A New Econometric Approach to Multivariate Count Data Modeling

Chandra R. Bhat*
The University of Texas at Austin
Dept 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

Rajesh Paleti
Parsons Brinckerhoff
One Penn Plaza, Suite 200
New York, NY 10119
Phone: 512-751-5341
Email: rajeshp@mail.utexas.edu

Marisol Castro
The University of Texas at Austin
Dept 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: m.castro@utexas.edu

*corresponding author

January 2013

ABSTRACT

In the current paper, we propose a modeling framework to explicitly link a count data model with an event type multinomial choice model. The proposed framework uses a multinomial probit kernel for the event type choice model and introduces unobserved heterogeneity in both the count and discrete choice components. Additionally, this paper establishes several new results regarding the distribution of the maximum of multivariate normally distributed variables, which form the basis to embed the multinomial probit model within a joint modeling system for multivariate count data. The model is applied for analyzing out-of-home non-work episodes pursued by workers, using data from the National Household Travel Survey.

Keywords: multivariate count data, generalized ordered-response, multinomial probit, multivariate normal distribution.

1. INTRODUCTION

Count data models are used in several disciplines to analyze discrete and non-negative outcomes without an explicit upper limit. These models assume a discrete probability distribution for the count variables, followed by the parameterization of the mean of the discrete distribution as a function of explanatory variables. The two most commonly used discrete probability distributions are the Poisson and the negative binomial (NB) distributions, though other distributions such as the binomial and logarithmic distributions have also been occasionally considered. Several modifications and generalizations of the Poisson and NB distributions have also been used, including zero-inflated count models (see, for example, Naya *et al*., 2008 and Musio *et al.*, 2010) and hurdle-count models (see, for example, Zhang *et al.*, 2008 and Bethell *et al.*, 2010). While these modifications and generalizations have been effective for use with univariate count models, they are difficult to implement in the case when there are inter-related multivariate counts at play (see Herriges *et al.*, 2008). On the other hand, multivariate count data are ubiquitous in consumer choice situations. For instance, households may patronize different shopping destinations with different frequencies, or may participate in episodes of different activity purposes with different frequencies, or may purchase different counts of brands for frequently purchased grocery items (such as cookies, ready-to-eat cereals, soft drinks, and yoghurt). Such multivariate count data also naturally arise in non-consumer choice settings such as crash frequencies by severity level and crash type.

In the current paper, we propose a parametric framework for multivariate count data that is based on linking a univariate count model for the total count across all possible event states with a discrete choice model for the choice among the event states. For example, the total count may be the total number of grocery shopping occasions within say a month, and the event states may be some discrete representation of locations of participation. In the next section, we discuss closely related efforts in the econometric literature, and position the current paper in the context of earlier research.[1]

[1] There have been several studies in the literature that ignore the joint nature of multivariate count data, and model each count independently from the other (see Terza and Wilson, 1990 and Cameron and Trivedi, 2013). We do not discuss such studies in the next section.

1.1. Earlier Related Research

Three broad approaches have been used in the literature to model multivariate count data: (1) multivariate count models, (2) multiple discrete-continuous models, and (3) joint discrete choice and count models.

1.1.1. Multivariate count models

A multivariate count model may be developed using multivariate versions of the Poisson or negative binomial (NB) discrete distributions (see Buck *et al.*, 2009 and Bermúdez and Karlis, 2011 for recent applications of these methods). These multivariate Poisson and NB models have the advantage of a closed form, but they become cumbersome as the number of events increases and can only accommodate a positive correlation in the counts. Alternatively, one may use a mixing structure, in which one or more random terms are introduced in the parameterization of the mean. The most common form of such a mixture is to include normally distributed terms within the exponentiated mean function, so that the probability of the multivariate counts then requires integration over these random terms (see, for example, Chib and Winkelman, 2001, Awondo *et al.*, 2011 and Haque *et al.*, 2010). The advantage of this method is that it permits both positive and negative dependency between the counts, but the limitations are that the approach gets quickly cumbersome in the presence of several mixing components. Besides, these multivariate count approaches are not based on an underlying utility-maximizing framework; rather they represent a specification for the statistical expectation of demand, and then use relatively mechanical statistical "stitching" devices to accommodate correlations in the multivariate counts. Further, the use of these models do not allow for potentially complex substitution and income effects that are likely to be present across event states in consumer choice decisions. For example, an increase in the price of groceries at one location (say A) may result in an increase in the attractiveness of other grocery locations due to a substitution effect, but also a decrease in total grocery shopping episodes because of an income effect. So, while the frequency of shopping instances to location A will reduce, the frequency of shopping instances to other locations may increase or decrease. The multivariate count models do not explicitly account for such substitution and income effects. Finally, such multivariate count models can be negatively affected by small sample sizes for each event count, and will, in general, necessitate the use of techniques to accommodate excess zeros in the count for each event category, which become difficult in a multivariate setting.

1.1.2. Multiple discrete-continuous models

Another approach that may be used for multivariate count data is to use an explicit utility maximizing framework based on the assumption that consumer preferences can be represented by a random utility function that is quasi-concave, increasing, and continuously differentiable with respect to the consumption quantity vector. Consumers maximize the stochastic utility function subject to one or more budget constraints. The use of a non-linear utility form that allows diminishing marginal utility (or satiation effects) with increasing consumption leads to the possibility of consumption of multiple alternatives and also provides the continuous quantity of the consumed alternatives. Bhat (2008) proposed a general Box-Cox transformation of the translated constant elasticity of substitution (or CES) additive utility function, and showed how the resulting constrained random utility maximization problem can be solved via standard Karush-Kuhn-Tucker (KKT) first order conditions of optimality (see Hanemann, 1978 and Wales and Woodland, 1983 for the initial conceptions of KKT-based model systems, and Kim *et al.*, 2002, von Haefen and Phaneuf, 2005, and Bhat, 2005 for specific implementations of the KKT framework in the past decade). The resulting multiple discrete-continuous (MDC) models have the advantage of being directly descendent from constrained utility maximizing principles, but fundamentally assume that alternatives can be consumed in non-negative and perfectly divisible (*i.e.*, continuous) units. On the other hand, the situation of multivariate counts is truly a discrete-discrete situation, where the alternatives are discrete and the consumption quantity of the consumed alternatives is also discrete. While the MDC model may be a reasonable approximation when the observation period of consumption is long (such as say a year in the context of grocery shopping episodes), a theoretically-consistent formulation that explicitly recognizes the discrete nature of consumption quantity would be more desirable.[2]

[2] von Haefen and Phaneuf (2003) consider a slightly revised version of the KKT-based utility maximization approach for handling multivariate count data. Specifically, they assume a deterministic utility function (rather than a random utility function), derive the implied deterministic continuous consumption vector using KKT conditions, then consider these continuous consumptions as the expected demands, and finally treat the consumer's observed demand for each alternative as an independent draw from a NB distribution with the expected demand function for the alternative as the mean. However, this method is a rather indirect way of accommodating discrete counts, and there is no guarantee that the predicted counts will satisfy the original budget constraint in the KKT framework.

1.1.3. Combined discrete choice and count model

A third approach uses a combination of a total count model to analyze multivariate count data and a discrete choice model for event choice that allocates the total count to different events. This approach has been adopted quite extensively in the literature. Studies differ in whether or not there is a linkage between the total count model and the discrete event choice model. Thus, many studies essentially model the total count using a count model system in the first step, and then independently (and hierarchically, given the total count) develop a multinomial choice model for the choice of event type at each instance of the total number of choice instances (as given by the total count). Since the multivariate count setting does not provide any information on the ordering of the choice instances, the probability of the observed counts in each event type, given the total count, takes a multinomial distribution form (see Terza and Wilson, 1990). This structure, while easy to estimate and implement, does not explicitly consider the substitution and income effects that are likely to lead to a change in total count because of a change in a variable that impacts any event type choice. An alternate and more appealing structure is one that explicitly links the event discrete choice model with the total count model. In this structure, the expected value of the maximum utility from the event type multinomial model is used as an explanatory variable in the conditional expectation for the total count random variable (see Mannering and Hamed, 1990, Hausman *et al.*, 1995, and Rouwendal and Boter, 2009). But a problem with this structure is that it fails to recognize the effects of unobserved factors in the event type alternative utilities on the total count (because only the expected value of maximum utility enters the count model intensity, with no mapping of the choice errors into the count intensity). On the other hand, the factors in the unobserved portions of utilities must also influence the count intensity just as the observed factors in the utilities do. This is essential to recognize the full econometric jointness between the event choice and the total count decision. In the case when a generalized extreme value (GEV) model is used for the event choice (as has been done in the past), the maximum over the utilities is also GEV distributed, but including the resulting error term in the count intensity leads to distributional mismatch issues. As indicated by Burda *et al.* (2012), while the situation may be resolved by using Bayesian augmentation procedures, these tend to be difficult to implement, particularly when random taste variations across individuals are also present in the event choice model.

1.2. The Current Paper

In the current paper, we use the third approach discussed above, while also accommodating the full jointness in the total count and event choice components of the model system. In this context, there are four aspects of the proposed model system that are novel in the literature. First, we use a multinomial probit (MNP) kernel for the event choice type model, rather than the traditional multinomial logit (MNL) or nested logit (NL) kernel used in earlier studies. The use of the MNP kernel allows a more flexible covariance structure for the event utilities relative to traditional GEV kernels, while also facilitating the linkage between the event choice and the total count components of our joint model system. Second, we allow random taste variations (or unobserved heterogeneity) in the sensitivity to exogenous factors in both the event choice model as well as the total count components. This is accomplished by recasting the total count model as a special case of a generalized ordered-response model in which a single latent continuous variable is partitioned into mutually exclusive intervals (see Castro, Paleti, and Bhat, 2012 or CPB in the rest of this paper). This recasting is a key precursor element of how we link the event type MNP model and the total count model (which is the focus of this paper). Third, we establish several new results regarding the distribution of the maximum of multivariate normally distributed random variables (with a general covariance matrix) as well as its stochastic affine transformations. These results constitute another core element in our approach to link the event and total count components, in addition to being important in their own right. In particular, the use of GEV structures in the past for event choice in joint models has ostensibly been because the exact form of the maximum of GEV distributed variables is well known. We show that similar results do also exist for the maximum of normally distributed variables. This obviates the need to adopt relatively cumbersome and indirect ways to deal with the complications arising from the introduction of the maximum utility term from the event component into the count model. Fourth, the resulting model for multivariate count data can be easily estimated using Bhat's (2011) frequentist MACML (for maximum composite marginal likelihood) approach. More broadly, the approach in this paper should open up a whole new set of applications in consumer choice modeling, because the analyst can now embed an MNP model within a joint modeling system for multivariate count data.

The rest of this paper is structured as follows. The next section presents the fundamental structure of the multivariate normal distribution and new results regarding the distribution of the maximum of normally distributed variables. Section 3 presents the model framework and estimation procedure for the proposed joint count and discrete choice model. Section 4 illustrates an application of the proposed model for analyzing out-of-home non-work episodes pursued by workers. Finally, Section 5 summarizes the key findings of the paper and identifies directions for further research.

2. RESULTS ON THE MAXIMUM OF NORMALLY DISTRIBUTED RANDOM VARIABLES

2.1. An Overview

Consider a vector of I absolutely continuous random variables $\boldsymbol{X} = (X_1, X_2, ...X_I)$. Under the assumption that the univariate random variables X_i are independent of one another, the asymptotic distribution of $\eta = \underset{i}{Max}\{X_i\} = Max(\boldsymbol{X})$ has received substantial attention in the operations research, statistics, and production literature (see, for example, Gumbel, 1958, Clark, 1961, and David, 1981, Chapter 8). Specifically, even if the exact density functions of the continuous univariate random variables are unknown, the asymptotic distribution of the maximum approaches the type-I extreme value distribution as long as the density function of each random variable decays in the upper tail as an exponential function. For the case when the density functions of the univariate random variables are exactly known, results exist for the exact distribution of the maximum of a finite set of independently distributed continuous variables (see Bose and Gupta, 1959 and David, 1981, Chapter 8). Approximations to the case of general dependence structures have been developed for the specific case when the random variables are multivariate normally distributed, mostly for the case of two or three random variables (Greig, 1967, Clark, 1961, Devroye, 1980). Later, Tong (1990) derived the exact probability density function of the maximum of normally distributed variables, but for the very specific case when the random variables follow an exchangeable multivariate normal distribution.[3] But it was not until the research of Arellano-Valle and Genton (2008) that an exact density function was obtained for η for arbitrarily dependent random variables. Jamalizadeh and Balakrishnan (2009, 2010) extended Arellano-Valle and Genton's results to obtain the moment generating function and moments of η for the case when the vector $\boldsymbol{X}$ has an elliptically contoured distribution.

[3] An exchangeable multivariate normally distributed vector $\boldsymbol{X}$ has a mean vector with the same element in all rows, and a covariance matrix with identical entries along the diagonal and identical entries on all the off-diagonals.

These works rest on showing that the density function of the maximum of an elliptically contoured distribution is a mixture of unified univariate skew-elliptical density functions. In particular, these studies show that the density function of η, when $\boldsymbol{X}$ has a general multivariate normal distribution, is a mixture of unified univariate skew-normal density functions.

In this paper, we revisit the problem of the density and cumulative distribution functions of η for a general multivariate normal distribution on $\boldsymbol{X}$. We provide another approach to deriving the density function for η that is based on a different way of writing the cumulative distribution function of η. Indeed, we have not seen our very simple expression for the cumulative distribution function of η appear in the literature discussed above. This alternative form provides a more direct expression for the density function of η, enables the derivation of new results on affine stochastic transformations of η that facilitates the linking of the total count and event choice models, and allows us to write the likelihood function for the resulting joint model system in a way that facilitates estimation.

