A New Utility-Consistent 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

Original: January 2013
1st Revision: August 2013
2nd Revision: February 2014

ABSTRACT

In the current paper, we propose a new utility-consistent 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 important 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.

In the current paper, we propose a parametric utility-consistent 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]

1.1. 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, and Haque *et al.*, 2010). The advantage of this method is that it permits both positive and negative

[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.

dependency between the counts, but the limitations are that the approach gets quickly cumbersome in the presence of several mixing components. Recently, Bhat and colleagues (see Castro *et al.*, 2012, Narayanamoorthy *et al.*, 2013, Bhat *et al.*, 2014) have addressed this problem by recasting count models as a special case of generalized ordered-response models with underlying continuous latent variables, and introducing multivariateness through the specification of the error terms in the continuous latent variables (this approach also happens to nest the copula approach proposed by van Ophem, 1999 as a special case). These models allow for a more "linear" introduction of the dependencies and, in combination with a new estimation technique proposed by the authors, lead to a simple way to estimate correlated count data models. But 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. Thus, these models are not of much use for economic welfare analysis, which can be very important in many recreational, cultural, and other empirical contexts. 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, Bhat, 2005, and Bhat *et al.*, 2009 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 utility-consistent formulation that explicitly recognizes the discrete nature of consumption quantity would be more desirable.[2]

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

[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.

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. This is because there is no linkage of any kind from the event type choice model back to the total count model. The structure without this linkage is also not consistent with utility theory, as we show in Appendix B in the online supplement to this paper. 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 and Hausman *et al.*, 1995, and Rouwendal and Boter, 2009). But a problem with the way this structure has been implemented in the earlier studies is that the resulting model is inconsistent with utility theory (more on this later) and/or 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, and not the full distribution of maximum utility, resulting in the absence of a 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 integrated nature of the event choice and the total count decisions. Unfortunately, if this were to be considered 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 extreme-value distributed, and including this maximum utility distribution form in the count model leads to difficult distributional mismatch issues in the count model component of the joint model (this is perhaps the reason that earlier models have not considered the full distribution of the maximum utility in the count model). 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 ensuring a utility-consistent model for multivariate counts that considers the linkage in the total count and event choice components of the model system by accommodating the complete distribution of maximum utility from the event type choice model to the total count model. To our knowledge, this is the first such joint model proposed in the literature. 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 GEV-based kernels (dominantly the multinomial logit (MNL) or the nested logit (NL) kernel) used in earlier studies. The use of the MNP kernel has several advantages, including allowing a more flexible covariance structure for the event utilities relative to traditional GEV kernels, ensuring that the resulting model is utility-consistent based on separability of the direct utility function (Hausman *et al*.'s (1995) model, while stated by the authors as being utility-consistent, is actually not utility-consistent because they use a GEV kernel for the choice model, as discussed later), and also facilitating the linkage between the event choice and the total count components of our proposed model system (this is because the cumulative distribution of the maximum over a multivariate normally distributed vector takes back the form of a cumulative multivariate normal distribution, which we exploit in the way we introduce the linkage between the event type choice model and the total count model in our modeling approach).[3] Second, and related to the first, 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). The recasting facilitates the inclusion of the linkage as well as easily accommodates random taste variations, because of the conjugate nature of the multivariate normal distribution of the linkage

[3] As a secondary contribution, the paper potentially opens up a whole new area of studies of welfare economics that use an MNP kernel for choice models, as opposed to the use of GEV-based models for welfare economics. Indeed, we have found no discussion in the literature on welfare economics of consumer surplus concepts in the context of MNP choice models, primarily because results regarding the distribution of the maximum of a multivariate normally distributed vector (with a general covariance matrix) have been recent and have been confined to the statistical literature. In this regard, the current paper brings these recent statistical results on the distribution of the maximum of multivariate normally distributed variables, along with new results that we establish, into the economic domain of utility-based models.

