A Closed-Form Multiple Discrete-Count Extreme Value (MDCNTEV) Model

Chandra R. Bhat
The University of Texas at Austin
Department of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St, Stop C1761, Austin, TX 78712, USA
Tel: 1-512-471-4535; Email: bhat@mail.utexas.edu
and
The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

ABSTRACT
In this paper, we propose a new two-stage budgeting-based utility-theoretic econometric multiple discrete-count model based on the linking of a fractional split MDCEV model component with a total count model. Through the strategic specification of error distributions in the model, we derive a multiple discrete-count extreme value (MDCNTEV) model that has a closed-form probability expression and that is estimable using straightforward maximum likelihood estimation. An application of the proposed model is demonstrated in the context of individuals' multivariate count of recreational episodes to each of multiple possible tourism destination locations. The results highlight the promise of the proposed model for a variety of multivariate count consumer choice settings. The model can also serve as a base model over which random heterogeneity may be superimposed to specify more advanced models.

Keywords: Multivariate count, MDCNTEV model, discrete-count, two-stage budgeting, utility theory, tourism trip generation and destination choice.

1. INTRODUCTION

Multiple discrete-continuous models (MDC) address choice situations that involve a portfolio or package choice of multiple elemental alternatives, along with the continuous intensity of consumption for the chosen elemental alternatives. Unlike typically analyzed single discrete choice models in the literature that are applicable for the case of the choice of a single alternative from a mutually exclusive set of elemental alternatives, the distinguishing characteristic of an MDC model is that it allows the choice of multiple alternatives simultaneously. Bhat (2005, 2008) developed the multiple discrete-continuous extreme value (MDCEV) model, which has since seen substantial use in different fields, thanks to a simple closed-form probability expression for the consumption pattern (see, for example, Ma and Ye, 2019, Shin et al., 2019, Varghese and Jana, 2019, and Mouter et al., 2021). The MDCEV model has its origins in a constrained direct utility maximization approach that is solved using the Karush-Kuhn-Tucker (KKT) first-order necessary conditions of optimality to obtain the consumption probability expression. However, extending the KKT approach to consider multiple discrete-count models has been rather elusive, because of the change from a continuous quantity to a discrete (indivisible) count quantity for the intensity of consumption. Instead, the two broad and commonly used approaches in the literature (spanning many different fields, such as the tourism, environmental economics, recreation, actuarial, transportation, ecology, and Biostatistics and epidemiology fields, to name just a few) to model multiple discrete-count data are the multivariate count model approach and a vertical accumulation-based choice approach, each of which is discussed in turn next.

1.1. Multivariate Count Models

A multivariate count model may be developed using multivariate versions of the Poisson or negative binomial (NB) discrete distributions (such as the negative multinomial, but these only allow positive correlations in the counts), or using the multinomial distribution directly (such as the multinomial, Dirichlet-Multinomial, and random-clumped multinomial that allow only negative correlation in the counts). The reader is referred to Zhang et al., 2017, Inouye et al., 2017, and Peyhardi et al., 2021 for recent reviews of these types of models and their compound distribution extensions. These multivariate count models have the advantage of a closed form probability expression, but become cumbersome as the level of multivariateness increases.[1] Of course, one can superimpose additional mixing structures over these basic models in the parameterization of the mean or other parameters in the model, relaxing the strict one-sided dependence between the counts, but such mixing approaches can also get unwieldy in the presence of several mixing components. Further, 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 statistical "stitching" devices to accommodate correlations in

[1] A similar-in-spirit approach to the multivariate count models is the Potts model (an extension of the binary Ising model), originating in the statistical mechanics field (see Potts, 1952, McCoy and Wu, 1973, and Razaee and Amini, 2020). The multivariateness in counts in such a model is accommodated through a set of interaction terms within a discrete multivariate distribution. However, such models are profligate in parameters, unless *a priori* restrictions are imposed on the nature and number of interaction terms introduced to engender multivariateness.

the multivariate counts. Thus, these models are not suitable for economic welfare analysis, an important issue in many recreational, cultural, and other empirical contexts. Additionally, these models do not allow for potentially complex substitution and income effects in consumer choice decisions. For example, an increase in the entry fee to a water park (say A) may result in an increase in the attractiveness of other water parks due to a substitution effect, but also a decrease in total water park recreation episodes as a whole because of an income effect. So, while the annual frequency of visit instances to water park A will reduce, the frequency of visit instances to other water parks may increase or decrease. The multivariate count models do not explicitly account for such substitution and income effects. Finally, in such multivariate count models, each count is modeled by a separate equation, requiring adequate sample points to estimate the parameters in each count equation. With many count variables, the sample size can whittle down quickly, and will also lead to a higher prevalence of zero counts, which then necessitates techniques to accommodate excess zeros in the count for each event category. Such techniques further increase the number of parameters to be estimated, and can also lead to challenging estimation problems.