2.2. Expressions Involving the Maximum of Random Variables with a Multivariate Normal Distribution

In this section, we list the important properties of η, with a particular emphasis on those that are important to the joint total count and event model system under consideration in this paper. However, some theorems are provided in general and have not appeared in the literature, and should be of use in many contexts well beyond the current methodological motivation. Here, we will assume that $\boldsymbol{X} \sim MVN_I(\boldsymbol{b}, \boldsymbol{\Sigma})$, where MVN_I stands for the multivariate normal distribution of *I* dimensions with mean vector $\boldsymbol{b}$ and covariance matrix $\boldsymbol{\Sigma}$. Other key notations that will be used here and the rest of this paper are as follows: $\mathbf{IDEN}_R$ for an identity matrix of dimension *R*, $\mathbf{1}_R$ for a column vector of ones of dimension *R*, $\boldsymbol{0}_R$ for a column vector of zeros of dimension *R*, $\mathbf{1}_{RR}$ for a matrix of ones of dimension *R*×*R*, $f(.;\mu,\sigma^2)$ for the univariate normal density function with mean μ and variance σ^2, $\phi(.)$ for the univariate standard normal density function, $f_R(.;\boldsymbol{\tau},\boldsymbol{\Gamma})$ for the multivariate normal density function of dimension *R* with mean vector $\boldsymbol{\tau}$ and covariance matrix $\boldsymbol{\Gamma}$, $\boldsymbol{\omega}_{\boldsymbol{\Gamma}}$ for the diagonal matrix of the standard deviations of $\boldsymbol{\Gamma}$, with its r^{th} element being $\omega_{\boldsymbol{\Gamma}r}$, $\phi_R(.;\boldsymbol{\Gamma}^*)$ for the multivariate standard normal density function of dimension

R and correlation matrix $\boldsymbol{\Gamma}^*$, $F(.;\mu,\sigma^2)$ for the univariate normal cumulative distribution function with mean μ and variance σ^2, $\Phi(.)$ for the univariate standard normal cumulative distribution function, $F_R(.;\boldsymbol{\tau},\boldsymbol{\Gamma})$ for the multivariate normal cumulative distribution function of dimension R with mean vector $\boldsymbol{\tau}$ and covariance matrix $\boldsymbol{\Gamma}$, and $\Phi_R(.;\boldsymbol{\Gamma}^*)$ for the multivariate standard normal cumulative distribution function of dimension R and correlation matrix $\boldsymbol{\Gamma}^*$. Also, the following well established results for the multivariate normal distribution are collected together in a single Lemma (without proof) for further use in the paper.

Lemma 1

1) The multivariate normal density function and cumulative distribution function of dimension R are respectively given by $f_R(\boldsymbol{z};\boldsymbol{\tau},\boldsymbol{\Gamma}) = \left(\prod_{r=1}^{R} \omega_{\Gamma r}\right)^{-1} \phi_R\left(\boldsymbol{\omega}_{\boldsymbol{\Gamma}}^{-1}[\boldsymbol{z}-\boldsymbol{\tau}];\boldsymbol{\Gamma}^*\right)$ and $F_R(\boldsymbol{z};\boldsymbol{\tau},\boldsymbol{\Gamma}) = \Phi_R\left(\boldsymbol{\omega}_{\boldsymbol{\Gamma}}^{-1}[\boldsymbol{z}-\boldsymbol{\tau}];\boldsymbol{\Gamma}^*\right)$, where $\boldsymbol{\Gamma}^* = \boldsymbol{\omega}_{\boldsymbol{\Gamma}}^{-1}\boldsymbol{\Gamma}\boldsymbol{\omega}_{\boldsymbol{\Gamma}}^{-1}$.

2) Let $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ be normally distributed vectors of dimension I_1 and I_2, respectively. The corresponding mean vector and covariance matrix of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are $(\boldsymbol{b}_1,\boldsymbol{\Sigma}_{11})$ and $(\boldsymbol{b}_2,\boldsymbol{\Sigma}_{22})$. Defining $\boldsymbol{X} = (\boldsymbol{X}_1',\boldsymbol{X}_2')'$, $\boldsymbol{b} = (\boldsymbol{b}_1',\boldsymbol{b}_2')$ and $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}' \\ \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$, where $\boldsymbol{\Sigma}_{12}$ is the covariance matrix between $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, the conditional distribution of $\boldsymbol{X}_2$ given $\boldsymbol{X}_1$ is $\boldsymbol{X}_2 \mid (\boldsymbol{X}_1 = \boldsymbol{x}_1) = MVN_{I_2}[\boldsymbol{b}_2 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{b}_1), \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}']$. Then,

$$\frac{\partial F_I(\boldsymbol{X}_1 = \boldsymbol{x}_1, \boldsymbol{X}_2 = \boldsymbol{x}_2)}{\partial \boldsymbol{x}_1} = f_{I_1}(\boldsymbol{x}_1;\boldsymbol{b}_1,\boldsymbol{\Sigma}_{11}) \times F_{I_2}(\boldsymbol{x}_2;\boldsymbol{b}_2 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{b}_1), \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}').$$

In what follows, we present and discuss four theorems that are key to the proposal in this paper.

Theorem 1

Let $\boldsymbol{b}_{-i}$ be the sub-vector of $\boldsymbol{b}$ without the i^{th} element, let b_i be the i^{th} element of $\boldsymbol{b}$, let $\boldsymbol{\Sigma}_{-i,-i}$ be the sub-matrix of $\boldsymbol{\Sigma}$ without the i^{th} row and the i^{th} column, let $\omega_{\Sigma i}^2$ be the diagonal entry at the i^{th} row and i^{th} column of $\boldsymbol{\Sigma}$, and let $\overline{\boldsymbol{\Sigma}}_{-i}$ be the i^{th} column of the matrix $\boldsymbol{\Sigma}$ minus the i^{th} row element.

Denote the univariate cumulative distribution function of $\eta = \text{Max}(\boldsymbol{X})$ by $G(z;\boldsymbol{b},\boldsymbol{\Sigma})$ and the probability density function of η by $g(z;\boldsymbol{b},\boldsymbol{\Sigma})$. Then:

$G(z;\boldsymbol{b},\boldsymbol{\Sigma}) = F_I\left(z\mathbf{1}_I;\boldsymbol{b},\boldsymbol{\Sigma}\right)$, and

$$g(z;\boldsymbol{b},\boldsymbol{\Sigma}) = \sum_{i=1}^{I} f\left(z;b_i,\omega_{\Sigma i}^2\right)\times F_{I\text{-}1}\left(z\mathbf{1}_{I-1};\boldsymbol{b}_{\text{-}i} + \overline{\boldsymbol{\Sigma}}_{\text{-}i}(\omega_{\Sigma i}^2)^{\text{-}1}(z-b_i),\boldsymbol{\Sigma}_{\text{-}i,\text{-}i} - \overline{\boldsymbol{\Sigma}}_{\text{-}i}(\omega_{\Sigma i}^2)^{\text{-}1}\overline{\boldsymbol{\Sigma}}'_{\text{-}i}\right).$$

Proof:

$$\begin{aligned} G(z;\boldsymbol{b},\boldsymbol{\Sigma}) &= Prob[\text{Max}(\boldsymbol{X})<z] = Prob[X_1 < z \text{ and } X_2 < z \text{ and}\ldots. X_I < z] \\ &= F_I(z\mathbf{1}_I;\boldsymbol{b},\boldsymbol{\Sigma}). \end{aligned}$$

The above simple approach to write the distribution function $G(z;\boldsymbol{b},\boldsymbol{\Sigma})$ does not seem to have appeared in the literature, with earlier studies deriving this function in a different (and more complicated) fashion by writing it as a mixture function of unified univariate skew-normal cumulative distribution functions (see Jamalizadeh and Balakrishnan, 2010 and Arellano-Valle and Genton, 2008). The proof that the density function takes the form as given above can be shown by differentiating $G(z;\boldsymbol{b},\boldsymbol{\Sigma})$ with respect to z and using the last result from Lemma 1.

Theorem 2

The density function of η may be written as the sum of log-concave density functions, which implies that the density function of η is a mixture of strongly unimodal density functions. This result has also appeared in Jamalizadeh and Balakrihnan (2010), though in a slightly different form. Appendix A provides the proof.

Theorem 3

The distribution of a stochastic transformation of $\eta = \text{Max}(\boldsymbol{X})$ as $\xi = \vartheta\eta + W$, where ϑ is a constant scalar parameter and W is a univariate normally distributed scalar ($W \sim N(\mu,\upsilon^2)$), has a cumulative distribution function and density function as below:

$H(z;\boldsymbol{b},\boldsymbol{\Sigma},\vartheta,\mu,\upsilon^2) = F_I(z\mathbf{1}_I;\vartheta\boldsymbol{b}+\mu\mathbf{1}_I,\vartheta^2\boldsymbol{\Sigma}+\mathbf{IDEN}_I\upsilon^2)$, and

$$h(z;\boldsymbol{b},\boldsymbol{\Sigma},\vartheta,\mu,\upsilon^2) = \sum_{i=1}^{I} f(z;\vartheta b_i+\mu,\vartheta^2\omega_{\Sigma i}^2+\upsilon^2)\times F_{I-1}(z\mathbf{1}_{I-1};\tilde{\boldsymbol{b}}_{i\xi},\tilde{\boldsymbol{\Sigma}}_{i\xi}),$$

where $\tilde{\boldsymbol{b}}_{i\xi} = \vartheta\boldsymbol{b}_{\text{-}i} + \mu\mathbf{1}_{I-1} + \vartheta^2\overline{\boldsymbol{\Sigma}}_{\text{-}i}(\vartheta^2\omega_{\Sigma i}^2+\upsilon^2)^{\text{-}1}(z-\vartheta b_i-\mu)$, and

$$\widetilde{\mathbf{\Sigma}}_{i\xi} = \vartheta^2 \mathbf{\Sigma}_{-i,-i} + \mathbf{IDEN}_{I-1}\upsilon^2 - \vartheta^2 \overline{\mathbf{\Sigma}}_{-i}(\vartheta^2\omega^2_{\Sigma i} + \upsilon^2)^{-1}(\vartheta^2\overline{\mathbf{\Sigma}}_{-i})' .$$

Appendix A provides the proof. The cumulative distribution form above has not appeared in the literature and will be useful in the model formulated in the current paper to handle multivariate count data. The proof that the density function takes the form as given in the theorem above can be shown by differentiating $H(z;\boldsymbol{b},\mathbf{\Sigma},\vartheta,\mu,\upsilon^2)$ with respect to z and using the last result from Lemma 1.

Theorem 4

The moment generating function of $\xi = \vartheta\eta + W$ is given by:

$$M_\xi(t) = \int_{z=-\infty}^{\infty} e^{tz} h(z;\boldsymbol{b},\mathbf{\Sigma},\vartheta,\mu,\upsilon^2)dz = \sum_{i=1}^{I} e^{t(\vartheta b_i+\mu)+\frac{1}{2}\omega^2_{\Sigma i\xi}t^2} \times F_{I-1}(\boldsymbol{s}_{\Sigma i\xi}\, t;\boldsymbol{\gamma}_i, \mathbf{\Psi}_{i\xi}) ,$$

where $\omega_{\Sigma i\xi} = \sqrt{\vartheta^2\omega^2_{\Sigma i} + \upsilon^2}$, $\boldsymbol{s}_{\Sigma i\xi} = \boldsymbol{\lambda}_i\,\omega_{\Sigma i\xi}$, $\boldsymbol{\lambda}_i = \omega_{\Sigma i\xi}\mathbf{1}_{I-1} - \dfrac{\vartheta^2\overline{\mathbf{\Sigma}}_{-i}}{\omega_{\Sigma i\xi}}$, $\boldsymbol{\gamma}_i = \vartheta(\boldsymbol{b}_{-i} - b_i\mathbf{1}_{I-1})$ and $\mathbf{\Psi}_{i\xi} = \widetilde{\mathbf{\Sigma}}_{i\xi} + \boldsymbol{\lambda}_i\,\boldsymbol{\lambda}_i'$. The above result allows the computation of the moments of the variable ξ, which can be helpful in many model formulations. The proof of this theorem is available in Appendix A.

Corollary to Theorem 4

Let $\boldsymbol{\gamma}_{i,-l}$ be the vector $\boldsymbol{\gamma}_i$ minus the l^{th} row element, γ_{il} the l^{th} element of the vector $\boldsymbol{\gamma}_i$, $s_{\Sigma i\xi l}$ the l^{th} element of the vector $\boldsymbol{s}_{\Sigma i\xi}$, $\mathbf{\Psi}_{i\xi,-l,-l}$ the sub-matrix of $\mathbf{\Psi}_{i\xi}$ without the l^{th} row and the l^{th} column, $\sigma^2_{\Psi il}$ be the diagonal entry at the l^{th} row and l^{th} column of $\mathbf{\Psi}_{i\xi}$, $\overline{\mathbf{\Psi}}_{i\xi,-l}$ be the l^{th} column of the matrix $\mathbf{\Psi}_{i\xi}$ minus the l^{th} row element, and the matrix $\mathbf{\Delta}_{il\xi} = \mathbf{\Psi}_{i\xi,-l,-l} - \overline{\mathbf{\Psi}}_{i\xi,-l}(\sigma^2_{\Psi il})^{-1}\overline{\mathbf{\Psi}}'_{i\xi,-l}$. Through straightforward, but tedious, differentiation, the following additional result may be obtained.

$$\begin{aligned} E(\xi) &= \left.\frac{dM_\xi(t)}{dt}\right|_{t=o} = \sum_{i=1}^{I}\Big\{(\vartheta b_i + \mu)F_{I-1}(\boldsymbol{0}_{I-1};\boldsymbol{\gamma}_i,\mathbf{\Psi}_{i\xi}) \\ &\quad + \sum_{l=1}^{I-1} s_{\Sigma i\xi\xi} f(0;\gamma_{il},\sigma^2_{\Psi il}) \times F_{I-2}\Big[\boldsymbol{0}_{I-2};\Big(\boldsymbol{\gamma}_{i,-l} - \overline{\mathbf{\Psi}}_{i\xi,-l}(\sigma^2_{\Psi il})^{-1}\gamma_{il}\Big),\mathbf{\Delta}_{il\xi}\Big]\Big\}. \end{aligned}$$

Higher order moments of ξ may be obtained by additional orders of differentiation.

3. THE JOINT EVENT TYPE-TOTAL COUNT MODEL SYSTEM

Let the total observed demand count over a certain period of interest for consumer q $(q = 1,2,...,Q)$ be n_q. Also, let there be I $(i = 1,2,...,I)$ event type possibilities (or alternatives) that the total count n_q may be allocated to (the number of event types may vary across decision agents; however, for ease in presentation and also because the case of varying number of event types does not pose any complications, we assume the same number of alternatives across all consumers). Each count unit contribution to the total count n_q corresponds to a choice occasion from among the I alternatives. Thus, one may view the choice situation as a case of repeated choice data, with n_q choice occasions and time-invarying independent variables.[4] The "chosen" alternative at each choice occasion is developed such that the total number of times an alternative is "chosen" across the n_q choice occasions equals the actual count in that alternative (the order of the assignment of the "chosen" alternatives across choice occasions is immaterial, and does not affect the estimation in any way). The resulting repeated choice data allows the estimation of individual-specific unobserved factors that influence the intrinsic preference for each alternative as well as the responsiveness to independent variables.