parameter (that includes the random taste variations in the event type choice model) and the multivariate normal distribution for the random taste variations in the count model. Further, the recasting can easily accommodate high or low probability masses for specific count outcomes without the need for zero-inflated or hurdle approaches, and allows the use of a specific estimation approach that very quickly evaluates multivariate normal cumulative distribution functions. Third, we establish a few new results regarding the distribution of the maximum of multivariate normally distributed random variables (with a general covariance matrix). These results constitute another core element in our utility-consistent 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, though these have simply not been invoked in econometric models. In doing so, we bring recent developments in the statistical field into the economic field. Fourth, we propose the estimation of our joint model for multivariate count data using Bhat's (2011) frequentist MACML (for maximum composite marginal likelihood) approach, which is easy to code and computationally time efficient (see also Bhat and Sidharthan, 2011). 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 modeling system for multivariate count data. In summary, it is the combination of multiple things that work in tandem that lead to our proposed new utility-consistent, flexible, and easy-to-estimate model, including the use of an MNP kernel for the event type choice, the recasting of traditional count models as generalized ordered-response models, the application of new statistical results for the maximum of multivariate normally distributed variables, and the use of the MACML estimation approach for estimation.

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 illustrates an application of the proposed model for analyzing out-of-home non-work episodes pursued by workers. Finally, Section 4 summarizes the key findings of the paper and identifies directions for further research.

2. 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 h_q. Also, let there be I $(i = 1,2,...,I)$ event type possibilities (or alternatives) that the total count h_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 h_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 h_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 h_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 econometric formulation for the event choice at each choice occasion, while the subsequent section develops the econometric formulation for the total count model (including the linkage between the event choice and the total count).

2.1. 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:

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

where $\boldsymbol{x}_{qi}$ is a $(D\times1)$-column vector of exogenous attributes (including a constant), and $\boldsymbol{\beta}_q$ is an individual-specific $(D\times1)$-column vector of corresponding coefficients that is a realization from a multivariate normal density function with mean vector $\boldsymbol{b}$ and covariance matrix $\boldsymbol{\Omega}$ (this

[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).

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 $\tilde{\boldsymbol{\varepsilon}}_{qt} \sim MVN_I(\boldsymbol{0}_I, \boldsymbol{\Theta})$, leading to a multinomial probit (MNP) model of event type choice ($MVN_I(\boldsymbol{0}_I, \boldsymbol{\Theta})$ stands for the multivariate normal distribution of I dimensions with mean vector $\boldsymbol{0}_I$ and covariance matrix $\boldsymbol{\Theta}$). 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 $\boldsymbol{\varepsilon}_{qt1} = (\varepsilon_{qt21}, \varepsilon_{qt31}, ..., \varepsilon_{qtI1})$. 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). Also, in MNP models where the variables are all specific to individuals (and whose values do not vary across alternatives), empirical identification issues need to be considered. In particular, as discussed by Keane (1992) and Munkin and Trivedi (2008), identification is tenuous unless exclusion restrictions are placed in the form of at least one individual characteristic being excluded from each alternative's utility in addition to being excluded from a base alternative (but appearing in the utilities of some other alternatives). In our application, this empirical identification issue does not arise because we do have alternative-specific variables in the event type choice model.

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

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

where $\boldsymbol{V}_q = \mathbf{1}_T \otimes \left[\mathbf{x}_q \boldsymbol{b}\right]$ and $\boldsymbol{\varepsilon}_q = \mathbf{1}_T \otimes \left[\mathbf{x}_q \tilde{\boldsymbol{\beta}}_q\right] + \tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{q}}$. 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:

$$\boldsymbol{u}_q^* = \boldsymbol{M}_q \boldsymbol{U}_q = \boldsymbol{M}_q \boldsymbol{V}_q + \boldsymbol{M}_q \boldsymbol{\varepsilon}_q. \quad (3)$$

