For example, Car = 35x100/(35+1.551) = 95.75%, Walk = 1.551x100/(35+1.551) = 4.25%. Thus 95.75% of the individual's commute duration is apportioned to car and 4.25% to walk. Thus, if total commute mileage consumption of the respondent is 2,000 miles for four weeks, 1,915.13 miles (95.75%) and 84.87 miles (4.25%) would be assigned to "Car" (answer – usually) and "Walk" (answer – occasionally), respectively.

This process helped get rid of unrealistic travel time, or distance, allocations to any mode. After commute mode duration and physical activity duration were computed for each individual, another round of data checks was performed to discard any missing/outlying data before proceeding to model estimation. Descriptive statistics of the model estimation sample are provided in Table 1. From the table, it can be seen that there is an equal split of male and female individuals in the data set. The percent of individuals living alone is in agreement with the single person household proportion seen in UK. It can also be seen that the majority of respondents are 'White', again in line with national level numbers. A look at the body mass index variable reveals that majority of the respondents are either overweight or obese, which might be related to the lower level of participation in physically active (transit, walk, bike) commuting and recreational time allocation to vigorous physical activity.

3. METHODOLOGY

3.1 Model Framework

Let there be G dependent variables of the multiple discrete-continuous (MDC) type and let g be the index for these variables $(g = 1,2,3,...,G)$. Also, let K_g be the number of alternatives corresponding to the g^{th} MDC dependent variable; and let k_g be the corresponding index. The number of alternatives K_g may vary across individuals, but we suppress the index for individuals to simplify the presentation, and assume that all the alternatives are available to all individuals.

Now, consider the g^{th} dependent variable. Following Bhat (2008), the individual is assumed to maximize his/her utility associated with this g^{th} dependent variable (and similarly for all other dependent variables) subject to a budget constraint, as below:

$$\max U_g(\boldsymbol{x}_g) = \sum_{k_g=1}^{K_g} \frac{\gamma_{gk_g}}{\alpha_{gk_g}} \psi_{gk_g} \left(\left(\frac{x_{gk_g}}{\gamma_{gk_g}} + 1 \right)^{\alpha_{gk_g}} - 1 \right)$$

$$s.t. \quad \sum_{k_g=1}^{K_g} p_{gk_g} x_{gk_g} = E_g, \qquad (1)$$

where the utility function $U_g(\boldsymbol{x}_g)$ is quasi-concave, increasing and continuously differentiable; $\boldsymbol{x}_g$ is the consumption quantity vector of dimension $(K_g \times 1)$ with elements x_{gk_g} such that $x_{gk_g} > 0\ \forall k_g$; and γ_{gk_g} , α_{gk_g} , and ψ_{gk_g} are parameters associated with alternative k_g . In the budget constraint, E_g is the total expenditure for the g^{th} dependent variable, and p_{gk_g} is the unit price of consumption for alternative k_g . Assume, for now, that there is no essential outside alternative (i.e., an alternative that is always consumed), so that corner solutions (*i.e.*, zero consumptions) are possible for all choice alternatives (relaxing this assumption is straightforward). ψ_{gk_g} represents the baseline marginal utility for alternative k_g (i.e., marginal utility of alternative k_g at the point of no consumption of it). γ_{gk_g} allows corner solutions for alternative k_g and also serves as a translation-based satiation parameter, while α_{gk_g} serves as an exponential-based satiation parameter. Only one parameter of the set γ_{gk_g} and α_{gk_g} will be empirically identified, so the analyst will have to estimate either a γ-profile (in which $\alpha_{gk_g} \to 0$) or an α-profile (in which the γ_{gk_g} terms are normalized to 1). Also, for the γ-profile, one need to ensure $\gamma_{gk_g} > 0\ \forall k_g$, and, for the α-profile, the condition is: $\alpha_{gk_g} \le 1\ \forall k_g$. In the current paper, we will retain the general utility form of Equation (1) to keep the presentation general.

Next, introduce stochasticity through the baseline marginal utility function ψ_{gk_g} , as:

$$\psi_{gk_g} = \exp(\boldsymbol{\beta}'_g \boldsymbol{z}_{gk_g} + \xi_{gk_g}), \tag{2}$$

where $\boldsymbol{z}_{gk_g}$ is a A_g-dimensional vector of attributes that characterizes alternative k_g (including a constant for each alternative except one, to capture intrinsic preferences for each alternative relative to a base alternative); $\boldsymbol{\beta}_g$ is a consumer-specific vector of coefficients (of dimension $A_g \times 1$) and ξ_{gk_g} captures the idiosyncratic (unobserved) characteristics that impact the baseline utility of alternative k_g . We assume that the error term $\boldsymbol{\xi}_g [= (\xi_{g1}, \xi_{g2}, ..., \xi_{gK_g})']$ is distributed multivariate normal. That is, $\boldsymbol{\xi}_g \sim \text{MVN}_{K_g}(\boldsymbol{0}_{K_g}, \boldsymbol{\Lambda}_g)$, where $\text{MVN}_K(\boldsymbol{0}_K, \boldsymbol{\Lambda})$ indicates a *K*-variate normal distribution with a mean vector of zeros denoted by $\boldsymbol{0}_K$ and a covariance matrix $\boldsymbol{\Lambda}$. Further, to incorporate taste heterogeneity, we consider $\boldsymbol{\beta}_g$ to be a realization from a multivariate normal distribution: $\boldsymbol{\beta}_g \sim f_{A_g}(\boldsymbol{b}_g, \boldsymbol{\Omega}_g)$. For future reference, we also write $\boldsymbol{\beta}_g = \boldsymbol{b}_g + \tilde{\boldsymbol{\beta}}_g$, where $\tilde{\boldsymbol{\beta}}_g \sim f_{A_g}(\boldsymbol{0}_{A_g}, \boldsymbol{\Omega}_{A_g})$.