The next section presents the formulation for the event choice at each choice occasion, while the subsequent section develops the formulation for the total count model (including the linkage between the event choice and the total count).

3.1. Event Type Choice Model

Consider the following random-coefficients formulation in which the utility U_{qti} that an individual q associates with alternative i at choice occasion t is given by:

$$U_{qti} = \boldsymbol{\beta}_q' \boldsymbol{x}_{qi} + \tilde{\varepsilon}_{qti}; \ \ \boldsymbol{\beta}_q = \boldsymbol{b} + \tilde{\boldsymbol{\beta}}_q \ , \ \tilde{\boldsymbol{\beta}}_q \sim MVN_D(\boldsymbol{0}_D, \boldsymbol{\Omega}) \ , \tag{1}$$

where $\boldsymbol{x}_{qi}$ is a (D×1)-column vector of exogenous attributes (including a constant), and $\boldsymbol{\beta}_q$ is an individual-specific (D×1)-column vector of corresponding coefficients that is a realization from a

[4] In many situations, the count by event type is explicitly based on observation or reported decisions at a choice occasion level (such as individuals reporting all the activity episodes by type of participation over a day, or recalling each recreational trip participated in over a period of time).

multivariate normal density function with mean vector $\boldsymbol{b}$ and covariance matrix $\boldsymbol{\Omega}$ (this specification allows taste variation as well as generic preference variations due to unobserved individual attributes). $\tilde{\varepsilon}_{qti}$ is assumed to be an independently and identically distributed (across choice occasions and across individuals) error term, but having a general covariance structure across alternatives at each choice occasion. Thus, consider the (I×1)-vector $\tilde{\boldsymbol{\varepsilon}}_{qt} = (\tilde{\varepsilon}_{qt1}, \tilde{\varepsilon}_{qt2}, \tilde{\varepsilon}_{qt3}, \ldots, \tilde{\varepsilon}_{qtI})'$. We assume that $\tilde{\boldsymbol{\varepsilon}}_{qt} \sim MVN_I(\boldsymbol{0}_I, \boldsymbol{\Theta})$. To accommodate the invariance in choice probabilities to utility function translations and scaling, appropriate identification considerations need to be imposed on $\boldsymbol{\Theta}$. An appealing approach is to take the differences of the error terms with respect to the first error term (the designation of the first alternative is arbitrary). Let $\varepsilon_{qti1} = (\tilde{\varepsilon}_{qti} - \tilde{\varepsilon}_{qt1})$, and let $\boldsymbol{\varepsilon}_{qt1} = (\varepsilon_{qt21}, \varepsilon_{qt31}, \ldots, \varepsilon_{qtI1})$. Then, up to a scaling factor, the covariance matrix of $\boldsymbol{\varepsilon}_{qt1}$ (say $\boldsymbol{\Theta}_1$) is identifiable. Next, scale the top left diagonal element of this error-differenced covariance matrix to 1. Thus, there are $[(I-1)\times(I/2)]-1$ free covariance terms in the $(I-1)\times(I-1)$ matrix $\boldsymbol{\Theta}_1$. Later on during estimation, we will take the difference of the utilities with respect to the chosen alternative (not the first alternative). But to ensure that, whenever differences are taken with respect to the chosen alternative, these differences are consistent with the same error covariance matrix $\boldsymbol{\Theta}$ for the undifferenced error term vector $\tilde{\boldsymbol{\varepsilon}}_{qt}$, $\boldsymbol{\Theta}$ is effectively constructed from $\boldsymbol{\Theta}_1$ by adding a top row of zeros and a first column of zeros (see Train, 2003; page 134).

We now set out some additional notation. Define $\boldsymbol{U}_{qt} = (U_{qt1}, U_{qt2}, \ldots, U_{qtI})'$ (I×1 vector), $\boldsymbol{U}_q = (\boldsymbol{U}'_{q1}, \boldsymbol{U}'_{q2}, \ldots, \boldsymbol{U}'_{qT})'$ (TI×1 vector), $\tilde{\boldsymbol{\varepsilon}}_q = (\tilde{\boldsymbol{\varepsilon}}'_{q1}, \tilde{\boldsymbol{\varepsilon}}'_{q2}, \ldots, \tilde{\boldsymbol{\varepsilon}}'_{qT})'$ (TI×1 vector), and $\mathbf{x}_q = (\boldsymbol{x}_{q1}, \boldsymbol{x}_{q2}, \ldots, \boldsymbol{x}_{qI})'$ (I×D matrix). Then, we can write:

$$\boldsymbol{U}_q = \left(\mathbf{1}_T \otimes [\mathbf{x}_q \boldsymbol{b}]\right) + \left(\mathbf{1}_T \otimes [\mathbf{x}_q \tilde{\boldsymbol{\beta}}_q] + \tilde{\boldsymbol{\varepsilon}}_q\right) = \boldsymbol{V}_q + \boldsymbol{\varepsilon}_q, \tag{2}$$

where $\boldsymbol{V}_q = \mathbf{1}_T \otimes [\mathbf{x}_q \boldsymbol{b}]$ and $\boldsymbol{\varepsilon}_q = \mathbf{1}_T \otimes [\mathbf{x}_q \tilde{\boldsymbol{\beta}}_q] + \tilde{\boldsymbol{\varepsilon}}_q$. Also, assume that individual q chooses alternative m_{qt} at the t^{th} choice instance. Define $\boldsymbol{M}_q$ as an $[(I-1)\times T]\times[TI]$ block diagonal matrix, with each block diagonal having $(I-1)$ rows and I columns corresponding to the q^{th} individual's t^{th} choice instance. This $(I-1)\times I$ matrix for individual q and observation time period t corresponds to an $(I-1)$ identity matrix with an extra column of -1's added as the m_{qt}^{th} column. In the utility differential form (where the utility differentials are taken with respect to the chosen alternative m_{qt} at each choice occasion), we may write Equation (2) as:

$$\boldsymbol{u}_q^* = \boldsymbol{M}_q \boldsymbol{U}_q = \boldsymbol{M}_q \boldsymbol{V}_q + \boldsymbol{M}_q \boldsymbol{\varepsilon}_q \,. \tag{3}$$

To determine the covariance matrix of $\boldsymbol{u}_q^*$, define $\tilde{\boldsymbol{\Omega}}_q = \mathbf{1}_{TT} \otimes [\mathbf{x}_q \boldsymbol{\Omega} \mathbf{x}_q]$ $(TI \times TI$ matrix) and $\tilde{\boldsymbol{\Theta}} = \mathbf{IDEN}_T \otimes \boldsymbol{\Theta}$ $(TI \times TI$ matrix). Let $\tilde{\mathbf{F}}_q = [\tilde{\boldsymbol{\Omega}}_q + \tilde{\boldsymbol{\Theta}}]$ and $\mathbf{F}_q = \boldsymbol{M}_q \tilde{\mathbf{F}}_q \boldsymbol{M}'_q$. Also, let $\boldsymbol{H}_q = \boldsymbol{M}_q \boldsymbol{V}_q$. Finally, we obtain the result below:

$$\boldsymbol{u}_q^* \sim MVN_{(I-1)\times n_q}(\boldsymbol{H}_q, \mathbf{F}_q) \,. \tag{4}$$

The parameters to be estimated in the event type model include the $\boldsymbol{b}$ vector, and the elements of the covariance matrices $\boldsymbol{\Omega}$ and $\boldsymbol{\Theta}$. The likelihood contribution of individual q from the event type choice model is the $[(I-1)\times n_q]$-dimensional integral below:

$$L_{q,event}(\boldsymbol{b}, \boldsymbol{\Omega}, \boldsymbol{\Theta}) = P(\boldsymbol{u}_q^* < 0) = \Phi_{(I-1)\times n_q}\left[(\boldsymbol{\omega}_{\mathbf{F}_q})^{-1}(-\boldsymbol{H}_q), (\boldsymbol{\omega}_{\mathbf{F}_q})^{-1}\mathbf{F}_q(\boldsymbol{\omega}_{\mathbf{F}_q})^{-1}\right], \tag{5}$$

where $\boldsymbol{\omega}_{\mathbf{F}_q}$ is the diagonal matrix of standard deviations of $\mathbf{F}_q$. The above likelihood function has a high dimensionality of integration, especially when the total number of counts n_q and/or the number of alternatives I is high. To resolve this, we use the MACML approach proposed by Bhat (2011), which involves the evaluation of only univariate and bivariate cumulative normal distribution evaluations. However, note that the parameters from the event type model also appear in the total count model, and hence we discuss the overall estimation procedure for the total count-event type model in Section 3.3 after first discussing the total count model formulation in the next section.

3.2. Total Count Model

A key to linking the event type choice model to the total count model is our recasting of the count model as a generalized ordered-response (GOR) model. Specifically, as discussed by CPB (2012), any count model may be reformulated as a special case of a GOR model in which a single latent continuous variable is partitioned into mutually exclusive intervals. Using this equivalent latent variable-based GOR framework for count data models, we are then able to gainfully and efficiently introduce the linkage from the event choice model to the count model through the latent continuous variable. The formulation also allows handling excess zeros in a straightforward manner.

We first provide a brief overview of CPB's recasting of the count model as a special case of the GOR probit (or GORP) model in Section 3.2.1, and then discuss the linkage with the event type model in Section 3.2.2.

3.2.1. The basic recasting

As earlier, let q $(q=1,2,...,Q)$ be the index for the consumer and let k $(k=0,1,2,...,\infty)$ be the index to represent the count level (n_q, the total observed count for consumer q, takes a specific value in the domain of k). Consider the following form of the GORP model system:

$$g_q^* = \boldsymbol{\theta}_q' \boldsymbol{w}_q + \zeta_q, \; g_q = k \text{ if } \delta_{q,k-1} < g_q^* < \delta_{qk}, \; \delta_{qk} = f_k(\boldsymbol{\varpi}_q) + \alpha_k, \quad (6)$$

where α_k is a scalar similar to the thresholds in a standard ordered-response model $(\alpha_{-1} = -\infty; \alpha_0 = 0$ for identification, and $0 < \alpha_1 < \alpha_2 < \alpha_3 < ...)$, and $f_k(\boldsymbol{\varpi}_q)$ is a non-linear function of a vector of consumer-specific variables $\boldsymbol{\varpi}_q$ that ensures that the thresholds δ_{qk} satisfy the ordering conditions ($\delta_{q,-1} = -\infty; -\infty < \delta_{q0} < \delta_{q1} < \delta_{q2} < \delta_{q,3} < ...$) in the usual ordered-response fashion. g_q^* in Equation (6) corresponds to the latent propensity underlying the observed count variable g_q, $\boldsymbol{w}_q$ is an (L×1)-column vector of exogenous attributes (excluding a constant), $\boldsymbol{\theta}_q$ is a corresponding (L×1)-column vector of individual-specific variable effects, and ζ_q is an idiosyncratic random error term assumed to be identically and independently standard normal distributed across individuals q.[5]

Several points about the GORP model of Equation (6) are noteworthy, as discussed by CPB. First, the model in Equation (6) can exactly reproduce any traditional count data model. Second, the analyst can accommodate high or low probability masses for specific count

[5] The use of the standard normal distribution rather than a non-standard normal distribution for the error term is an innocuous normalization to recognize the invariance of the probability in the GORP model to scaling of the latent propensity measure (see Zavoina and McKelvey, 1975; Greene and Hensher, 2010).

outcomes by estimating some of the α_k parameters in the threshold function. At the same time, the GORP model can estimate the probability for any arbitrary count value. All that needs to be done is to identify a count value *K* above which α_k is held fixed at α_K; that is, $\alpha_k = \alpha_K$ for all $k > K$. The analyst can empirically test different values of *K* and compare data fit to determine the optimal value of *K* to add flexibility over the traditional count specification (that constrains all α_k parameters to zero). Third, the interpretation of the GORP recasting is that consumers have a latent "long-term" (and constant over a certain time period) propensity g_q^* associated with the demand for the product/service under consideration that is a linear function of a set of consumer-related attributes $\boldsymbol{w}_q$. On the other hand, there may be some specific consumer contexts and characteristics (embedded in $\boldsymbol{\varpi}_q$) that may dictate the likelihood of the long-term propensity getting translated into a manifested demand at any given *instant of time* (there may be common elements in $\boldsymbol{w}_q$ and $\boldsymbol{\varpi}_q$). Further, as will be clear in the next section, our implicit assumption in linking the total count model to the event type choice model is that the maximum utility (or a measure of per unit consumer surplus) from the event type choice model affects the "long-term" latent demand propensity g_q^*, but does not play a role in the instantaneous translation of propensity to actual manifested demand. That is, the factors/constraints that are responsible for the instantaneous translation of propensity to manifested demand are not impacted by changes in the quality attributes of the consumer product alternatives (that is, of the event types), but the "long-term" demand propensity is.

3.2.2. Linkage with the event type choice model

To link the event type choice model with the count model, we need a measure of maximum utility from the event choice model in the count model. In this manner, an improvement in the quality or a reduction in price of any alternative in the choice model gets manifested as an increase in overall utility (or consumer surplus) per choice occasion, resulting in a higher propensity for the consumer product under consideration and an increase in the total count of units purchased. To develop this link, consider the utility expressions of each alternative in the event choice model at any choice occasion *t* $(t = 1,2,...,n_q)$. Since these expressions do not vary across choice occasions during the observation period, we can ignore the index *t*, as we now do.