To determine the covariance matrix of $\boldsymbol{u}_q^*$, define $\tilde{\boldsymbol{\Omega}}_q = \mathbf{1}_{TT} \otimes \left[\mathbf{x}_q \boldsymbol{\Omega} \mathbf{x}_q\right]$ ($TI \times TI$ matrix) and $\tilde{\boldsymbol{\Theta}} = \mathbf{IDEN}_T \otimes \boldsymbol{\Theta}$ ($TI \times TI$ matrix). Let $\tilde{\mathbf{F}}_q = \left[\tilde{\boldsymbol{\Omega}}_q + \tilde{\boldsymbol{\Theta}}\right]$ 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:

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

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 $\boldsymbol{\Theta}$. To write this, as well as for future use, we define several key notations 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 $\mathbf{\Gamma}^*$, $F(.;\mu,\sigma^2)$ for the univariate normal cumulative distribution function with mean μ and variance σ^2, $\Phi(.)$ for the univariate standard normal cumulative distribution function, $F_R(.;\boldsymbol{\tau},\mathbf{\Gamma})$ for the multivariate normal cumulative distribution function of dimension R with mean vector $\boldsymbol{\tau}$ and covariance matrix $\mathbf{\Gamma}$, and $\Phi_R(.;\mathbf{\Gamma}^*)$ for the multivariate standard normal cumulative distribution function of dimension R and correlation matrix $\mathbf{\Gamma}^*$ (these notations will also be used in Appendix A in the online supplement to this paper). The likelihood contribution of individual q from the event type choice model is then the $[(I-1)\times n_q]$-dimensional integral below:

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

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 2.3 after first discussing the total count model formulation in the next section.

2.2. 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 model. Specifically, as discussed by CPB (2012), any count model may be reformulated as a special case of a generalized ordered-response model in which a single latent continuous variable is partitioned into mutually exclusive intervals. Using this equivalent latent variable-based generalized-ordered response 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 generalized ordered-response probit model in Section 2.2.1, and then discuss the linkage with the event type model in Section 2.2.2.

2.2.1. 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 (h_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:

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

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 (a) ensures that the thresholds δ_{qk} satisfy the ordering conditions ($\delta_{q,-1} = -\infty$; $-\infty <$ $\delta_{q0} < \delta_{q1} < \delta_{q2} < \delta_{q,3} < ...$) and (b) allows identification for any variables that are common in $\boldsymbol{w}_q$ and $\boldsymbol{\varpi}_q$. g_q^* in Equation (6) corresponds to the latent propensity underlying the observed count variable g_q, $\boldsymbol{w}_q$ is an ($L\times 1$)-column vector of exogenous attributes (excluding a constant), $\boldsymbol{\theta}_q$ is a corresponding ($L\times 1$)-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.

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. For example, if $f_k(\varpi_q)$ = $f_k(\varpi_q) = \Phi^{-1}\left(e^{-\lambda_q} \sum_{l=0}^{k} \frac{\lambda_q^l}{l!} \right)$, $\lambda_q = e^{\varphi' \varpi_q}$ (φ is a parameter vector), $\alpha_k = 0\ \forall k$ and $\boldsymbol{\theta}_q = 0$, the result is the Poisson count model:

$$\begin{aligned} P[g_q = k] &= P\left[\Phi^{-1}\left(e^{-\lambda_q} \sum_{l=0}^{k-1} \frac{\lambda_q^l}{l!} \right) < g_q^* < \Phi^{-1}\left(e^{-\lambda_q} \sum_{l=0}^{k} \frac{\lambda_q^l}{l!} \right) \right] \\ &= \Phi\left(\Phi^{-1}\left(e^{-\lambda_q} \sum_{l=0}^{k} \frac{\lambda_q^l}{l!} \right) \right) - \Phi\left(\Phi^{-1}\left(e^{-\lambda_q} \sum_{l=0}^{k-1} \frac{\lambda_q^l}{l!} \right) \right) = \frac{e^{-\lambda_q} \lambda_q^k}{k!} = \frac{e^{-e^{\varphi'\varpi_q}} \left(e^{\varphi'\varpi_q} \right)^k}{k!} \end{aligned} \tag{7}$$

Second, the analyst can accommodate high or low probability masses for specific count 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).[5] Third, the interpretation of the generalized ordered-response 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.