The optimal consumption vector $\boldsymbol{x}_g$ can be solved based on the constrained optimization problem of Equation (1) by forming the Lagrangian function and applying the KKT conditions. The Lagrangian function for the problem is:

$$\ell_g = \sum_{k_g=1}^{K_g} \frac{\gamma_{gk_g}}{\alpha_{gk_g}} \exp(\boldsymbol{b}'_g \boldsymbol{z}_{gk_g} + \tilde{\boldsymbol{\beta}}'_g \boldsymbol{z}_{gk_g} + \xi_{gk_g}) \left\{ \left(\frac{x_{gk_g}}{\gamma_{gk_g}} + 1 \right)^{\alpha_{gk_g}} - 1 \right\} - \lambda_g \left[\sum_{k_g=1}^{K_g} p_{k_g} x_{gk_g} - E_g \right], \tag{3}$$

where λ_g is a Lagrangian multiplier associated with the expenditure constraint of the g^{th} dependent variable. The KKT first-order conditions for the optimal consumptions $x^*_{gk_g}$ are:

$$\begin{aligned}
&\exp(\boldsymbol{b}'_g \boldsymbol{z}_{gk_g} + \tilde{\boldsymbol{\beta}}'_g \boldsymbol{z}_{gk_g} + \xi_{gk_g})\left(\frac{x^*_{gk_g}}{\gamma_{gk_g}}+1\right)^{\alpha_{gk_g}-1} - \lambda_g\, p_{gk_g} = 0 \text{, if } x^*_{gk_g} > 0 \text{, } k_g = 1,2,...,K_g \\
&\exp(\boldsymbol{b}'_g \boldsymbol{z}_{gk_g} + \tilde{\boldsymbol{\beta}}'_g \boldsymbol{z}_{gk_g} + \xi_{gk_g})\left(\frac{x^*_{gk_g}}{\gamma_{gk_g}}+1\right)^{\alpha_{gk_g}-1} - \lambda_g\, p_{gk_g} < 0 \text{, if } x^*_{gk_g} = 0 \text{, } k_g = 1,2,...,K_g .
\end{aligned} \tag{4}$$

The optimal demand satisfies the above conditions and the budget constraint $\sum_{k_g=1}^{K_g} p_{gk_g} x^*_{gk_g} = E_g$. The budget constraint implies that only $(K_g - 1)$ of the $x^*_{gk_g}$ values need to be estimated. To accommodate this singularity, let m_g be, without loss of generality, the consumed alternative with the lowest value of k_g (note that the consumer must consume at least one alternative given $E_g > 0$). For this m_g^{th} alternative, $x^*_{gm_g} > 0$, which, in conjunction with the first set of KKT conditions in Equation (4), implies the following expression for λ_g:

$$\lambda_g = \frac{\exp(\boldsymbol{b}'_g \boldsymbol{z}_{gm_g} + \tilde{\boldsymbol{\beta}}'_g \boldsymbol{z}_{gm_g} + \xi_{gk_g})}{p_{gm_g}}\left(\frac{x^*_{gm_g}}{\gamma_{gm_g}}+1\right)^{\alpha_{gm_g}-1}. \tag{5}$$

Substituting λ_g back in Equation (4) for the other alternatives k_g ($k_g = 1,2,...,K_g ; k_g \neq m_g$), and taking logarithm simplifies the KKT conditions as follows:

$$\begin{aligned}
&y^*_{gk_g m_g} = 0 \text{, if } x^*_{gk_g} > 0 \text{, } k_g = 1,2,...,K \text{, } k_g \neq m_g \\
&y^*_{gk_g m_g} < 0 \text{, if } x^*_{gk_g} = 0 \text{, } k_g = 1,2,...,K_g \text{, } k_g \neq m_g .
\end{aligned} \tag{6}$$

where, $y^*_{gk_g m_g} = y_{gk_g} - y_{gm_g}$; $y_{gk_g} = V_{gk_g} + \tilde{\boldsymbol{\beta}}'_g \boldsymbol{z}_{gk_g} + \xi_{gk_g}$;and

$V_{gk_g} = \boldsymbol{b}'_g \boldsymbol{z}_{gk_g} + (\alpha_{gk_g} - 1)\ln\left(\frac{x^*_{gk_g}}{\gamma_{gk_g}}+1\right) - \ln p_{gk_g}$.