From Equation (1), the utility expression for alternative i at any choice occasion is then as follows:

$$\breve{U}_{qi} = \boldsymbol{\beta}'_q \mathrm{x}_{qi} + \breve{\varepsilon}_{qi}; \ \boldsymbol{\beta}_q = \boldsymbol{b} + \tilde{\boldsymbol{\beta}}_q, \tilde{\boldsymbol{\beta}}_q \sim MVN_D(\boldsymbol{0}_D, \boldsymbol{\Omega}). \tag{8}$$

Define $\breve{\boldsymbol{U}}_q = (\breve{U}_{q1}, \breve{U}_{q2}, ..., \breve{U}_{qI})'$ ($I \times 1$ vector) and $\breve{\boldsymbol{\varepsilon}}_q = (\breve{\varepsilon}_{q1}, \breve{\varepsilon}_{q2}, ..., \breve{\varepsilon}_{qI})'$ ($I \times 1$ vector). With other definitions as earlier, we may write:

$$\breve{\boldsymbol{U}}_q = [\mathbf{x}_q \boldsymbol{b}] + \left([\mathbf{x}_q \tilde{\boldsymbol{\beta}}_q] + \breve{\boldsymbol{\varepsilon}}_q \right). \tag{9}$$

This vector $\breve{\boldsymbol{U}}_q$ is normally distributed as follows: $\breve{\boldsymbol{U}}_q \sim MVN_I(\boldsymbol{d}_q, \boldsymbol{\Sigma}_q)$, where $\boldsymbol{d}_q = \mathbf{x}_q \boldsymbol{b}$ and $\boldsymbol{\Sigma}_q = \mathbf{x}_q \boldsymbol{\Omega} \mathbf{x}'_q + \boldsymbol{\Theta}$. Following the notation in Section 2, let $\eta_q = \text{Max}(\breve{\boldsymbol{U}}_q)$. We introduce this variable in the total count model of Equation (6) as follows:

$$g_q^* = \left(\boldsymbol{\theta} + \tilde{\boldsymbol{\theta}}_q\right)' \boldsymbol{w}_q + \vartheta \eta_q + \zeta_q, \ g_q = k \text{ if } \delta_{q,k-1} < g_q^* < \delta_{qk}, \ k \in \{0, 1, 2, ..., \infty\},$$
$$\text{with } \delta_{qk} = \Phi^{-1}\left(e^{-\lambda_q} \sum_{l=0}^{k} \frac{\lambda_q^l}{l!} \right) + \alpha_k, \text{ where } \lambda_q = e^{\boldsymbol{\varphi}' \boldsymbol{\varpi}_q}, \ \delta_{q,-1} = -\infty, \text{ and } \alpha_0 = 0. \tag{10}$$

$\tilde{\boldsymbol{\theta}}_q$ in the equation above is an individual-specific coefficient vector introduced to account for unobserved heterogeneity in the demand propensity, and is assumed to be distributed multivariate normal: $\tilde{\boldsymbol{\theta}}_q \sim MVN_L(\boldsymbol{0}_L, \boldsymbol{\Xi})$. It is assumed that $\tilde{\boldsymbol{\theta}}_q$ is independent of ζ_q. The long-term propensity in Equation (10) may be re-written as follows:

$$g_q^* = \vartheta \eta_q + W_q, \text{ where } W_q \sim N(\mu_q, \upsilon_q^2), \ \mu_q = \boldsymbol{\theta}' \boldsymbol{w}_q, \ \upsilon_q^2 = \boldsymbol{w}'_q \boldsymbol{\Xi} \boldsymbol{w}_q + 1. \tag{11}$$

Using the results in Theorem 3 from Section 2.2, the cumulative distribution function of g_q^* is:

$$H(z; \boldsymbol{d}_q, \boldsymbol{\Sigma}_q, \vartheta, \mu_q, \upsilon_q^2) = F_I\left[z\mathbf{1}_I; \left(\vartheta \boldsymbol{d}_q + \mu_q \mathbf{1}_I\right), \left(\vartheta^2 \boldsymbol{\Sigma}_q + \mathbf{IDEN}_I \upsilon_q^2\right)\right] \tag{12}$$

Finally, the likelihood function from the total count model, given that the observed count level of consumer q is n_q, may be written as:

$$L_{q,count}(\boldsymbol{b}, \boldsymbol{\Omega}, \boldsymbol{\Theta}, \boldsymbol{\theta}, \boldsymbol{\Xi}, \varphi, \vartheta) = H(\delta_{n_q}; \boldsymbol{d}_q, \boldsymbol{\Sigma}_q, \vartheta, \mu_q, \upsilon_q^2) - H(\delta_{n_q - 1}; \boldsymbol{d}_q, \boldsymbol{\Sigma}_q, \vartheta, \mu_q, \upsilon_q^2). \tag{13}$$

The likelihood function above involves the computation of an I-dimensional integral.

3.3. Estimation Technique

The overall likelihood function for the joint count-event type model may be obtained from Equations (5) and (13) as follows:

$$L_q(\boldsymbol{b}, \boldsymbol{\Omega}, \boldsymbol{\Theta}, \boldsymbol{\theta}, \boldsymbol{\Xi}, \varphi, \vartheta) = L_{q,event}(\boldsymbol{b}, \boldsymbol{\Omega}, \boldsymbol{\Theta}) \times L_{q,count}(\boldsymbol{b}, \boldsymbol{\Omega}, \boldsymbol{\Theta}, \boldsymbol{\theta}, \boldsymbol{\Xi}, \varphi, \vartheta)\,. \tag{14}$$

To address the issue of the high dimensionality of integration in $L_{q,event}$ (of dimension $n_q(I-1)$) in the above function, we replace the log-likelihood from the event model with a composite marginal likelihood (CML), $L_{CML,q,event}$. The CML approach has been proposed for and applied to various binary and ordered response model forms in the past (see Varin *et al.*, 2011 for a recent extensive review of CML methods; Lindsay *et al.*, 2011 and Yi *et al*., 2011 are also useful references), and Bhat (2011) extended it recently to unordered choice models. The CML approach, which belongs to the more general class of composite likelihood function approaches (see Lindsay, 1988), may be explained in a simple manner as follows. In the event type choice model, instead of developing the likelihood of the entire sequence of repeated choices from the same consumer, consider developing a surrogate likelihood function that is the product of the probability of easily computed marginal events. For instance, one may compound (multiply) pairwise probabilities of a consumer *q* choosing alternative *i* at time *t* and choosing alternative i' at time t', of the consumer *q* choosing alternative *i* at time *t* and choosing alternative i'' at time t'', and so forth. The CML estimator (in this instance, the pairwise CML estimator) is then the one that maximizes the compounded probability of all pairwise events. The properties of the CML estimator may be derived using the theory of estimating equations (see Cox and Reid, 2004, Yi *et al*., 2011). Specifically, under usual regularity assumptions (Molenberghs and Verbeke, 2005, page 191, Xu and Reid, 2011), the CML estimator is consistent and asymptotically normal distributed, and its covariance matrix is given by the inverse of Godambe's (1960) sandwich information matrix (see Zhao and Joe, 2005).

Letting the individual *q*'s choice at time *t* be denoted by the index C_{qt}, the CML function for the event type choice model for consumer *q* may be written as:

$$
\begin{aligned}
L_{CML,q,event} &= \prod_{t=1}^{n_q-1}\prod_{t'=t+1}^{n_q} Prob(C_{qt}=m_{qt}, C_{qt'}=m_{qt'}) \\
&= \prod_{t=1}^{n_q-1}\prod_{t'=t+1}^{n_q} Prob(\boldsymbol{u}^*_{qt}<0 \text{ and } \boldsymbol{u}^*_{qt'}<0) = \prod_{t=1}^{n_q-1}\prod_{t'=t+1}^{n_q} Prob(\vec{\boldsymbol{u}}^*_{qtt'}<0)
\end{aligned}
\quad (15)
$$

where $\vec{\boldsymbol{u}}^*_{qtt'} = \left[\left(\boldsymbol{u}^*_{qt}\right)', \left(\boldsymbol{u}^*_{qt'}\right)'\right]'$. Then,

$$
P(\vec{\boldsymbol{u}}^*_{qtt'}<0) = \Phi_{2\times(I-1)}\left((\vec{\boldsymbol{\omega}}_{\mathbf{F}_q tt'})^{-1}(-\vec{\boldsymbol{H}}_{\boldsymbol{qtt'}}); (\vec{\boldsymbol{\omega}}_{\mathbf{F}_q tt'})^{-1}\mathbf{F}_{qtt'}(\vec{\boldsymbol{\omega}}_{\mathbf{F}_q tt'})^{-1}\right), \quad (16)
$$

where $\vec{\boldsymbol{H}}_{\boldsymbol{qtt'}} = (\boldsymbol{H}'_{\boldsymbol{qt}}, \boldsymbol{H}'_{\boldsymbol{qt'}})'$, $\mathbf{F}_{qtt'}$ is the 2×2-sub-matrix of $\mathbf{F}_q$ that includes elements corresponding to the t^{th} and $t'^{\text{ th}}$ choice occasions of individual q, and $\vec{\boldsymbol{\omega}}_{\mathbf{F}_q tt'}$ is the diagonal matrix of the standard deviations of $\mathbf{F}_{qtt'}$. Finally, the function to be maximized to obtain the parameters is:

$$
L_{CML,q}(\boldsymbol{b},\boldsymbol{\Omega},\boldsymbol{\Theta},\boldsymbol{\theta},\boldsymbol{\Xi},\varphi,\vartheta) = L_{CML,q,event}(\boldsymbol{b},\boldsymbol{\Omega},\boldsymbol{\Theta}) \times L_{q,count}(\boldsymbol{b},\boldsymbol{\Omega},\boldsymbol{\Theta},\boldsymbol{\theta},\boldsymbol{\Xi},\varphi,\vartheta). \quad (17)
$$

The $L_{CML,q,event}$ component in the equation above entails the evaluation of a multivariate normal cumulative distribution (MVNCD) function of dimension equal to $[(I-1)\times 2]$, while the $L_{q,count}$ component involves the evaluation of a MVNCD function of dimension I. But these may be evaluated using the approximation part of the maximum approximate composite marginal likelihood (MACML) approach of Bhat (2011), leading to solely bivariate and univariate cumulative normal function evaluations.

One additional issue still needs to be dealt with. This concerns the positive definiteness of several matrices in Equation (17). Specifically, for the estimation to work, we need to ensure the positive definiteness of the following matrices: $\boldsymbol{\Omega}$, $\boldsymbol{\Theta}$, and $\boldsymbol{\Xi}$. This can be guaranteed in a straightforward fashion using a Cholesky decomposition approach (by parameterizing the function in Equation (17) in terms of the Cholesky-decomposed parameters).

4. AN EMPIRICAL APPLICATION TO WEEKDAY NON-WORK ACTIVITY EPISODE GENERATION AND SCHEDULING

4.1. Background

The joint count-event type choice model proposed in this paper can be used in a wide variety of multivariate count data settings. In the current research, we demonstrate an application to examine the total number of out-of-home non-work episodes pursued by a worker and the organization of these episodes across five time-of-day blocks. The time-of-day blocks are defined based on the worker's schedule, recognizing that the work activity tends to be a "peg" around which other activities typically get scheduled (see Damm, 1980, Rajagopalan *et al.*, 2009). The five time-of-day blocks are as follows:

- Before-work (BW), representing the time from 3 AM in the morning to the individual's departure from home on the first home-to-work trip in the day.
- During home-to-work commute (HWC), representing the time between the individual's departure from home on her/his first home-to-work trip in the day to the individual's arrival time at work at the end of this first home-to-work trip (for presentation ease, we will refer to this latter clock time as the work start time of the individual).
- Work-based (WB), representing the time between the individual's work start time to the individual's departure time from work on the last trip of the day from work-toward home (we will refer to this departure time as the work end time of the individual).
- During work-to-home commute (WHC), representing the period between the individual's work end time to the arrival time at home at the end of the chain of trips that began at work at the work start time (we will label this arrival time at home as the home arrival time).
- After home arrival from work (AH), representing the period from the home arrival time to 3AM the next day.

The joint model of total non-work episodes and organization in the five time blocks identified above can provide important insights for travel demand forecasting and policy analysis (see Damm, 1980 and McGuckin *et al*., 2005). More broadly, modeling the organization of episodes to different time blocks allows the generation of activity episode demand for different times of the day, and can capture interaction effects in episode generation across different time blocks. Besides, the total count of non-work episodes, and the variability in this count across the population, is important in its own right, given that increases in overall travel demand in recent years may be largely attributed to non-work travel growth, with corresponding greenhouse gas emissions implications and global climate change effects. It is therefore no surprise that there has been an explosion in studies in the past few years on studying non-work-related activity participation and scheduling, and the interlinking of non-work participation with work activity, to inform the subsequent and finer modeling of activity purpose/types, time allocation, destinations, and travel mileages within each time-of-day block defined in relation to the work schedule (see, for example, Jou *et al.*, 2010, Castro *et al.*, 2011, Van Acker and Witlox, 2011, and Lachapelle and Noland, 2012).

4.2. Data Source and Sample Description

The data used in this study is derived from the 2009 National Household Travel Survey (NHTS) conducted in the United States, which collected information on more than one million trips undertaken by 320,000 individuals from 150,000 households sampled from all over the country for one day of the week. The survey also collected detailed information on individual and household socio-demographic and employment-related characteristics. For this study, we employed the NHTS California add-on dataset for the Southern California (SC) region comprising Imperial, Los Angeles, Orange, Riverside, San Bernardino and Ventura counties. The SC region was chosen because the California add-on dataset has geocoded home and work location Census tract information, and because the research team has detailed accessibility measures computed at the census tract level by time of day for the SC region.[6] The accessibility measures are opportunity-based indicators that measure the number of activity opportunities by fifteen different industry types (such as agriculture, construction, manufacturing, and entertainment) that can be reached within 20 minutes from each Census tract during each of four time periods: (1) morning-peak period (6am-9am), (2) off-peak period (9am-3pm), (3) afternoon-peak period (3pm-7pm), and (4) night-time period (7pm-6am).

The sample formation included several steps. First, only individuals over 18 years of age, and who participated in at least one work activity episode during the survey day on a weekday (Monday to Friday), were selected. Second, we eliminated individuals whose trip diary did not start or end at home. Third, records that contained incomplete information on individual, household, employment-related, and activity and travel characteristics of relevance to the current analysis were removed from the sample. Fourth, several consistency checks were performed and records with missing or inconsistent data were eliminated. The final estimation sample contained

[6] These accessibility measures were computed by Prof. Konstadinos Goulias's research group at the University of California at Santa Barbara. The reader is referred to Chen *et al.* (2011) for details of the construction of these Census tract-based accessibility measures. The time-of-day variation arises because the number of activity opportunities varies based on the open/closed-times of the activity centers and also because the number of people employed varies by time of day. Thus, the accessibility for entertainment opportunities (arts, entertainment, recreation, accommodation and food services) would see an increase in the evening periods compared to the morning periods.