2.2.2. 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.

[5] It should be noted that the analyst needs to place the restriction $\alpha_k = \alpha_K$ for some value of K in order for the GORP reformulation to be able to predict count outcomes beyond those observed in the estimation data.

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

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

Define $\breve{\boldsymbol{U}}_q = (\breve{U}_{q1}, \breve{U}_{q2}, ..., \breve{U}_{qI})'$ ($I \times 1$ vector) and $\breve{\boldsymbol{\varepsilon}}_q = (\tilde{\varepsilon}_{q1}, \tilde{\varepsilon}_{q2}, ..., \tilde{\varepsilon}_{qI})'$ ($I \times 1$ 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 $\boldsymbol{\Sigma}_q = \mathbf{x}_q \boldsymbol{\Omega} \mathbf{x}_q' + \boldsymbol{\Theta}$. Let $\eta_q = \mathrm{Max}(\breve{\boldsymbol{U}}_q)$. η_q, when divided by the marginal utility of income (assuming constant marginal utility of income), is a measure of per choice occasion consumer surplus for individual q. That is, η_q represents the utility that individual q receives from each choice occasion characterizing her/his total demand count (this is because the individual, at each choice occasion, chooses the alternative with the highest utility). Now, it is reasonable and natural to assume that the individual's count choice is a function of the per-choice occasion utility accrued by the individual (as we will show later, and because of our use of an MNP kernel for the event type choice, this assumption also makes our joint model consistent with two-stage budgeting). As the per choice occasion utility for an individual increases, the individual will have a higher count. Equivalently, the introduction of the per-choice occasion consumer surplus or maximum utility measure η_q in the count model is equivalent to the introduction of a single (scalar) price index represented by η_q for the commodity group represented by the count. Note, however, that this is a stochastic variable to the analyst, because the analyst does not observe the utility vector $\breve{\boldsymbol{U}}_q$. Thus, it is important to consider the full distribution of η_q in the count model, as opposed to using simply the expected value of η_q (as has been done by earlier studies, including Hausman *et al.*, 1995 and Rouwendal and Boter, 2009). We introduce the η_q 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:

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

To develop the likelihood function from the total count model, we need the cumulative distribution function of g_q^*, which we obtain from the following theorem:

Theorem 1: The distribution of a stochastic transformation of $\eta_q = \text{Max}(\breve{\boldsymbol{U}}_q)$ as $g_q^* = \vartheta\eta_q + W_q$, , where ϑ is a constant scalar parameter and W_q is a univariate normally distributed scalar $(W_q \sim N(\mu_q, \upsilon_q^2))$ has a cumulative distribution function as below:

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

Appendix A provides the proof in the online supplement to this paper.

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

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

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

2.3. Estimation Technique

As we show in Appendix B in the online supplement to this paper, the choice of the MNP model as the basis for the event type choice, combined with the use of the maximum utility measure η_q from the event type choice model in the count model, makes our overall model of total count and event type choice consistent with a two-stage budgeting approach within a direct utility maximizing planning framework. This allows us to write the count for event type *i* as the product of the total count observed (across all event types) and the probability of observing event type *i* (see Equation B.2 in Appendix B). The net econometric consequence for estimation purposes is that the total count model can be separately analyzed in a first stage (as long as η_q is introduced at this first stage), and the event type choice can be separately analyzed in a second stage independent of the choice of the total count. However, η_q is a random variable with a distribution (because of the presence of individual-level heterogeneity), and has a role in the count model estimation. Specifically, η_q serves as the vehicle that transmits the effect of event type choice determinants and modeling errors into the total count model. Thus, the appropriate likelihood function to maximize in the two-stage budgeting approach corresponds to the product of the likelihood function of the count model (considering the randomness in the η_q variable) and the likelihood of the MNP model. This overall likelihood function for our two-stage total count-event type model may be obtained from Equations (13) and (5) as follows:

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

Note that the two components of the likelihood have common parameters.