Two important identification issues need to be noted here. <u>First</u>, a constant cannot be identified in the $\boldsymbol{b}'_g \boldsymbol{z}_{gk_g}$ term for one of the alternatives. Similarly, consumer-specific variables that do not vary across alternatives can be introduced for $(K_g - 1)$ alternatives, with the remaining alternative being the base. <u>Second</u>, only the covariance matrix of the error differences is estimable. Taking the difference with respect to the first alternative, only the elements of the covariance matrix $\breve{\boldsymbol{\Lambda}}_g$ of $\varepsilon_{gk_g 1} = \xi_{gk_g} - \xi_{g1}$, $k_g \neq 1$ are estimable. However, the KKT conditions take the difference against the first consumed alternative m_g. Thus, in translating the KKT conditions to the consumption probability, the covariance matrix of the error differences with respect to m_g is desired. Since m_g will vary across consumers, this covariance matrix will also vary across consumers. But all the covariance matrices must originate from the same covariance matrix $\boldsymbol{\Lambda}_g$ for the original error term vector $\boldsymbol{\xi}_g$. To achieve this consistency, the error covariance matrix is constructed in a specific way that will be explained later in this section.

Now, the jointness in the unobserved portion of the utility of different MDC variables may be generated as follows: define $\boldsymbol{\varepsilon}_g = (\varepsilon_{g21}, \varepsilon_{g31}, ..., \varepsilon_{gK_g1})'$ $\lfloor (K_g - 1) \times 1 \rfloor$ and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1', \boldsymbol{\varepsilon}_2', ..., \boldsymbol{\varepsilon}_G')'$ of size $\left[\sum_{g=1}^{G}\left(K_g - 1\right)\right] \times 1$. Then the distribution of the vector $\boldsymbol{\varepsilon}$ can be written as:

$$\breve{\boldsymbol{\Lambda}} = \begin{bmatrix} \breve{\boldsymbol{\Lambda}}_1 & \breve{\boldsymbol{\Lambda}}_{12} & \cdots & \breve{\boldsymbol{\Lambda}}_{1G} \\ \breve{\boldsymbol{\Lambda}}_{12}' & \breve{\boldsymbol{\Lambda}}_2 & \cdots & \breve{\boldsymbol{\Lambda}}_{2G} \\ \vdots & \vdots & \ddots & \vdots \\ \breve{\boldsymbol{\Lambda}}_{1G}' & \breve{\boldsymbol{\Lambda}}_{2G}' & \cdots & \breve{\boldsymbol{\Lambda}}_G \end{bmatrix} \tag{7}$$

where $\breve{\boldsymbol{\Lambda}}_g$, as mentioned earlier, captures the covariance between error differences (with respect to the first alternative) of the g^{th} MDC variable and $\breve{\boldsymbol{\Lambda}}_{gg'}$ captures the dependence between the error differences (with respect to the first alternative) of g and g' dependent variables. Further, if there is no price variation across alternatives for each consumer (*i.e.*, if $p_{gk_g} = \tilde{p}_g \; \forall k_g$), an additional scale normalization needs to be imposed on $\breve{\boldsymbol{\Lambda}}_{\mathbf{g}}$ (Bhat, 2008). For instance, one can normalize the element of $\breve{\boldsymbol{\Lambda}}_g$ in the first row and first column to the value of one. But, if there is some price variation across alternatives for even a subset of consumers, there is no need for this scale normalization and all the $K_g(K_g - 1)/2$ parameters of the matrix $\breve{\boldsymbol{\Lambda}}_g$ are estimable. In the general case, this allows the estimation of $\left(\dfrac{K_g * (K_g - 1)}{2} - 1\right)$ terms embedded in each $\breve{\boldsymbol{\Lambda}}_g$ matrix and the $\sum_{g=1}^{G-1}\sum_{l=g+1}^{G}(K_g - 1) \times (K_l - 1)$ covariance terms in the off-diagonal matrices of the matrix $\breve{\boldsymbol{\Lambda}}$ characterizing the dependence between the latent utility differentials (with respect to the first alternative) across the MDC variables. Note that the matrix $\breve{\boldsymbol{\Lambda}}$ represents the covariance of error difference taken with respect to the first alternative for each of the dependent variables. For estimation, the corresponding error differentials with respect to the m_g^{th} alternative (i.e., a chosen alternative) for each MDC variable, say $\overset{\leftrightarrow}{\boldsymbol{\Lambda}}$, is needed. For this purpose, a general covariance matrix $\boldsymbol{\Lambda}$ needs to be created depending on the value of m_g for all MDC variables. To do so, define a matrix $\mathbf{D}$ of size $\left[\sum_{g=1}^{G} K_g\right] \times \left[\sum_{g=1}^{G} K_g - 1\right]$ whose elements are zero. Then insert an identity matrix of size $(K_g - 1)$ for the every g^{th} MDC variable in the rows $\left(\sum_{i=1}^{g-1} K_i\right) + 2$ to $\left(\sum_{i=1}^{g} K_i\right)$ and columns $\left(\sum_{i=1}^{g-1}(K_i-1)\right)+1$ to $\left(\sum_{i=1}^{g}(K_i-1)\right)$ where $\left(\sum_{i=1}^{0}K_i\right)=0$. Then, the covariance matrix for the original error terms may be developed as $\mathbf{\Lambda}=\mathbf{D}\breve{\mathbf{\Lambda}}\mathbf{D}'$. All the parameters in this matrix are identifiable by virtue of the way this matrix is constructed based on error differences and, at the same time, it provides a consistent means to obtain the covariance matrix $\overleftrightarrow{\mathbf{\Lambda}}$ that is needed for estimation.