2,113 person observations. Fifth the trip diaries of these 2,113 individuals were processed to obtain, for each individual, the total number of out-of-home non-work episodes undertaken during the survey day, along with the number of these episodes pursued during each of the five time-of-day blocks identified in Section 4.1. Finally, the accessibility measures by the fifteen different industry types were appended to each time-of-day block for each individual as follows. For the before-work (BW) block, the accessibility measures (by industry type) are based off the time the individual would have had to leave home if s/he went directly to work (computed as the individual's work start time minus the estimated direct home-to-work commute time assuming auto mode of travel and an average speed of 30 mph). That is, the accessibility measures corresponding to the individual's estimated departure time from home to work (assuming a direct home-to-work trip) and for the residential Census tract of the individual are designated as the home end accessibilities for the BW block. For the home-to-work commute (HWC) block, the accessibility measures are based off the individual's work start time. For this block, we create two sets of accessibility measures, one for the home end (based on the Census tract of residence) and another for the work end (based on the Census tract of the individual's workplace location). For the work-based (WB) block, the accessibility measures are based on the off-peak period for the work location Census tract. For the work-to-home commute (WHC) block, the accessibility measures are based off the individual's work end time. For this block, we once again create both a home end set of accessibilities as well as a work end set of accessibilities. For the after home arrival from work (AH) block, the accessibilities are based off the time the individual would have arrived home if s/he went directly back home from work (computed as the individual's work end time plus the estimated direct work-to-home commute time assuming auto mode of travel and an average speed of 30 mph). That is, the accessibility measures corresponding to the estimated arrival time back home and for the residential Census tract of the individual (assuming a direct work-to-home trip) are designated as the home end accessibilities.

Table 1 provides a summary of select individual, household, work-related and activity and travel characteristics of the final sample. Among *individual characteristics*, Table 1 reveals a high percentage of non-Hispanic Caucasian workers (almost 72%), a higher proportion of men than women, a vast majority of individuals with a driver's license, a highly educated sample, and work being characterized as the primary activity in the past week for most individuals (as opposed to non-work activities such as vacation, studying, shopping and recreation), and about

43% of individuals shopping over the internet in the past month. The descriptive statistics of age in the middle panel of the table indicate an average age of about 47 years, with a minimum of 18 years and a maximum of 86 years. In the category of *household characteristics*, the table shows a rather high household income in the sample (compared to the overall California population), a high percentage of individuals residing in an urban cluster, an average household size of 3.14 individuals (an average of 2.4 adults and 0.74 non-adults per household), about the same average household number of drivers and vehicles as the number of adults, and an average of 1.84 workers per household. *Work-related characteristics* capture the nature of work schedules and the flexibility associated with the schedules. Table 1 shows that almost half of the individuals work in professional, managerial or technical jobs, with less than 10% of the workforce being self-employed and holding more than one job. Also, less than 45% of individuals have a flexible work schedule and about 13% have the option to telecommute. The average distance to work (see the middle panel) is 13.52 miles, with a variation about equal to the mean. The *activity and travel characteristics* of individuals in the sample reveal a slightly less than the expected fifth of individuals completing their survey on a Friday, a small fraction using public transportation on the survey day or bicycling in the week prior to the survey day, and a rather high percentage of workers who pursued at least one trip completely by walk.

The bottom panel of Table 1 shows the sample statistics of the number of out-of-home non-work episodes by time-of-day blocks, which is the dependent variable in the current empirical analysis. The statistics clearly reveal the higher inclination to undertake non-work activities after work (during the WHC and AW blocks). This is consistent with the findings from earlier literature (Strathman and Dueker, 1995; Bhat and Sardesai, 2006). On average, workers participate in 1.54 non-work activities per day, with a standard deviation which is larger than the mean value (see the last row).

4.3. Estimation Results

4.3.1. Variable Specification

The selection of variables included in the final model specification in Table 2 was based on previous research, intuitiveness, and parsimony considerations. For categorical exogenous variables, if a certain level of the variable did not have sufficient observations, it was combined with another appropriate level; and if two levels had similar effects, they were combined into one level. For continuous variables, we tested alternative linear and non-linear functional forms, including dummy variables for different ranges. The exogenous variables described in Section 4.2 were considered both in the count model specification (threshold and long-term propensity) and in the event type choice model specification, except for the time of day block-specific accessibility measures that were introduced in the time-of-day block choice (*i.e.*, event type) model. The accessibility measures constructed at the home end were used in the BW, HWC, WHC and AH blocks, while the accessibility measures constructed at the workplace end were used in the HWC, WB, and WHC blocks.

The final estimation results are presented in Table 2 (for the count data model component) and Table 3 (for the event type choice model component). In some cases, we have retained variables that are not statistically significant at a 0.05 significance level because of their intuitive effects and to inform future research efforts in the field.

4.3.2. Count data model component

The first main numeric column of Table 2 provides the coefficients associated with the latent propensity, while the second main numeric column presents the threshold coefficients. In these tables, for categorical variables, the base category is presented in parenthesis. For example, for the "race and ethnicity" variables, the base category is "non-Hispanic and non-Asian". Also, a positive sign for a latent propensity coefficient indicates that an increase in the corresponding variable results in an increased propensity to undertake non-work activity episodes, while a negative sign indicates the reverse. For the threshold variables, a positive coefficient shifts the threshold toward the left of the propensity scale, which has the effect of reducing the probability of the zero-trip outcome (increasing the overall probability of the non-zero outcome). A negative coefficient, on the other hand, shifts the threshold toward the right of the propensity scale, which has the effect of increasing the probability of the zero-trip outcome (decreasing the overall probability of the non-zero outcome; see CPB).

The first row panel in Table 2 presents the constant in the φ vector, as well as the threshold-specific constants (α_k values). These constants do not have any substantive interpretations, though the threshold specific constants (α_k) provide flexibility in the count model to accommodate high or low probability masses for specific outcomes. As indicated in

Section 3.2.1, identification is achieved by specifying $\alpha_0 = 0$ and $\alpha_k = \alpha_K \; \forall \; k \geq K$. In the present specification, we initially set *K* = 13 (which is the maximum value of the total number of non-work episodes in the sample) and progressively reduced *K* based on statistical significance considerations and general data fit. We also combined the threshold constants when they were not statistically significantly different to gain estimation efficiency. The final specification in Table 2 is based on setting *K* = 6.

The next row panel of Table 2 provides the effects of individual characteristics. Hispanic and non-Hispanic Asians are less likely to pursue non-work episodes during the day relative to other race-ethnicity groups (primarily dominated by non-Hispanic Caucasians). Women, on average, pursue more non-work episodes than males, a consistent finding in the literature attributable to the typically larger role played by women in maintenance, shopping, and serve-passenger activities (see Crane and Takahashi, 2009 and Bernardo *et al*., 2012). However, there is substantial variation in this gender effect, as evidenced by the large standard deviation estimate on the female dummy variable. The mean and standard deviation estimates indicate that about 60% of employed women participate in more non-work activities than their male counterparts, while 40% of employed women participate in less activities that their male counterparts. Individuals who characterized their primary activity last week as being non-work related have a higher non-work episode making propensity, as expected, while the internet shopping variable indicates complementarity between internet shopping and in-person shopping out-of-home (see Bhat *et al*., 2003 and Farag, 2006 for a similar result).

Among *household characteristics*, individuals whose home location is not in an urban cluster are less inclined to undertake non-work activities. The household composition effects are interesting, and reflect the higher levels of in-home activity participation and/or economies of scale in non-work participation when there are multiple adults in the household. Also, on average, a higher number of non-adults in the household leads to higher shopping and care-related needs of non-adults (see McDonald, 2008), as evidenced by the positive sign on the mean effect of this variable. However, there is also substantial variation in the magnitude of this effect, though the positive sign is retained for almost all individuals. The number of workers in the employee's household is found to positively influence non-work episode frequency through the threshold specification that governs the "instantaneous" translation of the non-work participation propensity to whether or not a non-work episode is participated on any given day. This positive effect is a reflection perhaps of spontaneous non-work stops by employed individuals made during the work commute.

In the category of w*ork-related characteristics,* self-employed workers have a higher propensity to participate in non-work episodes relative to those not self-employed, while those who have the option to work from home make more spontaneous non-work stops than those who do not have the option to work from home. The former result is suggestive of the overall flexibility enjoyed by those who are self-employed, while the latter result may be an indication of the "on-the-spur" decision-making ability of those who work from home. Workers with multiple jobs have a higher propensity to make non-work stops, perhaps a reflection of juggling tasks and having many non-work responsibilities (see Dickey *et al*., 2011 and Khan *et al.*, 2012). In addition, those with long commutes have less time for non-work activity participation than those with short commutes, which may explain the negative sign on the "distance to work" variable (see also Lyons and Chatterjee, 2008 and Sandow, 2011 for a similar result).

The effects of the *mobility and situational characteristics* are also reasonable. Employed individuals who use some form of public transportation on the survey day have a lower non-work participation propensity than other individuals, possibly due to schedule inflexibility and less time available for non-work participation among those who use public transportation. Also, workers who walked or biked at least once in the past week are more likely to undertake non-work episodes, a result that can be associated with the active life style of individuals who use non-motorized modes (Merom *et al*., 2010 also observe this result).

Finally, the parameter that links the event type choice model with the count model in our final model specification is highly statistically significant, supporting the hypothesis that workers jointly decide the frequency of non-work activities (count model) and the organization of these activities across time-of-day blocks (event type choice model). That is, the total count of non-work episodes is endogenous to the time-of-day participation in the episodes, and variables that affect the time-of-day of participation also impact the total count of episodes.

4.3.3. Time-of-day block (i.e., event type) choice model component

Table 3 presents the results of the time-of-day block choice model component. The first row panel of Table 3 presents the alternate specific constants, with the base alternative being the before-work (BW) block. These constants do not have any substantive interpretation because of the presence of continuous explanatory variables (the accessibility measures). However, several of these constants have a significant standard deviation, indicating individual-specific heterogeneity in the preferences for the time-of-day alternatives for non-work episode participation.

The *accessibility measures* by industry type and time block are significant determinants of time-of-day block, both at the home end and the work end. In general, workers are less likely to participate in non-work episodes during time blocks when their homes/work locations are highly accessible to traditionally work-focused industry centers (such as natural resources, manufacturing, information, financial services, and educational services), and more likely, in general, to participate in non-work episodes during time blocks when their home/work locations are highly accessible to service and entertainment related industry opportunities (wholesale trade, health, and entertainment). The significant standard deviation on the entertainment accessibility indicates variation in this effect, though the mean and standard deviation estimates imply an increase in entertainment accessibility in a specific time-of-day block increases non-work activity participation in the time block for over 92% of employed individuals. The results also indicate the marginally higher propensity of women to participate in non-work episodes during time blocks that have a high accessibility to retail trade, a finding consistent with the higher shopping tendency of women relative to men (Brunow and Gründer, 2012).

In the category of *work-related characteristics*, self-employed workers are more likely to participate in non-work activity episodes during the WB block and less likely to participate during the WHC block. This is intuitive, given the independence and flexibility offered by self-employment during the WB period, and the consequent reduction in WHC (van Ommeren and van der Straaten, 2008). The finding that workers who have a flexible work start time have a lower propensity (than those with rigid work start times) to undertake non-work episodes in the BW block is interesting, and needs further exploration.

Within the category of *mobility and situational characteristics*, workers are more likely to pursue non-work episodes during the WHC and AH blocks on Fridays than on other weekdays, highlighting the spike in social-recreational activity pursuits on Friday evenings (Stone *et al.*, 2012). Workers who use public transportation on the survey day are less likely to participate in non-work activities in the BW block, presumably because of difficulty in coordinating non-work activities with the public transportation schedules and the work start time.

As described in Section 3.1, we optimize the likelihood function with respect to the elements of the differenced covariance matrix $\mathbf{\Theta}$ during model estimation. However, the elements of the differenced covariance matrix are not intuitive and cannot be interpreted directly. To make meaningful inferences, it is essential to impute the dependencies between utilities of alternatives directly. So, we constructed an equivalent un-differenced covariance matrix which results in the differenced covariance matrix that we obtained at the end of the model estimation process (this final specification of the differenced covariance matrix was a restrictive version of the fully free differenced covariance matrix with the single scale restriction; the restrictive version provided as good a fit, from a statistical standpoint, as the fully free covariance matrix). Table 4 presents the estimation results corresponding to the equivalent un-differenced covariance matrix of the type-of-day block choice model component. It can be seen from the table that only two elements are significant from their corresponding values in an independent MNP model at 95% confidence level. All the remaining elements are fixed as shown in the table (the diagonal elements of the covariance matrix are fixed to 0.5 while the off-diagonal elements are fixed to zero). We found that there is high positive covariance in the unobserved factors affecting the WHC and AW time-of-day blocks. This suggests that there are common unobserved factors which simultaneously increase (decrease) the utility associated with these two time-of-day blocks. This is intuitive given that there are no rigid space and time constraints after the end of work (such as fixed work start time, minimum work hours, and presence at the work place) resulting in considerable available time for activity participation during both WHC and AW time-of-day blocks. It is also possible that the evening time after work is perceived to be more conducive for participating in several out-of-home activities (including shopping, dining, and recreation) with family and friends. The magnitude of the variance element corresponding to the AH time-of-day block is 0.5695 and is significantly different from 0.5, indicating larger variability in the unobserved factors impacting the utility associated with AH time-of-day block compared to other time-of-day blocks.

4.4. Model Fit

The composite log-likelihood (CL) measure of the model system proposed in this paper that retains the linkage between the total count model and the event type model (the joint model) is $-14{,}441.3$ with 50 parameters. The corresponding figure for the model system that unlinks the total count model and the event type model (the independent model) is $-14{,}488.8$ with 49 parameters. These CL measures can be statistically compared by computing the adjusted composite likelihood ratio test (*ADCLRT*) statistic, which serves the same role as the likelihood ratio test in traditional maximum likelihood estimation (see Pace *et al.*, 2011 and Bhat, 2011 for details of the computation of this *ADCLRT* statistic). This *ADCLRT* statistic returns a value of 66.23, which is larger than the table chi-squared value with one degree of freedom at any reasonable level of significance.