To address the issue of the high dimensionality of integration in $L_{q,event}$ (of dimension $h_q(I-1)$) in Equation (14), we replace the log-likelihood from the event model with a composite marginal likelihood (CML), $L_{CML,q,event}$ (this CML is not an approximation of the true likelihood nor does it make any restrictive assumptions regarding the total count and event type models beyond the separability of the likelihood components made possible by two-stage budgeting; rather, the CML is simply a different inference approach that also leads to good asymptotic properties, as we discuss later). 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), 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. Almost all earlier research efforts employing the CML technique have used the pairwise approach. Alternatively, the analyst can also consider larger subsets of observations, such as triplets or quadruplets or even higher dimensional subsets. However, it is generally agreed that the pairwise approach is a good balance between statistical and computational efficiency.

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 (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). Of course, the CML estimator loses some asymptotic efficiency from a theoretical perspective relative to a full likelihood estimator (Lindsay, 1988; Zhao and Joe, 2005). On the other hand, when simulation methods have to be used to evaluate the likelihood function (as would be needed to compute $L_{q,event}$ in Equation (5)), there is also a loss in asymptotic efficiency in the maximum simulated likelihood (MSL) estimator relative to a full likelihood estimator (see McFadden and Train, 2000).

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:

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

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

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

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

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

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).

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

3.1. 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 over the course of a weekday, and the organization of these episodes across five time-of-day blocks during the day. An episode, which is a commonly used term in the travel modeling field, refers to a single instance of participation in a specific activity. An example would be an episode of participation in shopping activity. Note that there can be multiple episodes of non-work activity within a given day.

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 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 McGuckin *et al.*, 2005).

3.2. 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 to/from each out-of-home episode undertaken by 320,000 individuals from 150,000 households sampled from all over the country for one day of the week. The purpose (such as work, shopping, recreation, etc.) of each out-of-home episode was provided by the respondent. 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 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 measures take the following general form for Census tract *i*, industry type *e*, and time period *t*: $A_{iet} = \sum_{j \in L_{it}} O_{je}$, where L_{it} is the set of all Census tracts that are reachable within ten minutes of highway travel from tract *i* during time period *t*, and O_{je} is the number of activity opportunities of industry type *e* at Census tract *j*. The details of the approach to develop L_{it} and O_{je} for each Census tract is provided in Chen *et al*. (2011).[7]

The sample formation included several steps, which are presented in Appendix C in the online supplement to this paper. The table in Appendix C provides an unweighted summary of select individual, household, work-related and activity and travel characteristics of the final sample.

3.3. Estimation Results

3.3.1. Variable Specification

The exogenous variables described in Section 3.2 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

---

[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.

[7] The fifteen industry types used in the accessibility computations are (1) Agriculture (including forestry, fishing and hunting and mining, (2) Construction, (3) Manufacturing, (4) Wholesale trade, (5) Retail trade, (6) Transportation and warehousing and utilities, (7) Information, (8) Finance services (including insurance, real estate and rental and leasing), (9) Professional, scientific, management, administrative, and waste management services, (10) Educational, (11) Health, (12) Entertainment (including arts, entertainment, recreation, accommodation and food services), (13) Other services (except public administration), (14) Public administration, and (15) Armed forces.

accessibility measures constructed at the workplace end were used in the HWC, WB, and WHC blocks.

The final estimation results are presented in Table 1 (for the count data model component) and Table 2 (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.

3.3.2. Count data model component

The first main numeric column of Table 1 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 1 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 2.2.1, 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 1 is based on setting K = 6 (so $\alpha_k = \alpha_6 \forall \; k \geq 6$).