3.2 Model Estimation

Let $\boldsymbol{\lambda}$ be the vector of all parameters to be estimated for all the dependent variables under consideration. To develop the likelihood function, define the following vector and matrices: $\boldsymbol{y}_g=(y_{g1},y_{g2},...,y_{gK_g})'\ \left[K_g\times 1\right]$, $\boldsymbol{y}=(\boldsymbol{y}_1',\boldsymbol{y}_2',...,\boldsymbol{y}_G')'\ \left[\left(\sum_{i=1}^{G}K_i\right)\times 1\right]$, $\boldsymbol{V}_g=(V_{g1},V_{g2},...,V_{gK_g})'\ \left[K_g\times 1\right]$

$,\boldsymbol{V}=(\boldsymbol{V}_1',\boldsymbol{V}_2',...,\boldsymbol{V}_G')'\ \left[\left(\sum_{i=1}^{G}K_i\right)\times 1\right]$, $\boldsymbol{\xi}_g=(\xi_{g1},\xi_{g2},...,\xi_{gK_g})'\ \left[K_g\times 1\right]$,

$\boldsymbol{\xi}=(\boldsymbol{\xi}_1',\boldsymbol{\xi}_2',...,\boldsymbol{\xi}_{G_g}')'\ \left[\left(\sum_{i=1}^{G}K_i\right)\times 1\right]$, $\tilde{\boldsymbol{\beta}}=(\tilde{\boldsymbol{\beta}}_1',\tilde{\boldsymbol{\beta}}_2',...,\tilde{\boldsymbol{\beta}}_G')'\ \left[\left(\sum_{i=1}^{G}K_i\right)\times\left(\sum_{i=1}^{G}A_i\right)\text{matrix}\right]$,

$\mathbf{z}_g=(\mathbf{z}_{g1},\mathbf{z}_{g2},...,\mathbf{z}_{gK_g})'\ \left[K_g\times A_g\text{ matrix}\right]$ and $\mathbf{z}=(\mathbf{z}_1',\mathbf{z}_2',...,\mathbf{z}_G')'\ \left[\left(\sum_{i=1}^{G}K_i\right)\times\left(\sum_{i=1}^{G}A_i\right)\text{matrix}\right]$.

Then, we may write, in matrix notation, $\boldsymbol{y}=\boldsymbol{V}+\tilde{\boldsymbol{\beta}}'\mathbf{z}+\boldsymbol{\xi}$. It is easy to see that vector $\boldsymbol{y}$ is distributed multivariate normal with mean $\boldsymbol{V}$ and covariance $(\widehat{\mathbf{\Omega}}+\mathbf{\Lambda})$ where

$$\widehat{\mathbf{\Omega}}=\begin{bmatrix}\widehat{\mathbf{\Omega}}_1 & \mathbf{0}_{K_1K_2} & \cdots & \mathbf{0}_{K_1K_G}\\ \mathbf{0}_{K_2K_1} & \widehat{\mathbf{\Omega}}_2 & \cdots & \mathbf{0}_{K_2K_G}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0}_{K_GK_1} & \mathbf{0}_{K_GK_2} & \cdots & \widehat{\mathbf{\Omega}}_G\end{bmatrix}\quad\left[\sum_{g=1}^{G}K_g\right]\times\left[\sum_{g=1}^{G}K_g\right]\text{matrix}\qquad(8)$$

where, $\widehat{\mathbf{\Omega}}_g=\mathbf{z}_g\mathbf{\Omega}_g\mathbf{z}_g'$ and $\mathbf{0}_{K_GK_{G'}}$ represents a $\left[G\times G'\right]$ matrix with all its elements being zero.

For model estimation, we need to derive the distribution of the vector $\boldsymbol{y}$. To do so, define a matrix $\mathbf{M}$ of size $\left(\sum_{g=1}^{G}(K_g-1)\right)\times\left(\sum_{g=1}^{G}K_g\right)$ whose elements are all zero. Then insert an identity matrix of size (K_g-1) after supplementing with a column of '-1' values in the column corresponding to the value of m_g in the rows $\left(\sum_{i=1}^{g-1}(K_i-1)\right)+1$ to $\left(\sum_{i=1}^{g}(K_i-1)\right)$ and columns $\left(\sum_{i=1}^{g-1}K_i\right)+1$ to $\left(\sum_{i=1}^{g}K_i\right)$ where $\left(\sum_{i=1}^{0}K_i\right)=0$.