The model fit of our proposed model can also be evaluated using other more intuitive measures by obtaining predictive distributions. To do so, we first define $\mathbf{R}_i$ $(i = 1,2,...,I)$ as an $(I-1)\times I$ matrix that corresponds to an $(I-1)$ identity matrix with an extra column of -1's added as the i^{th} column. Following the notation in Equation (10) and immediately after, define $\mathbf{G}_{qi} = \mathbf{R}_i \mathbf{\Sigma}_q \mathbf{R}'_i$. We can then write the probability that individual (consumer) q chooses alternative i at any choice occasion as:

$$P_{qi} = P[\boldsymbol{M}_{qi}\breve{\boldsymbol{U}}_q < \boldsymbol{0}_{I-1}] = \Phi_{(I-1)}\left[(\boldsymbol{\omega}_{\mathbf{G}_q})^{-1}(-\boldsymbol{d}_q), (\boldsymbol{\omega}_{\mathbf{G}_q})^{-1}\mathbf{G}_{qi}(\boldsymbol{\omega}_{\mathbf{G}_q})^{-1}\right]. \tag{18}$$

Next, since this probability does not change across choice occasions, and the individual-specific preferences are already embedded in $\breve{\boldsymbol{U}}_q$ (through the $\boldsymbol{\beta}_q$ vector), the multivariate probability of counts in each time-of-day block (*i.e.*, event type), conditional on the total count level k_q ($k_q > 0$), takes the usual multinomial distribution form:

$$P[(g_{q1} = k_{q1}), (g_{q2} = k_2), ...(g_{qI} = k_{qI}) \mid k_q] = \frac{k_q!}{\prod_{i=1}^{I} k_{qi}!} \prod_{i=1}^{I} (P_{qi})^{k_{qi}}. \tag{19}$$

In our joint model of multivariate counts, the unconditional multivariate probability then takes the form indicated below ($k_q = \sum_{i=1}^{I} k_{qi}$, $k_{qi} = 0,1,2,...,\infty$, $k_q = 0,1,2,...,\infty$):

$$P[(g_{q1} = k_{q1}), (g_{q2} = k_2), ...(g_{qI} = k_{qI})] = P[g_q = k_q] \times \left(\frac{k_q!}{\prod_{i=1}^{I} k_{qi}!} \prod_{i=1}^{I} (P_{qi})^{k_{qi}} \right), \tag{20}$$

with $P[g_q = k_q]$ as in Equation (13) after replacing n_q (the actual observed total count for individual q in the estimation sample) with an arbitrary value k_q. Using the properties of the multinomial distribution, the marginal probability of k_{qi} counts for time-of-day block i is:

$$P[g_{qi} = k_{qi}] = \sum_{k_q=0}^{\infty} \left[P[g_q = k_q] \times \left(\frac{k_q!}{k_{qi}!(k_q - k_{qi})!} (P_{qi})^{k_{qi}} (1 - P_{qi})^{(k_q - k_{qi})} \right) \right] \tag{21}$$

In the above expression, the upper bound of the summation is $k_q = \infty$, though the probability values fade very rapidly beyond a k_q value of 10. For the purposes of this paper, we carry the summation up to $k_q = 50$.

With the above preliminaries, the model predictions can be used to evaluate data fit at both the disaggregate and aggregate levels, as well as for both the multivariate count distribution and the marginal count distribution. At the disaggregate level, we estimate the probability of the observed multivariate count outcome for each individual using Equation (20), and compute an average probability of correct prediction. Similarly, we also estimate the probability of the observed marginal count outcome separately for each time-of-day period using Equation (21), and compute an average probability of correct prediction. At the aggregate level, we design a heuristic diagnostic check of model fit by computing the predicted aggregate share of individuals for specific multivariate outcome cases (because it would be infeasible to provide this information for each possible multivariate outcome). In particular, we compute the aggregate share of consumers for each of six combinations. The first combination corresponds to no participation in any non-work episodes (which we will refer to as the "no participation" combination). The other five combinations correspond to participation in one or more episodes during a specific time-of-day block and no participation in any other time-of-day period (which we will refer to using such labels as the "BW participation only" combination or the "HWC participation only" combination). In addition to these aggregate shares of multivariate outcomes, we also compute the aggregate shares of the marginal outcomes of count values of 0, 1, 2, 3, and 4+ for each time-of-day period, as well as for the total count. As a yardstick to evaluate the performance of the joint model proposed here, we compare the predictions from the joint model with the independent model using the absolute percentage error (APE) statistic for each count value, and then compute a mean weighted APE value across the count values (of 0 1, 2, 3, and 4+) using the observed number for each count value as the weight for that count value. .

The disaggregate-level data fit measures indicate an average probability of correct prediction of 13.9% for the multivariate counts and an average probability of correct prediction of 67.6% for the marginal counts. The corresponding values for the independent model are 13.6% and 65.0%, respectively, which are smaller in magnitude than those from the joint model. The aggregate fit measures are provided in Table 5. The joint model provides a better (lower) APE value for all the multivariate outcomes in Table 5 (see upper panel of the table), except for the WB participation only outcome. The APE values are sizeable for both the joint and independent values, but it should be noted that these predictions are for multivariate outcomes. Overall, the mean weighted APE value is about 12% higher for the independent model relative to the joint model. As expected, the APE values are lower for the marginal outcomes (see lower panel of Table 5) than for the multivariate outcomes. The total count predictions from the joint model are much better than the total count predictions from the independent model. Also, the predictions for the other marginal counts are better from the joint model relative to the independent model (except for the WB block count). These results clearly show that the joint model proposed here outperforms the traditional independent model in the disaggregate level and aggregate level comparisons.

4.5. Model Application

The joint model estimated in this paper can be used to examine the impact of changes in socio-demographic characteristics over time as well as the effects of policy actions that involve a change in the accessibility measures and work-related characteristics. In this paper, we demonstrate the application of this model by studying the effects of changes in three selected variables: distance to workplace, retail trade accessibility at the home location, and entertainment accessibility at the home location. These three variables are increased by 20% across all workers. The impact on the frequency and organization of non-work activities is estimated by determining the percentage change in the expected number of non-work episodes (across all workers) for the entire day (*i.e.*, total count) and for each time-of-day block. To demonstrate the potentially misleading inferences from the independent model, we compute the change as predicted by both the joint model as well as the independent model. The emphasis here is not on substantive empirical inferences as much as it is on demonstrating the differences in the inferences from the two models. Table 6 provides the results.

Three observations may be made from Table 6. First, in the independent model, a change in the retail trade and entertainment accessibility variables do not have any impact on the total count of non-work episodes over the entire day. This is, of course, because these variables appear only in the event discrete choice model and not the total count model (and the independent model does not have any link between the discrete choice model and the total count model). As indicated earlier in the paper, it is natural to expect that changes in the attributes impacting the attractiveness of alternatives in the choice model (retail trade and entertainment accessibilities in the specific case under discussion) will result not only in substitution among the counts of each discrete choice alternative, but also an overall change in the total count, as appropriately recognized by the joint model. Second, the positive effect (of an increase in the number of retail trade jobs) on the number of non-work tours during the BW, HWC, WHC, and AH periods is underestimated by the independent model, while the negative effect of the variable on WB non-work tours is overestimated by the independent model. Third, a similar result holds also for the influence of the number of entertainment jobs at the home location. Indeed, for this variable, the directionality of the effect on WHC non-work tours is itself different between the independent and joint models. These differences between the models highlight the potentially misinformed policy analyses that result from ignoring the joint nature of frequency of non-work episodes and their organization across time-of-day blocks.

5. CONCLUSIONS

Count data models are used in several disciplines to analyze discrete and non-negative outcomes, but their implementation has been mostly restricted to univariate or bivariate count systems. In the current paper, we develop an approach to multivariate count data modeling that combines a total count model with a discrete choice model for event choice that allocates the total count to different events. While previous studies have used such an approach, most do not consider the linkage between the event choice and the total count. And those that consider this linkage do so by including the expected value of the maximum utility from the event type multinomial model as an explanatory variable in the conditional expectation for the total count random variable.

They ignore the effect of the event choice errors on the total count, which is critical to recognize the full econometric jointness of the two decisions

In the current paper, we have proposed a joint model of total count and event type choice for multivariate count data analysis that (a) uses a flexible MNP structure for the event type choice, (b) develops and uses new results regarding the distribution of the maximum of multivariate normally distributed random variables (with a general covariance matrix) as well as its stochastic affine transformations, and (c) employs a latent variable framework for modeling the total count variable that, at once, enables the linkage of the event type choice and total count, recognizes the presence of unobserved individual-specific preference and taste variations, and accommodates excess zeros (or excess number of any count value for that matter) without the need for zero-inflated or hurdle devices.

The modeling framework is applied to examine the total number of out-of-home non-work episodes pursued by a worker and the organization of these episodes across five time-of-day blocks. The data used is derived from the 2009 National Household Travel Survey (NHTS) for the South California region. The results show the importance of recognizing the joint nature of total count and event type choice decisions, from both a data fit perspective as well as for forecasting and policy analysis.

ACKNOWLEDGEMENTS

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

REFERENCES

Arellano-Valle, R.B., Genton, M.G., 2008. On the exact distribution of the maximum of absolutely continuous dependent random variables. *Statistics & Probability Letters* 78(1), 27-35.

Awondo, S., Egan, K.J., Dwyer, D.F., 2011. Increasing beach recreation benefits by using wetlands to reduce contamination. *Marine Resource Economics* 26(1), 1-15.

Azzalini, A., Regoli, G., 2012. Some properties of skew-symmetric distributions. *Annals of the Institute of Statistical Mathematics* 64(4), 857-879.

Bermúdez, L., Karlis, L., 2011. Bayesian multivariate Poisson models for insurance ratemaking. *Insurance: Mathematics and Economics* 48(2), 226-236.

Bernardo, C., Paleti, R., Hoklas, M., Bhat, C.R., 2012. An empirical investigation into the time-use and activity patterns of dual-earner couples with and without young children. Technical paper, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin.

Bethell, J., Rhodes, A.E., Bondy, S.J., Lou, Wendy, W.Y., Guttmann, A., 2010. Repeat self-harm: Application of hurdle models. *The British Journal of Psychiatry* 196(3), 243-244.

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

Bhat, C.R., Sivakumar, A., Axhausen, K.W., 2003. An analysis of the impact of information and communication technologies on non-maintenance shopping activities. *Transportation Research Part B* 37(10), 857-881.

Bose, R.C., Gupta, S.S., 1959. Moments of order statistics from a normal population. *Biometrika* 46(3-4), 433-440.

Brunow, S., Gründer, M., 2012. The impact of activity chaining on the duration of daily activities. *Transportation*, forthcoming.

Buck, A., Blackstone, E., Hakim, S., 2009. A multivariate poisson model of consumer choice in a multi-airport region. *iBusiness* 1(2), 85-98.

Burda, M., Harding, M., Hausman, J.A., 2012. A Poisson mixture model of discrete choice. *Journal of Econometrics* 166(2), 184-203.

Cameron, A.C., Trivedi, P.K., 2013. Regression analysis of count data, 2nd edition. *Econometric Society Monograph*, Cambridge University Press, forthcoming.

Castro, M., Eluru, N., Bhat, C.R., Pendyala, R.M., 2011. A joint model of participation in non-work activities and time-of-day choice set formation for workers. *Transportation Research Record* 2254, 140-150.

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

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

Chib, S., Winkelmann, R., 2001. Markov chain Monte Carlo analysis of correlated count data. *Journal of Business and Economic Statistics* 19(4), 428-435.

Clark, C., 1961. The greatest of a finite set of random variables. *Operations Research* 9(2), 145-162.

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

Crane, R., Takahashi, L., 2009. Sex changes everything: The recent narrowing and widening of travel differences by gender. *Public Works Management and Policy* 13(4), 328–337.

Damm, D. 1980. Interdependencies in activity behavior. *Transportation Research Record* 750, 33-40.

David, H.A., 1981. *Order statistics*. 2nd edition. Wiley, New York.

Devroye, L., 1980. A note on finding convex hulls via maximal vectors. *Information Processing Letters* 11(1), 53-56.

Dickey, H., Watson, V., Zangelidis, A., 2011. Is it all about money? An examination of the motives behind moonlighting. *Applied Economics* 43(26), 3767-3774.

Farag, S., 2006. E-shopping and its interactions with in-store shopping. Ph.D. Thesis, Department of Human Geography and Planning, Faculty of Geosciences, Utrecht University.

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

Greene, W.H., Hensher, D.A., 2010. *Modeling ordered choices: A primer*. Cambridge University Press, Cambridge.

Greig, M., 1967. Extremes in a random assembly. *Biometrika* 54(1-2), 273-282.

Gumbel, E. J., 1958. *Statistic of extremes*. Columbia University Press, New York.

Gupta, A.K., González-Farías, G., Domínguez-Monila, J.A., 2004. A multivariate skew normal distribution. *Journal of Multivariate Analysis* 89(1), 181-190.

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

Haque, M., Chin, H., Huang, H., 2010. Applying Bayesian hierarchical models to examine motorcycle crashes at signalized intersections. *Accident Analysis and Prevention* 42(1), 203-212.

Hausman, J.A., Leonard, G.K., McFadden, D., 1995. A utility-consistent, combined discrete choice and count data model: Assessing recreational use losses due to natural resource damage. *Journal of Public Economics* 56(1), 1-30.

Herriges, J.A., Phaneuf, D.J., Tobias, J.L., 2008. Estimating demand systems when outcomes are correlated counts. *Journal of Econometrics* 147(2), 282-298.

Jamalizadeh, A., Balakrishnan, N., 2009. Order statistics from trivariate normal and t-distributions in terms of generalized skew-normal and skew-t-distributions. *Journal of Statistical Planning and Inference* 139(11), 3799-3819.

Jamalizadeh, A., Balakrishnan, N., 2010. Distributions of order statistics and linear combinations of order statistics from an elliptical distribution as mixtures of unified skew-elliptical distributions. *Journal of Multivariate Analysis* 101(6), 1412-1427.

Jou, R-C., Hensher, D.A., Liu, Y., Chiu, C-S., 2010. Urban commuters' mode-switching behaviour in Taipai, with an application of the bounded rationality principle. *Urban Studies* 47(3), 650-665.

Khan, M., Paleti, R., Bhat, C.R., Pendyala, R.M., 2012. A joint household level analysis of work arrangement choices of individuals. *Transportation Research Record*, forthcoming.