The next row panel of Table 1 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). 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, with a higher number of non-adults in the household leading to a lower level of non-work episodes for almost 26% of 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 *work-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 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 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.

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

Table 2 presents the results of the time-of-day block choice model component. The first row panel of Table 2 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, 2013).

In the category of *work-related characteristics*, self-employed workers are more likely to participate in non-work activity episodes during the work-based (WB) block and less likely to participate during the work-to-home commute (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 2.1, 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 3 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 a 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.

3.4. 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. Due to space constraints, we relegate the presentation of these alternative model fit measures to Appendix D in the online supplement to this paper. Also, in Appendix E of the online supplement, we provide an application of the joint model.

4. CONCLUSIONS

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 acknowledge the helpful comments of four anonymous reviewers on an earlier version of the paper. The authors are grateful to Lisa Macias for her help in formatting this document.

REFERENCES

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

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

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

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

Bhat CR, Born K, Sidharthan R, Bhat P. 2014. A count data model with endogenous covariates: Formulation and application to roadway crash frequency at intersections. *Analytic Methods in Accident Research* **1**: 53-71.

Brunow S, Gründer M. 2013. The impact of activity chaining on the duration of daily activities. *Transportation* **40**(5): 981-1001.

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 JA. 2012. A poisson mixture model of discrete choice. *Journal of Econometrics* **166**(2): 184-203.

Cameron AC, Trivedi PK. 2013. Regression analysis of count data, 2nd edition. *Econometric Society Monograph*. Cambridge University Press: Cambridge.

Castro M, Paleti R, Bhat CR. 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 SY, Lei T, Goulias KG, Pendyala PM, Bhat CR, 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.

Cox DR, 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.

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 VP. 1960. An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics* **31**(4): 1208-1211.

Hanemann WM. 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 JA. Leonard GK, 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.

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

Khan M, Paleti R, Bhat CR, Pendyala RM. 2012. Joint household-level analysis of individuals' work arrangement choices. *Transportation Research Record* **2323**: 56-66.

Kim J, Allenby GM, Rossi PE. 2002. Modeling consumer demand for variety. *Marketing Science* **21**(3): 229-250.

Lindsay BG. 1988. Composite likelihood methods. *Contemporary Mathematics* **80**: 221-239.

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

McDonald NC. 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.

McFadden D, Train K. 2000. Mixed MNL models for discrete response. *Journal of Applied Econometrics* **15**(5): 447-470.

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 HP, Corpuz G, Bauman AE, 2010. Public health perspectives on household travel surveys active travel between 1997 and 2007. *American Journal of Preventive Medicine* **39**(2): 113-121.

Munkin MK, Trivedi PK. (2008). Bayesian analysis of the ordered probit model with endogenous selection. *Journal of Econometrics* **143**(2): 334-348.

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

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

Rajagopalan BS, Pinjari AR, Bhat CR. 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 AA, Schneider S, Harter JK. 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.

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

Train K. 2003. *Discrete Choice Methods with Simulation*. 1st edition. Cambridge University Press, Cambridge.

van Ommeren JN, 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.

van Ophem H. 1999. A general method to estimate correlated discrete random variables. *Econometric Theory* **15**(2): 228-237.

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

von Haefen RH, Phaneuf DJ. 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 RH, Phaneuf DJ. 2005. Kuhn-Tucker demand system approaches to nonmarket valuation. In *Applications of Simulation Methods in Environmental and Resource Economics*, Scarpa R, Alberini AA (eds). Springer, Dordrecht.

Wales TJ, Woodland AD. 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 GY, Zeng L, Cook RJ. 2011. A robust pairwise likelihood method for incomplete longitudinal binary data arising in clusters. *Canadian Journal of Statistics* **39**(1): 34-51.

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

Table 1. 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 2. Joint Model Estimation Results - Event Type Choice Model Component

Variables	Coefficient Estimate	Coefficient t-stat	Standard Deviation Estimate	Standard Deviation 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 3. 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