Then $\boldsymbol{y}^* = \mathbf{M}\boldsymbol{y} \sim MVN_{\sum_{g=1}^{G}(K_g-1)}(\boldsymbol{H},\boldsymbol{\Psi})$, where $\boldsymbol{H} = \mathbf{M}\boldsymbol{V}$ and $\boldsymbol{\Psi} = \mathbf{M}\left(\hat{\boldsymbol{\Omega}} + \boldsymbol{\Lambda}\right)\mathbf{M}'$. Also, define $L_{g,NC}$ as the number of non-consumed alternatives for the g^{th} MDC variable ($0 \le L_{g,NC} \le K_g - 1$), and $L_{g,C}$ as the number of consumed alternatives (other than the m_g^{th} alternative) for the g^{th} MDC variable ($0 \le L_{g,C} \le K_g - 1$). Next, partition the vector $\boldsymbol{y}^*$ into a sub-vector $\tilde{\boldsymbol{y}}_{NC}^*$ of length $\left(\sum_g L_{g,NC}\right) \times 1$ of all the non-consumed alternatives (across all the G dependent variables), and another sub-vector $\tilde{\boldsymbol{y}}_C^*$ of length $\left(\sum_g L_{g,C}\right) \times 1$ of all the consumed alternatives (across all the G dependent variables, except the alternative m_g for each dependent variable). Let $\tilde{\boldsymbol{y}}^* = \left(\left[\boldsymbol{y}_{NC}^*\right]', \left[\boldsymbol{y}_C^*\right]'\right)'$, which may be obtained as $\tilde{\boldsymbol{y}}^* = \mathbf{R}\boldsymbol{y}^*$, where $\mathbf{R}$ is a re-arrangement matrix of dimension $\left[\sum_{g=1}^{G}(K_g - 1)\right] \times \left[\sum_{g=1}^{G}(K_g - 1)\right]$ with zeros and ones: $\mathbf{R} = \begin{bmatrix} \mathbf{R}_{NC} \\ \mathbf{R}_C \end{bmatrix}$. $\mathbf{R}_{NC}$ itself is a matrix of as many rows as $\sum_g L_{g,NC}$ and as many columns as $\sum_{g=1}^{G}(K_g - 1)$. Each column corresponds to an alternative (except the m_g^{th} alternative for each dependent variable *g*). Then, for every row, $\mathbf{R}_{NC}$ has a value of one over of the columns corresponding to an alternative that is not consumed (across all the *g* dependent variables), and a value of zero in every other column. A similar construction is involved in creating the matrix $\mathbf{R}_C$, which is of dimension $\left[\sum_{g=1}^{G} L_{g,C}\right] \times \left[\sum_{g=1}^{G}(K_g - 1)\right]$.

Using the above rearrangement matrices, one may partition $\boldsymbol{y}^*$ as $\boldsymbol{y}_{NC}^* = \mathbf{R}_{NC}\boldsymbol{y}^*$ and $\boldsymbol{y}_C^* = \mathbf{R}_C\boldsymbol{y}^*$. Consistent with this partitioning, define $\tilde{\boldsymbol{H}} = \mathbf{R}\boldsymbol{H}$, $\tilde{\boldsymbol{H}}_{NC} = \mathbf{R}_{NC}\boldsymbol{H}$, $\tilde{\boldsymbol{H}}_C = \mathbf{R}_C\boldsymbol{H}$, and $\tilde{\boldsymbol{\Psi}} = \mathbf{R}\boldsymbol{\Psi}\mathbf{R}' = \begin{bmatrix} \tilde{\boldsymbol{\Psi}}_{NC} & \tilde{\boldsymbol{\Psi}}'_{NC,C} \\ \tilde{\boldsymbol{\Psi}}_{C,NC} & \tilde{\boldsymbol{\Psi}}_C \end{bmatrix}$, where $\tilde{\boldsymbol{\Psi}}_{NC} = \mathbf{R}_{NC}\boldsymbol{\Psi}\mathbf{R}'_{NC}$, $\tilde{\boldsymbol{\Psi}}_C = \mathbf{R}_C\boldsymbol{\Psi}\mathbf{R}'_C$, and $\tilde{\boldsymbol{\Psi}}_{NC,C} = \mathbf{R}_{NC}\boldsymbol{\Psi}\mathbf{R}'_C$. Define $\boldsymbol{x}_g^* = (x_{g1}^*, x_{g2}^*, \ldots, x_{gK_g}^*)'$ and $\boldsymbol{x}^* = (\boldsymbol{x}_1^{*\prime}, \boldsymbol{x}_2^{*\prime}, \ldots, \boldsymbol{x}_G^{*\prime})'$. Then, the likelihood function corresponding to the consumption quantity vector $\boldsymbol{x}^*$ may be written as:

$$L(\boldsymbol{x}^*) = \det(\mathbf{J}) \int_{\boldsymbol{h}_{NC}=-\infty}^{\boldsymbol{0}} f_{\sum_{g=1}^{G}(K_g-1)}\left(\boldsymbol{h}_{NC}, \boldsymbol{0}_{L_C} \mid \tilde{\boldsymbol{H}}, \tilde{\boldsymbol{\Psi}}\right) d\boldsymbol{h}_{NC}, \qquad (9)$$

where $\mathbf{J}$ is the block diagonal Jacobian matrix (of dimension $(L_C+G)\times(L_C+G)$, with $L_C=\sum_{g=1}^{G}L_{g,C}$), with each block matrix (of size $(L_{g,C}+1)\times(L_{g,C}+1)$) corresponding to a specific dependent variable g. Due to the block diagonal nature of $\mathbf{J}$, and using Bhat's (2008) derivation,

$$\det(\mathbf{J})=\prod_{g=1}^{G}\left[\left\{\prod_{k_g\in C_g}\frac{1-\alpha_{gk_g}}{x^*_{gk_g}+\gamma_{gk_g}}\right\}\left\{\sum_{k_g\in C_g}\left(\frac{x^*_{gk_g}+\gamma_{gk_g}}{1-\alpha_{gk_g}}\right)\left(\frac{p_{gk_g}}{p_{gm_g}}\right)\right\}\right], \tag{10}$$

where C_g is the set of all alternatives consumed in dimensions g (including alternative m_g).