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

Lachapelle, U., Noland, R.B., 2012. Does the commute mode affect the frequency of walking behavior? The public transit link. *Transport Policy* 21, 26-36.

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

Lindsay, B.G., Yi, G.Y., Sun, J., 2011. Issues and strategies in the selection of composite likelihoods. *Statistica Sinica* 21(1), 71-105.

Lyons, G., Chatterjee, K., 2008. A human perspective on the daily commute: Costs, benefits and trade-offs. *Transport Reviews* 28 (2), 181-198.

Mannering, F.L., Hamed, M., 1990. Occurrence, frequency and duration of commuters' work-to-home departure delay. *Transportation Research Part B* 24(2), 99-109.

Marsaglia, G., 1963. Expressing the normal distribution with covariance matrix A + B in terms of one with covariance matrix A. *Biometrika* 50(3-4) 535–538.

McDonald, N.C., 2008. Household interactions and children's school travel: The effect of parental work patterns on walking and biking to school. *Journal of Transport Geography* 16(5), 324 -331.

McGuckin, N., Zmud, J., Nakamoto, Y., 2005. Trip-chaining trends in the United States: Understanding travel behavior for policy making. *Transportation Research Record* 1917, 199-204.

Merom, D., van der Ploeg, H.P., Corpuz, G., Bauman, A.E., 2010. Public health perspectives on household travel surveys active travel between 1997 and 2007. *American Journal of Preventive Medicine* 39(2), 113-121.

Molenberghs, G., Verbeke, G., 2005. *Models for discrete longitudinal data*. Springer Series in Statistics, Springer Science + Business Media, Inc., New York.

Musio, M., Sauleau, E., Buemi, A., 2010. Bayesian semi-parametric ZIP models with space-time interactions: An application to cancer registry data. *Mathematical Medicine and Biology* 27(2), 181-194.

Naya, H., Urioste, J.I., Chang, Y.M., Rodriguez-Motta, M., Kremer, R., Gianola, D., 2008. A comparison between Poisson and zero-inflated Poisson regression models with and application to number of black spots in Corriedale sheep. *Genetics, Selection, Evolution* 40(4), 379-394.

Pace, L., Salvan, A., Sartori, N., 2011. Adjusting composite likelihood ratio statistics. *Statistica Sinica* 21(1), 129-148.

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

Rouwendal, J., Boter, J., 2009. Assessing the value of museums with a combined discrete choice/count data model. *Applied Economics* 41(11), 1417–1436.

Sandow, E., 2011. On the road: Social aspects of commuting long distances to work. Unpublished doctoral dissertation, Umea University, Sweden.

Stone, A.A., Schneider, S., Harter, J.K., 2012. Day-of-week mood patterns in the United States: On the existence of ‘Blue Monday’, ‘Thank God it's Friday’ and weekend effects. *The Journal of Positive Psychology* 7(4), 306-314.

Strathman, J.G., Dueker, K.J., 1995. Understanding trip chaining. 1990 NPTS Special Reports on Trips and Vehicle Attributes. Report FHWA-PL-95-033. Federal Highway Administration, US Department of Transportation, Washington, D.C.

Terza, J.V., Wilson, P.W., 1990. Analyzing frequencies of several types of events: A mixed multinomial-Poisson approach. *Review of Economics and Statistics* 72(1), 108-115.

Tong, Y.L., 1990. *The multivariate normal distribution*. Springer, New York.

Train, K., 2003. *Discrete choice methods with simulation*. 1st edition. Cambridge University Press, Cambridge.

Van Acker, V., Witlox, F., 2011. Commuting trips within tours: How is commuting related to land use? *Transportation* 38(3), 465-486.

van Ommeren, J.N., van der Straaten, W., 2008. The effects of search imperfections on commuting behavior: evidence from employed and self-employed workers. *Regional Science and Urban Economics* 38(2), 127-147.

Varin, C., Reid, N., Firth, D., 2011. An overview of composite likelihood methods. *Statistica Sinica* 21(1), 5-42.

von Haefen, R.H., Phaneuf, D.J., 2003. Estimating preferences for outdoor recreation: A comparison of continuous and count data demand system frameworks. *Journal of Environmental Economics & Management* 45(3), 612-630.

von Haefen, R.H., Phaneuf, D.J., 2005. Kuhn-Tucker demand system approaches to nonmarket valuation. *In Applications of Simulation Methods in Environmental and Resource Economics*, R. Scarpa and A.A. Alberini (Eds.), Springer.

Wales, T.J., Woodland, A.D., 1983. Estimation of consumer demand systems with binding non-negativity constraints. *Journal of Econometrics* 21(3), 263-85.

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

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

Zavoina, R., McKelvey, W., 1975. A statistical model for the analysis of ordinal-level dependent variables. *Journal of Mathematical Sociology* 4, 103-120.

Zhang, F., Huang, C.L., Lin, B-H., Epperson, J.E., 2008. Modeling fresh organic produce consumption with scanner data: A generalized double hurdle model approach. *Agribusiness: An International Journal* 24(4), 510-22.

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

APPENDIX A

Proof of Theorem 2

From Theorem 1, the density function is:

$$g(z;\boldsymbol{b},\boldsymbol{\Sigma}) = \sum_{i=1}^{I} g_i(z;\tilde{\boldsymbol{b}}_i,\tilde{\boldsymbol{\Sigma}}_i),$$

where $\tilde{\boldsymbol{b}}_i = \boldsymbol{b}_{\text{-}i} + \overline{\boldsymbol{\Sigma}}_{\text{-}i}(\omega_{\Sigma i}^2)^{-1}(z-b_i)$, and $\tilde{\boldsymbol{\Sigma}}_i = \boldsymbol{\Sigma}_{\text{-}i,\text{-}i} - \overline{\boldsymbol{\Sigma}}_{\text{-}i}(\omega_{\Sigma i}^2)^{-1}\overline{\boldsymbol{\Sigma}}'_{\text{-}i}$.

Now consider the i^{th} component of the density function $g(z;\boldsymbol{b},\boldsymbol{\Sigma})$:

$$\begin{aligned} g_i(z;\tilde{\boldsymbol{b}}_i,\tilde{\boldsymbol{\Sigma}}_i) &= f(z;b_i,\omega_{\Sigma i}^2)\times F_{I\text{-}1}(z\mathbf{1}_{I-1};\tilde{\boldsymbol{b}}_i,\tilde{\boldsymbol{\Sigma}}_i) \\ &= f(z;b_i,\omega_{\Sigma i}^2)\times F_{I\text{-}1}\left[z\left(\mathbf{1}_{I-1} - \overline{\boldsymbol{\Sigma}}_{\text{-}i}(\omega_{\Sigma i}^2)^{-1}\right);(\boldsymbol{b}_{\text{-}i} - \overline{\boldsymbol{\Sigma}}_{\text{-}i}(\omega_{\Sigma i}^2)^{-1}b_i),\tilde{\boldsymbol{\Sigma}}_i\right] \\ &= f(z;b_i,\omega_{\Sigma i}^2)\times F_{I\text{-}1}\left(z\vartheta_i;\hat{\boldsymbol{b}}_i,\tilde{\boldsymbol{\Sigma}}_i\right) \end{aligned}$$

where, $\vartheta_i = \mathbf{1}_{I-1} - \overline{\boldsymbol{\Sigma}}_{\text{-}i}(\omega_{\Sigma i}^2)^{-1}$ and $\hat{\boldsymbol{b}}_i = \boldsymbol{b}_{\text{-}i} - \overline{\boldsymbol{\Sigma}}_{\text{-}i}(\omega_{\Sigma i}^2)^{-1}b_i$.

The univariate normal density function, $f(z;b_i,\omega_{\Sigma i}^2)$, is strictly log-concave, as can be observed by twice-differentiating the logarithm of this density function and showing that it is negative for any value of z:

$$\frac{\partial\, ln\, f(z;b_i,\omega_{\Sigma i}^2)}{\partial z} = -\frac{1}{\omega_{\Sigma i}}\left(\frac{z-b_i}{\omega_{\Sigma i}}\right), \text{ and } \frac{\partial^2 f(z;b_i,\omega_{\Sigma i}^2)}{\partial z^2} = -\frac{1}{\omega_{\Sigma i}^2}.$$

Next, consider $F_{I\text{-}1}\left(z\vartheta_i;\hat{\boldsymbol{b}}_i,\tilde{\boldsymbol{\Sigma}}_i\right)$. The multivariate normal cumulative distribution function is a non-decreasing function. It is also log-concave, as can be shown through the log-concavity of the multivariate normal density function (if a density function is unimodal and log-concave, then the corresponding cumulative distribution function is also log-concave). This is done by taking the second partial derivative and showing that the net result is a negative definite matrix:

$$\frac{\partial\, ln\, f_R(\boldsymbol{z};\boldsymbol{\tau},\boldsymbol{\Gamma})}{\partial \boldsymbol{z}} = -(\boldsymbol{z}-\boldsymbol{\tau})'\boldsymbol{\Gamma}^{-1}, \text{ and } \frac{\partial^2 ln\, f_R(\boldsymbol{z};\boldsymbol{\tau},\boldsymbol{\Gamma})}{\partial\boldsymbol{z}\partial\boldsymbol{z}'} = -\boldsymbol{\Gamma}^{-1}.$$

Note that $\boldsymbol{\Gamma}$ is a positive definite matrix, and thus $-\boldsymbol{\Gamma}^{-1}$ is a negative definite matrix. Finally, the function $h(z) = z\vartheta_i$ embedded within the multivariate normal cumulative distribution function $F_{I\text{-}1}\left(z\vartheta_i;\hat{\boldsymbol{b}}_i,\tilde{\boldsymbol{\Sigma}}_i\right)$ is concave (because the second derivative of $z\vartheta_i$ with respect to z is non-positive). Thus, by proposition 8 (iii) of Azzalini and Regoli (2012), we have the result that $F_{I\text{-}1}\left(z\vartheta_i;\hat{\boldsymbol{b}}_i,\tilde{\boldsymbol{\Sigma}}_i\right)$ is log-concave and, by Corollary 1 of Azzalini and Regoli (2012), we have the result that $g_i(z;\tilde{\boldsymbol{b}}_i,\tilde{\boldsymbol{\Sigma}}_i)$ $(i=1,2,...,I)$ is log-concave. Thus, the density function of is the sum of log-concave density functions.

Note that, for an exchangeable multivariate normally distributed vector $\boldsymbol{X}$, each $g_i(z;\tilde{b}_i,\tilde{\Sigma}_i)$ component is the same (across *i*), and thus the density function of $g(z;\boldsymbol{b},\boldsymbol{\Sigma})$ is simply *I* times a log-concave function. In this situation, the density function of $Max(\boldsymbol{X})$ is log-concave, which implies strong unimodality.

Proof of Theorem 3

$$H(z;\boldsymbol{b},\boldsymbol{\Sigma},\vartheta,\mu,\upsilon^2) = P\left[\vartheta Max(\boldsymbol{X})+W<z\right] \;=\; P\left[Max(\vartheta\boldsymbol{X}+W\mathbf{1}_I)<z\right]$$
$$= P\left[\vartheta X_1+W<z \text{ and } \vartheta X_2+W<z \text{ and } ,..., \ \vartheta X_I+W<z\right]$$
$$= F_I(z\mathbf{1}_I;\vartheta\boldsymbol{b}+\mu\mathbf{1}_I,\vartheta^2\boldsymbol{\Sigma}+\mathbf{IDEN}_I\upsilon^2)$$

The proof that the density function takes the form in Theorem 3 can be shown by differentiating the above function with respect to z and using the last result from Lemma 1.

Proof of Theorem 4

Consider the density function of $\xi=\vartheta\eta+W$ as provided in Theorem 3 (with all notations as defined in the text):

$$h(z;\boldsymbol{b},\boldsymbol{\Sigma},\vartheta,\mu,\upsilon^2) = \sum_{i=1}^{I} f(z;\vartheta b_i+\mu,\vartheta^2\omega_{\boldsymbol{\Sigma}i}^2+\upsilon^2)\times F_{I-1}(z\mathbf{1}_{I-1};\tilde{\boldsymbol{b}}_{i\xi},\tilde{\boldsymbol{\Sigma}}_{i\xi})$$

$$= \sum_{i=1}^{I} f(z;\vartheta b_i+\mu,\omega_{\boldsymbol{\Sigma}i\xi}^2)\times F_{I-1}\left(\left(\frac{z\mathbf{1}_{I-1}\text{ - }(\vartheta b_i+\mu)\mathbf{1}_{I-1}}{\omega_{\boldsymbol{\Sigma}i\xi}}\right)\omega_{\boldsymbol{\Sigma}i\xi}-\frac{\vartheta^2\overline{\boldsymbol{\Sigma}}_{\text{-}i}}{\omega_{\boldsymbol{\Sigma}i\xi}}\left(\frac{z-(\vartheta b_i+\mu)}{\omega_{\boldsymbol{\Sigma}i\xi}}\right);\boldsymbol{\gamma}_i,\tilde{\boldsymbol{\Sigma}}_{i\xi}\right)$$

$$= \sum_{i=1}^{I}\frac{1}{\omega_{\boldsymbol{\Sigma}i\xi}}\phi(y_i)\times F_{I-1}\left(\boldsymbol{\lambda}_i\, y_i;\boldsymbol{\gamma}_i,\tilde{\boldsymbol{\Sigma}}_{i\xi}\right)$$

where $y_i=\dfrac{z-(\vartheta b_i+\mu)}{\omega_{\boldsymbol{\Sigma}i\xi}}$ and $\boldsymbol{\lambda}_i=\omega_{\boldsymbol{\Sigma}i\xi}\mathbf{1}_{I-1}-\dfrac{\vartheta^2\overline{\boldsymbol{\Sigma}}_{\text{-}i}}{\omega_{\boldsymbol{\Sigma}i\xi}}$.