1.2. Vertical Accumulation-Based Choice Approach

The second approach, vertical accumulation-based choice, essentially presumes a single discrete choice model at each of several choice occasions (by single discrete choice, we refer to the situation where, at each of the choice occasions, the individual is assumed to select one and only one alternative from the available set of outcome categories). However, preferences may vary within the same individual across multiple choice instances, leading to the observation of multiple discreteness over the vertical stream of choice instances. Such models, by themselves, do not explicitly model the total count of positive participation/consumption decisions (for example, the total number of annual vacation/recreational trips, or the total number of monthly purchases within a product category of goods, such as coffee or pizza). This total count is determined in one of three broad ways, as discussed in turn in the next few paragraphs.

The first method within the vertical accumulation-based choice approach is applicable when the choice made at each of several observed instances of participation/consumption (such as the destination visited in each of several vacation trips) is available in the form of longitudinal data. In this case, each instance (say of a vacation trip) is considered as a single discrete choice occasion, assuming independence across the single discrete choice occasions or using unobserved heterogeneity and state dependence effects across the choice occasions. In this approach, only those with a positive participation/consumption decision are considered, leaving out the decision of whether to participate/consume at all or not in the product category of interest. However, the approach may be combined with a separate total count model, with or without a linkage up from the single discrete choice model to the total count model. The former is not consistent with two-stage budgeting utility theory, while the latter (with the linkage term appropriately developed) can be shown to be consistent with two-stage budgeting utility theory (see Hausman et al., 1995, Mannering and Hamed, 1990, Rouwendal and Boter, 2009, Bhat et al., 2015, and Wiśniewska et al., 2020).

The second and third methods within the vertical accumulation-based choice approach are used when the actual sequence of choices made are not observable, but only the total count of positive participation/consumption occasions and the allocation of that total count to each of many alternatives are observed. The second method, very similar to the first method, considers a vertical string of single discrete choice instances. But since only the aggregate count across all choice occasions is observed (rather than information on the ordering of the choice instances), the probability of the observed counts for each alternative, given the total count, takes a multinomial distribution form (see Terza and Wilson, 1990). An added layer of unobserved heterogeneity may be considered by assuming that the single discrete choice probabilities for any alternative (at any choice instance) is itself Dirichlet distributed, which then leads to a more flexible Dirichlet-Multinomial (DM) model (see Murteira and Ramalho, 2016). Then, a total count model can be estimated, with or without linkage up from the multinomial/DM model.

In the third method, the analyst pre-fixes a certain number of "synthetic" choice occasions, at each of which an individual is assumed to make a single discrete choice among the alternatives of deciding not to participate/consume or choosing to select one of multiple possible choice alternatives. Typically, the single discrete choice at one choice occasion is assumed to be independent of the single discrete choices at other occasions, so that, effectively, the choice for each alternative at each synthetic choice occasion is constructed as a series of "0" or "1" dependent variables (with the sequence being immaterial) such that the observed aggregate shares of alternatives is preserved across choice occasions (though sometimes, the observed aggregate shares are represented as synthetic choices in a specific sequence over the total choice instances, with the possibility of accommodating state dependence effects). Examples include Morey et al., 1993 (who assumes an arbitrary total count (budget) value of 50 fishing trips in each fishing season for each individual), Paleti et al., 2014 (who set the number of synthetic choice occasions in a "vertical" vehicle body/fuel type choice setting as the number of driving age members in the household plus 2 to exploit the fact that the number of vehicles owned by a household is virtually never greater than the number of driving age members plus two), and Hendel, 1999 and Dubé, 2004 (who consider the case of "multiple discreteness" in the purchase of multiple brands within a particular consumer product category at a single shopping instance as the result of a stream of expected, but unobserved to the analyst, future consumption occasions between successive shopping purchase occasions, and determine the number of such total consumption occasions using an independent count model). In this third approach, a zero count is assumed to implicitly emerge though a string of non-participation decisions across all the synthetic choice occasions.