The integral in the likelihood function in Equation (10) may be rewritten as a product of a marginal and conditional multivariate normal (MVN) distribution, where the marginal distribution is an MVN probability density function (which has a closed form expression) and the conditional distribution is an MVN cumulative density (MVNCD) function (see Bhat et al., 2013 for details). To evaluate the MVNCD function for estimating the parameters embedded in the likelihood function in Equation (10), we use the maximum approximated composite marginal likelihood (MACML) approach proposed by (Bhat, 2011).

Finally, the presence of outside good in one of the MDC variables has no impact on the overall methodology apart from the fact that the utility expression for alternatives in presence of an outside good changes slightly. We do not discuss the details of MDC construction for the outside good and the reader is referred to Bhat (2008) for a detailed discussion.

4. ESTIMATION RESULTS

Model estimation results are presented in Table 2. For the physically active recreational time allocation model component, the base alternative is passive activities (the outside good in which everybody participates). For the commute mode time allocation model component, the base alternative is car. The model coefficients indicate the extent to which different explanatory variables contribute positively or negatively to time allocation in a specific category.

Lower income individuals are less likely to pursue moderate and vigorous activities, with the lowest income group least prone to pursue such activities. It is possible that this group cannot afford to undertake such activities or do not reside in neighborhoods with good amenities to pursue such activities. The natural logarithm of age is associated with a higher level of moderate activities, suggesting that individuals beyond a certain advanced age are more likely to “slow down” and increase pursuit of moderate activities – they wish to stay active, and yet cannot participate in vigorous activities. Males are more inclined than females to allocate time to moderate and vigorous activities, suggesting that there is a gender difference in recreational pursuits and time availability to do so. The body mass index indicating obesity is negatively associated with the time allocation to vigorous activities. Although the direction of causality is not clear, it is apparent that there is an indirect relationship between BMI and vigorous activity time allocation. Those who consume at least five portions of fruits and vegetables and consumed fish during the month are likely to undertake moderate and vigorous activities more than others. These individuals are likely to be health conscious and eat healthy foods; their health consciousness also manifests itself in the form of increased moderate and vigorous activities.

There is a significant endogenous effect involving bicycle commuting duration. Those who bicycle to work also tend to allocate more time to moderate and vigorous activities, pointing to a strong positive and symbiotic relationship between the use of an active commuting mode and the pursuit of moderate and vigorous physical activities. This relationship may be recursive in nature. An individual who chooses to allocate more time to bicycle commuting may deliberately choose to undertake less vigorous activities (relative to moderate activities) because he or she may believe that the exercise obtained through commuting by bicycle provides the necessary vigorous activity. Thus, the choice of commute time allocation influenced time allocation to different types of physically active recreational pursuits. This type of recursive relationship can be modeled using the multivariate MDCP. Another way to view the recursive relationship is through the impacts of commute distance. Longer commute distances are associated with less time allocation to bicycling and walking modes. The reduction in bicycle commuting time will in turn contribute to a lowering of moderate physical activity relative to passive and vigorous physical activities.

The commute time allocation model also shows behaviorally intuitive results. Lower income individuals are more likely to bicycle or walk to work with the tendency more pronounced for the lowest income bracket. Limited car ownership in these market segments likely contributes to these findings. Individuals older than 55 years of age are less prone to walk to work; as individuals get older, they may be less inclined to walk to work due to the physical requirements of doing so. Males are more inclined to allocate time to bicycling than females. It appears that males are more willing to take on physically strenuous activities, although it may be possible that females continue to shoulder greater household responsibilities and chauffeuring duties, leaving less time and making it more inconvenient to commute by bicycle. Non-whites are more likely than other races to allocate time to transit. Finally, those who consumed fish were more likely to walk, once again suggesting that these individuals are health conscious and hence depict these behaviors.

Although the model estimation results are quite intuitive, the findings should not be construed as providing authoritative information on the factors affecting time allocation to physical activities and non-motorized commuting modes. Despite the richness of the data on the health and nutrition aspects, the data does not have many variables that would serve as explanatory variables in the model specification. The data set includes only a limited set of socio-economic and demographic variables, no built environment and contextual variables, and very limited mobility and transportation related data. As such, it was not possible to estimate a MMDCP model system with a rich specification that includes many socio-economic variables. The goodness of fit of the model is very low and it is clear that there are many other attributes that affect this behavior. Although the goodness of fit measures suggest that the MMDCP model system is statistically significantly better in fitting the data when compared with an independent MDCP model, the degree of improvement is not very large. In the absence of a rich specification, it is not possible to draw conclusive inferences regarding commuting mode use and physical activity engagement and the relationships that exist between and govern these dimensions. This paper is not intended to serve as a confirmatory empirical study of physical activity and commuting mode use; rather the empirical application is merely serving to illustrate and demonstrate the capabilities of the methodological advances presented in this paper. The significant error difference covariances seen in Table 3 also indicate the merits of a multivariate MDCP over an independent MDCP model that ignores such correlations (leading to biased parameter estimates).