The moment generating function of ξ is given by:

$$M_\xi(t)=\int_{z=-\infty}^{\infty} e^{tz}h(z;\boldsymbol{b},\boldsymbol{\Sigma},\vartheta,\mu,\upsilon^2)\,dz=\int_{z=-\infty}^{\infty} e^{tz}\sum_{i=1}^{I}\frac{1}{\omega_{\boldsymbol{\Sigma}i\xi}}\phi(y_i)\times F_{I-1}\left(\boldsymbol{\lambda}_i\, y_i;\boldsymbol{\gamma}_i,\tilde{\boldsymbol{\Sigma}}_{i\xi}\right)dz$$

$$= \sum_{i=1}^{I} \int_{y_i=-\infty}^{\infty} e^{t\left(y_i \omega_{\Sigma i\xi} + \vartheta b_i + \mu\right)} \phi(y_i) \times F_{I-1}\left(\boldsymbol{\lambda}_i\, y_i ; \boldsymbol{\gamma}_i , \tilde{\boldsymbol{\Sigma}}_{i\xi}\right) dy_i$$

$$= \sum_{i=1}^{I} \int_{y_i=-\infty}^{\infty} e^{t\left(\vartheta b_i + \mu\right)} e^{t\left(y_i \omega_{\Sigma i\xi}\right)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} y_i^2} \times F_{I-1}\left(\boldsymbol{\lambda}_i\, y_i ; \boldsymbol{\gamma}_i , \tilde{\boldsymbol{\Sigma}}_{i\xi}\right) dy_i$$

$$= \sum_{i=1}^{I} \int_{y_i=-\infty}^{\infty} e^{t\left(\vartheta b_i + \mu\right)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(y_i^2 - 2 y_i \omega_{\Sigma i\xi} t + \omega_{\Sigma i\xi}^2 t^2\right) + \frac{1}{2}\omega_{\Sigma i\xi}^2 t^2} \times F_{I-1}\left(\boldsymbol{\lambda}_i\, y_i ; \boldsymbol{\gamma}_i , \tilde{\boldsymbol{\Sigma}}_{i\xi}\right) dy_i$$

$$= \sum_{i=1}^{I} \int_{y_i=-\infty}^{\infty} e^{t\left(\vartheta b_i + \mu\right) + \frac{1}{2}\omega_{\Sigma i\xi}^2 t^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(y_i - \omega_{\Sigma i\xi} t\right)^2} \times F_{I-1}\left(\boldsymbol{\lambda}_i\, y_i ; \boldsymbol{\gamma}_i , \tilde{\boldsymbol{\Sigma}}_{i\xi}\right) dy_i$$

$$= \sum_{i=1}^{I} e^{t\left(\vartheta b_i + \mu\right) + \frac{1}{2}\omega_{\Sigma i\xi}^2 t^2} \int_{u_i=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} u_i^2} \times F_{I-1}\left(\boldsymbol{\lambda}_i (u_i + \omega_{\Sigma i\xi} t) ; \boldsymbol{\gamma}_i , \tilde{\boldsymbol{\Sigma}}_{i\xi}\right) du_i \text{, where } u_i = y_i - \omega_{\Sigma i\xi} t$$

$$= \sum_{i=1}^{I} e^{t\left(\vartheta b_i + \mu\right) + \frac{1}{2}\omega_{\Sigma i\xi}^2 t^2} \int_{u_i=-\infty}^{\infty} \phi(u_i) \times F_{I-1}\left(\boldsymbol{\lambda}_i (u_i + \omega_{\Sigma i\xi} t) ; \boldsymbol{\gamma}_i , \tilde{\boldsymbol{\Sigma}}_{i\xi}\right) du_i$$

$$= \sum_{i=1}^{I} e^{t\left(\vartheta b_i + \mu\right) + \frac{1}{2}\omega_{\Sigma i\xi}^2 t^2} \times E_{u_i}\left[F_{I-1}\left(\boldsymbol{\lambda}_i (u_i + \omega_{\Sigma i\xi} t) ; \boldsymbol{\gamma}_i , \tilde{\boldsymbol{\Sigma}}_{i\xi}\right)\right]$$

$$= \sum_{i=1}^{I} e^{t\left(\vartheta b_i + \mu\right) + \frac{1}{2}\omega_{\Sigma i\xi}^2 t^2} \times F_{I-1}\left(\boldsymbol{\lambda}_i\, \omega_{\Sigma i\xi} t ; \boldsymbol{\gamma}_i , \tilde{\boldsymbol{\Sigma}}_{i\xi} + \boldsymbol{\lambda}_i \boldsymbol{\lambda}_i'\right)$$

since $E_{u_i}\left[F_{I-1}\left(a + \boldsymbol{B}\, u_i ; \boldsymbol{\gamma}_i , \tilde{\boldsymbol{\Sigma}}_i\right)\right] = F_{I-1}\left(a ; \boldsymbol{\gamma}_i , \tilde{\boldsymbol{\Sigma}}_i + \boldsymbol{B}\boldsymbol{B}'\right)$ for all scalar *a,* vector ***B*** and random variable $u_i \sim N(0,1)$ (see Marsaglia, 1963 and Gupta *et al.*, 2004).

Finally, $M_{\xi}(t) = \sum_{i=1}^{I} e^{t(\vartheta b_i + \mu) + \frac{1}{2}\omega_{\Sigma i\xi}^2 t^2} \times F_{I-1}(\boldsymbol{s}_{\Sigma i\xi}\, t ; \boldsymbol{\gamma}_i , \boldsymbol{\Psi}_{i\xi})$.

Table 1. Sample Characteristics

Variable	Share [%]	Variable	Share [%]
Individual characteristics		**Household characteristics**	
Race and ethnicity		*Household income [US$/year]*	
Non-Hispanic Caucasian	71.56	Less than 80,000	46.66
Hispanic	9.99	80,000 or more	53.34
Non-Hispanic Asian	9.37	*Home location*	
Non-Hispanic African-American	4.45	Urban cluster	94.18
Non-Hispanic Other	4.63	Not in urban cluster	5.82
Gender		**Work-related characteristics**	
Male	52.25	*Employment Industry*	
Female	47.75	Professional, managerial or technical	48.62
Driver status		Sales or services	23.32
Has driver's license	98.58	Clerical or administrative support	14.59
Does not have a driver's license	1.42	Other	13.47
Highest education level		Is self-employed	9.51
At least some college education	76.53	Has flexible work start time	44.87
No college education	23.47	Has more than one job	9.13
Past week primary activity		Has the option to work at home	13.06
Work	94.18	**Activity and travel characteristics**	
Other activity	5.82	Survey day is Friday	17.79
Shopped via internet in past month		Used public transportation on survey day	3.98
No	57.31	At least one walk trip in past week	63.98
Yes	42.69	At least one bike trip in past week	6.58

Descriptive statistics

Variable	Mean	Std. Dev.	Min.	Max.
Individual characteristics				
Age [years]	46.67	12.70	18.00	86.00
Household characteristics				
Number of adults	2.40	0.92	1.00	7.00
Number of non-adults	0.74	1.05	0.00	6.00
Number of drivers	2.33	0.92	0.00	7.00
Number of vehicles	2.59	1.30	0.00	12.00
Number of workers	1.84	0.82	1.00	5.00
Work-related characteristics				
Distance to work [miles]	13.52	12.56	0.11	97.00

Dependent variable: Number of out-of-home non-work episodes

Time-of-day block	Mean	Std. Dev.	Min.	Max.
Before-work (BW)	0.12	0.44	0.00	6.00
Home-to-work commute (HWC)	0.20	0.56	0.00	11.00
Work-based (WB)	0.23	0.47	0.00	4.00
Work-to-home commute (WHC)	0.43	0.83	0.00	6.00
After-home (AH)	0.56	1.12	0.00	12.00
Total non-work episodes	1.54	1.67	0.00	13.00

Table 2. Joint Model Estimation Results - Count Data Model Component

Variables	Latent Propensity Coefficients		Threshold Coefficients	
	Estimate	t-stat	Estimate	t-stat
Constant in φ vector			-0.3733	-1.683
Threshold specific constants				
α_1			0.0837	1.222
α_1 to α_5			0.0887	0.787
α_6			0.1447	0.827
Individual characteristics				
Race and ethnicity (non-Hispanic and non-Asian)				
Hispanic	-0.1787	-1.500		
Non-Hispanic Asian	-0.1796	-1.470		
Gender (male)				
Female - mean effect	0.1933	2.217		
- std. deviation	0.8789	8.200		
Past week primary activity (work)				
Other activity	0.3393	2.304		
Shopped via internet in past month (no)				
Yes	0.3442	4.426		
Household characteristics				
Home location (urban cluster)				
Not in urban cluster	-0.5824	-3.668		
Household composition				
Number of adults	-0.1670	-2.886		
Number of non-adults - mean effect	0.1952	5.453		
- std. deviation	0.3018	5.097		
Number of workers			0.1059	5.701
Work-related characteristics				
Is self-employed *(not self-employed)*	0.2707	2.277		
Has the option to work at home *(cannot work from home)*			0.3577	4.189
Has more than one job *(has only one job)*	0.2557	2.222		
Distance to work [miles/100]	-1.6488	-5.444		
Mobility and situational characteristics				
Used public transportation on survey day *(not used public transportation on survey day)*	-0.3927	-2.098		
At least one walk trip in past week *(no walk trip in past week)*	0.2562	2.996		
At least one bike trip in past week *(no bike trip in past week)*	0.1643	1.437		
Linkage parameter ϑ	1.0660	6.020		

Table 3. Joint Model Estimation Results - Event Type Choice Model Component

Variables	Coefficient		Standard Deviation	
	Estimate	t-stat	Estimate	t-stat
Constants				
HWC	-0.4717	-5.457	0.6888	4.440
WB	-0.8882	-7.609		
WHC	0.3764	3.261	0.2739	1.639
AH	0.5233	7.334		
Accessibility measures at the home location for BW, HWC, WHC and AH time-of-day blocks [number of jobs/100,000]				
For the entire population				
Natural resources	-0.9339	-1.843		
Manufacturing	-0.0773	-2.015		
Information	-0.1487	-1.596		
Financial services	-0.0847	-1.307		
Educational	-0.8455	-4.161		
Wholesale trade	0.4065	2.259		
Health	0.2268	2.298		
Entertainment	0.2781	2.967	0.2757	5.170
For females only				
Retail trade	0.0490	1.114		
Accessibility measures at the workplace location for HWC, WB and WHC time-of-day blocks [number of jobs/100,000]				
For the entire population				
Manufacturing	-0.0363	-2.202		
Information	-0.0702	-1.258		
Financial services	0.0999	1.460		
For females only				
Retail trade	0.0360	1.934		
Work-related characteristics				
Is self-employed				
WB	0.3045	2.021		
WHC	-0.0615	-0.853		
Has flexible work start time				
BW	-0.6257	-7.040		
Mobility and situational characteristics				
Survey day is Friday				
WHC and AH	0.1827	2.115		
Used public transportation on survey day				
BW	-1.8864	-11.974		

Table 4. Covariance Matrix for the Event Type Choice Model Component

Time-of-Day Block	BW	HWC	WB	WHC	AH
BW	0.5				
HWC	0.0	0.5			
WB	0.0	0.0	0.5		
WHC	0.0	0.0	0.0	0.5	
AH	0.0	0.0	0.0	**0.5146 (29.153)***	**0.5695 (11.535) ****

* t-stat computed with respect to zero
** t-stat computed with respect to 0.5

Table 5. Aggregate Data Fit Measures

Aggregation Level	Combination Event		Observed	Joint Model		Independent Model	
				Predicted	APE	Predicted	APE
Multivariate	No participation		676	669.7	0.9	656.1	2.9
	BW participation only		67	90.5	35.0	95.2	42.1
	HWC participation only		67	85.2	27.1	87.0	29.9
	WB participation only		168	63.0	62.5	67.4	59.9
	WHC participation only		230	153.6	33.2	137.0	40.4
	AH participation only		279	345.4	23.8	347.7	24.6
	Overall mean weighted APE			**19.9**		**22.2**	
Marginal	Total count	0	676	669.7	0.9	656.1	2.9
		1	593	589.1	0.7	585.7	1.2
		2	388	379.1	2.3	377.2	2.8
		3	208	225.6	8.4	253.5	21.9
		4+	248	249.5	0.6	240.5	3.0
	Weighted APE			1.8		4.3	
	BW block count	0	1926	1756.8	8.8	1745.3	9.4
		1	147	295.8	101.2	305.7	108.0
		2	24	47.8	99.1	50.8	111.5
		3	13	7.6	41.5	8.1	37.7
		4+	3	5.0	66.7	3.1	3.3
	Weighted APE			16.5		17.6	
	HWC block count	0	1792	1739.1	3.0	1729.8	3.5
		1	250	317.8	27.1	326.4	30.6
		2	57	45.8	19.6	48.1	15.7
		3	10	5.9	41.0	6.2	38.3
		4+	4	4.4	10.0	2.5	37.5
	Weighted APE			6.5		7.3	
	WB block count	0	1660	1831.5	10.3	1809.9	9.0
		1	421	244.8	41.9	261.2	38.0
		2	29	29.2	0.7	35.0	20.7
		3	2	3.6	77.5	4.7	133.2
		4+	1	3.9	290.0	2.2	120.0
	Weighted APE			16.7		15.1	
	WHC block count	0	1516	1560.2	2.9	1593.4	5.1
		1	397	423.4	6.6	408.7	3.0
		2	131	97.1	25.8	87.2	33.5
		3	45	22.0	51.0	17.6	60.8
		4+	24	10.3	57.1	6.1	74.6
	Weighted APE			6.7		8.4	
	AH block count	0	1465	1216.0	17.0	1201.7	18.0
		1	394	576.9	46.4	589.8	49.7
		2	90	214.3	138.1	221.7	146.3
		3	103	71.9	30.2	71.6	30.5
		4+	61	33.9	44.4	28.2	53.8
	Weighted APE			29.1		31.0	

Table 6. Aggregate Change in Expected Number of Non-Work Episodes

Effect of 20% increase in …	Time-of-day block	Joint Model	Independent Model
Distance to work	All day	-2.51	-2.30
	BW	-2.64	-2.44
	HWC	-2.52	-2.32
	WB	-2.30	-2.08
	WHC	-2.46	-2.27
	AH	-2.46	-2.27
Number of retail trade jobs at the home location	All day	0.53	0.00
	BW	0.75	0.31
	HWC	0.85	0.40
	WB	-1.30	-2.34
	WHC	0.62	0.06
	AH	0.69	0.31
Number of entertainment jobs at the home location	All day	3.44	0.00
	BW	8.31	5.23
	HWC	5.93	2.32
	WB	-2.41	-6.94
	WHC	1.52	-1.75
	AH	3.27	0.37