5. CONCLUSIONS

There is considerable interest in understanding the interrelationships between disparate multiple discrete continuous choice phenomena. This is because of the role played by transportation in affecting choice dimensions in various aspects of life. Whether it be monetary expenditures, time expenditures, or mileage consumptions, transportation is filled with examples of multiple discrete continuous phenomena where a decision maker can choose multiple alternatives from a choice set. The model formulation is motivated by the fact that disparate multiple discrete continuous variables may be measured along different scales or units, and thus it is impossible to combine such variables into a single multiple discrete continuous choice variable. The multivariate model system can also account for endogeneity effects where correlated unobserved attributes affect multiple choice dimensions simultaneously. Finally, the model formulation may be of value even when the multiple choice dimensions are measured in the same units. If it is posited that there is a clear sequential, recursive, and causal relationship between two dimensions of interest, then the multivariate multiple discrete continuous model formulation can effectively capture such cause-and-effect relationships. If the dimensions were all combined into a single multiple discrete continuous model, then patterns of substitution may be seen in a correlational framework without any insights into the cause-and-effect relationships that drive the phenomena of interest.

This paper demonstrates the capabilities of the methodological advances by presenting a joint analysis of individuals' monthly time allocation to commuting via different modes of travel and monthly time allocation to recreational activities. The motivation for jointly analyzing such seemingly disparate aspects of individuals' travel and activity patterns stems from the recognition that both commuting and recreation may involve physical activity – physically active commuting (walking and bicycling) and physically active recreational activities, respectively. While the former is a utilitarian travel, the latter involves recreational activities and non-utilitarian travel, thereby warranting the need to analyze commuting and recreation as a joint bundle. In addition, the model is based on a monthly time allocation (as opposed to analysis over limited timeframes such as a day) using a dataset that collected information on individuals' monthly time allocation to a number of physically recreational activities and monthly commuting mode choice. The dataset also includes a number of rich health-related variables (e.g., body mass index) and nutritional variables (e.g., consumptions of different types of foods). For the joint analysis, a novel multivariate multiple discrete continuous choice model (labeled the MV-MDCP model) that explicitly recognizes that monthly commuting may involve travel by (and time allocation to) multiple modes of travel and that monthly recreational activities may involve time allocation to potentially multiple types of physical activities. In addition, the model recognizes the interrelationships between the two MDC variables via an endogenous influence of commute time on physical activities as well as correlations due to common unobserved factors influencing them.

Model estimation results suggest the role of demographics, body mass index, and nutritional habits on both commute time allocation and physically active recreational activity. Interesting findings include: (a) the positive association of healthy nutritional habits with both physically active recreation and physically active commuting, (b) the deterrence of longer commute distance on physically active participation, (c) the negative association between overweight (obese) and physical activity participation, and (d) the substitutive influence of physically active utilitarian travel on physical activity participation. Equally important are the findings relevant to the presence of common unobserved effects influencing both monthly commute mode choice and time allocation and monthly recreational activity participation. These results highlight the need for a multivariate model system that also recognizes that each variable is multiple discrete-continuous in nature.

REFERENCES

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., M. Castro, and M. Khan (2013), A New Estimation Approach for the Multiple Discrete-Continuous Probit (MDCP) Choice Model, *Transportation Research Part B*, Vol. 55, pp. 1-22

Center for Disease Control and Prevention. (2015). General Physical Activities Defined by Level of Intensity. Available at: http://www.cdc.gov/nccdphp/dnpa/physical/pdf/PA_Intensity_table_2_1.pdf, Accessed July 31, 2015.

Passive

- Fishing or hunting
- Snooker, billiards, or darts
- Musical instrument playing or singing
- TV, DVD or video viewing
- Computer use at home

Moderate

- Swimming leisurely
- Walking for pleasure
- Cycling for pleasure
- Mowing
- Watering lawn or garden
- Weeding, pruning
- DIY
- Other impact aerobics
- Exercise with weights
- Conditioning exercises (bike/rowing machine)
- Floor exercises (Stretching/yoga)
- Dancing
- Bowling
- Table tennis
- Golf
- Cricket
- Rowing
- Netball, volleyball, basketball
- Horse-riding
- Ice skating
- Sailing, wind-surfing, or boating

Vigorous

- Swimming - competitive
- Backpacking or mountain climbing
- Racing or rough terrain cycling
- Heavy gardening (Digging, shoveling, chopping wood)
- High impact aerobics
- Competitive running
- Jog
- Tennis
- Squash
- Badminton
- Football, rugby, hockey
- Martial arts, boxing, or wrestling

The categorization is adopted from CDC website (http://www.cdc.gov/nccdphp/dnpa/physical/pdf/PA_Intensity_table_2_1.pdf)

FIGURE 1 Aggregation of Physical Activities into Three Groups

TABLE 1 Sample Characteristics (N=590)

Demographic, health and nutrition variables	Categories	Proportion
Household annual income	£0-£24,999 £25,000-£34,999 >£35,000-£79,999	28.6 20.4 51.0
Gender	Male	49.0
Educational attainment	High school or less Higher education, below degree level College degree or graduate degree in	53.1 14.9 32.0
Individual living alone	Yes	17.3
Ethnicity	White (other category is non-white)	91.4
Number of adults	One adult Two adults Three or more adults	22.9 59.2 17.9
Number of children	No kids One kid Two kids or more	58.3 12.4 29.3
Age	Between 16 and 29 years old Between 30 and 55 years old Older than 55	17.5 67.6 14.9
Body mass index	Underweight Normal weight Overweight Obese	1.0 33.1 36.6 29.3
Consumption of fruits and	At least 5 portions per day	31.4
Consumption of fish	Consumed fish during the month	63.2
Commute distance	Average in miles (standard deviation)	12.7 (26.2)

Dependent variables		Participation	Duration*
Commute time allocation	Car Transit Bike Walk	80.2 16.3 6.1 19.0	15.5 17.9 13.6 13.1
Recreational time allocation	Passive activities Moderate physical Vigorous physical	100.0 72.0 34.2	119.2 27.9 10.5

*: Durations are computed only for individuals participating in the corresponding activity.

TABLE 2 Estimation Results of the Multivariate MDCP Model

Variable	Physically Active Recreational Time Allocation in hours/month (*base: Passive Activities*)				Commute Time Allocation in hours/month (*base: Car*)					
	Moderate Activities		Vigorous Activities		Transit		Bike		Walk	
	Coeff.	T-stat	Coeff.	T-stat	Coeff.	T-stat	Coeff.	T-stat	Coeff.	T-stat
Baseline Utility Parameters										
Constants	-5.759	-12.86	-5.664	-35.60	-1.276	-14.35	-1.491	-10.44	-0.871	-6.26
Household annual income (base: over £35,000)										
£0,000-£24,999	-0.236	-2.04	-0.416	-2.63	--	--	0.314	1.86	0.214	1.40
£25,000-£34,999	--	--	-0.257	-1.88	--	--	--	--	0.385	2.53
Age										
Natural logarithm of age	0.377	3.10	--	--	--	--	--	--	--	--
Older than 55 years (dummy)	--	--	--	--	--	--	--	--	-0.337	-1.61
Gender (base: female)										
Male	0.179	1.74	0.460	3.29	--	--	0.637	3.31	--	--
Household structure (base: more than one person)										
Individual living alone	--	--	--	--	0.390	2.74	--	--	--	--
Ethnicity (*base: white*)										
Non-white	--	--	--	--	0.604	3.32	--	--	--	--
Commute distance (in miles)	--	--	--	--	0.003	1.91	-0.050	-3.72	-0.040	-8.23
Body mass index (base: underweight, normal weight or overweight)										
Obese (BMI ≥ 30)	--	--	-0.276	-2.14	--	--	--	--	--	--
Nutritional habits										
Consumed at least 5 portions per day of fruit and vegetables (dummy)	0.185	1.61	0.459	3.24	--	--	--	--	--	--
Consumed fish during the month (dummy)	0.237	2.24	0.390	2.72	--	--	--	--	0.224	1.76
*Endogenous effect***										
Bike commute time	0.017	2.55	--	--	NA	NA	NA	NA	NA	NA
Walk commute time	--	--	--	--	NA	NA	NA	NA	NA	NA
Satiation Parameters*	10.546	7.20	5.381	8.81	68.048	3.64	24.398	3.62	48.583	4.15

--: not significant. NA: not applicable.

Number of observations: 590. Log-likelihood at convergence: -4,172.06. Log-likelihood at only constants: -4,264.51. Adjusted rho square w.r.t constants: 0.0155

***:** Since all the individuals in the sample participate in passive activities, no satiation parameter was estimated for that category. Satiation parameter estimate corresponding to the car alternative in the commute time allocation model is equal to 57.720 with a t-stat of 3.46.

**: The independent model also has a positive effect of bike commute time on moderate activity time allocation. However, unlike the joint model, the independent model has a negative and significant effect of walking commute time on moderate activities time allocation.

- Several socio-demographic and nutrition related variables were tested as explanatory variables of the satiation effects but none of them came out significant.
- The log-likelihood at convergence of the independent model is -4,176.14 and the corresponding adjusted rho square w.r.t constants is 0.0148.
- **Log-likelihood ratio test (between Joint and Independent models):** 8.472 compared to a χ^2 with 2 degrees of freedom (5.99 at 95.0% confidence). To perform the nested test, we included in the joint model the non-significant effect of walk commute time on moderate activity time allocation.

TABLE 3 Estimated Covariance Matrix of Error Differences in the Multivariate MDCP Model (variables significant at 5% level of significance unless otherwise noted)

	Physically Active Recreation (*difference with respect to Passive Activities*)		**Commute time (*difference with respect to Car*)**		
	Moderate Activities	**Vigorous Activities**	**Transit**	**Bike**	**Walk**
Moderate Activities	1.000*				
Vigorous Activities	0.211	1.277			
Transit	0.000*	0.000*	1.000*		
Bike	0.000*	-0.112	0.500*	1.000*	
Walk	0.000*	0.168	0.500*	0.500*	1.000*

* Fixed