1.3. The Current Paper

The current paper uses a different "Horizontal Choice" approach to multiple discrete-count modeling relative to the vertical accumulation-based approach. In particular, our approach views the fractional splits among the many choice alternatives as arising from a fundamentally "at-once" horizontal choice, rather than on the premise that the multivariate discreteness and count data are the result of the "vertical accumulation" of single discrete choice decisions at each of the many choice instances. The vertical process considers that the goods under consideration for consumption within the product category of interest are perfect substitutes of one another, and does not consider the possibility that the observations may originate from a "horizontal" choice of multiple alternatives that are viewed as being imperfect substitutes of one another. Indeed, earlier theoretical studies of motivations for leisure travel and leisure product consumption indicate that individuals and families do not make disconnected instance-specific decisions about each leisure trip, but plan for such trips over longer periods of time such as a month or even a year; in doing so, they seek a social-psychological sense of "optimal arousal" in consumption patterns based on stability (psychological security) as well as change (novelty) (see Woodside and Lysonski, 1989, and Iso-Ahola, 1983), thus supporting the notion of a horizontal choice process. That is, even as the environmental stimulus of consumption patterns may bring stability and comfort, consumers also appreciate that the expected familiarity of the same experience (as encoded in past mental information) can lead to boredom and a lack of adventure/novelty. Thus, there is an intentional and purposeful "horizontal at-once" choice process at play right from the outset in consumption patterns that reflects satiation behavior, based on the recognition that different alternatives serve different functionalities and provide different kinds of utility "arousal" (see LaMondia et al., 2008 for an extended discussion). For example, in the recreational literature, the focus may be on how individuals decide to allocate their annual fishing trips to different angler sites based on different characteristics of the sites, or how consumers, over an extended time horizon, choose to allocate a fixed number of vacation trips to different vacation destinations. Similarly, in the transportation field, the focus may be on the body/fuel type composition split (for example, split between hybrid SUV, electric sedan, and gasoline pickup vehicles) of the count of vehicles in a household, or the split of working days in a month that individuals work from home versus from a regular work place versus from a third work place. In all these choice situations, the alternatives may be considered as imperfect substitutes for one another. That is, individuals are likely to be variety-seekers, who appreciate the value of different functional needs being satisfied through investing in different alternatives. Equivalently, satiation effects set in as the investment in any single alternative increases.

Based on the above discussion, we formulate a new econometric multiple discrete-count model. There are two components to our proposed model. The first component is a total count model, framed within a generalized ordered-response (GOR) framework, with a linking function from the second fractional split MDCEV model component appearing in this total count model. In this second MDCEV component, the discrete component corresponds to whether or not an individual chooses (consumes) a specific alternative over the total count, and the continuous component refers to the fractional split of investment in each of the consumed alternatives over the total count. This formulation explicitly accommodates a non-linear utility function (with decreasing marginal utility at higher fractional consumption levels) for each alternative. This second component is relevant only if the observed total count is non-zero. Specifically, we start from Bhat's (2008) utility formulation for the case when only inside goods are present (with the budget of the fractions across all alternatives being equal to one), but adopt a reverse Gumbel distributional assumption for the stochastic terms in the baseline preferences of each of the alternatives. We also consider a reverse Gumbel distribution for the random error term within the GOR framework of the total count model, as well as reverse Gumbel error terms to include unobserved heterogeneity in the linking function. With these assumptions, we obtain a closed-form multiple discrete-count extreme value (MDCNTEV) model, which is shown to be consistent with a two-stage budgeting utility-theoretic framework. Unlike the two-stage budgeting model proposed recently by Bhat (2022) for the case of multiple discrete-continuous data, the first stage model here is a count model rather than a Tobit model. The count model poses interesting challenges, necessitating appropriate assumptions and formulations that continue to keep the resulting model within a closed-form formulation. The formulation we propose here, to our knowledge, is a first of its kind in the count model literature.[2]

2. MODEL FORMULATION AND STATISTICAL SPECIFICATION

2.1. Reverse Gumbel MDCEV Model of Fractional Split

In this section, we start with the second stage of the allocation among inside goods within the product category under consideration (say product category *G*). We use the MDCEV framework as a fractional allocation model. This model comes into play only if there is positive count consumption in the product category, as determined in the first stage count model. Thus, the fractional model is estimated only for those with a positive count in the product category, Consider the following constrained direct utility form from Bhat (2008):

$$U(\tilde{\mathbf{f}}) = \sum_{k=1}^{K} \gamma_k \psi_k \ln\left(\frac{\tilde{f}_k}{\gamma_k} + 1\right) \tag{1}$$

$$s.t. \; \sum_{k=1}^{K} \tilde{f}_k = 1,$$

where $\tilde{f}_k$ is the fraction of the total count (of consumption on the product group under consideration) allocated to the inside good *k* (an inside good is any good within the product group under consideration). In the above utility function, $U(\tilde{\mathbf{f}})$ is a quasi-concave, increasing, and continuously differentiable function with respect to the fractional consumption quantity ($K\times1$)-vector $\tilde{\mathbf{f}}$ ($0 \le \tilde{f}_k \le 1$ for all *k*), and ψ_k and γ_k are parameters associated with good *k*. The function

[2] Our model is different from those of Lee and Allenby, 2014 and Kuriyama and Hanemann, 2006. These earlier studies assume away stochasticity in the baseline utility preference for the outside good, while we explicitly consider the more realistic case that there could be individual-level unobserved variations in the baseline preference for all goods, inside and outside (see Bhat, 2008 for an extended discussion). Further, the integer programming approach of the earlier papers does not produce a closed-form solution as our model does. Also, both studies use a linear utility structure for the outside good, which does not guarantee the positivity of the outside good (see Bhat et al., 2022 and Saxena et al., 2022). Finally, both these earlier studies are based on a Hicksian composite good approach, That is, by defining the goods of interest as inside goods, changes in exogenous variables directly impact the consumptions of these inside goods (even if the true effect is an indirect impact through budget changes), co-mingling strict budget effects (that is, income effects) and strict allocation effects (that is, substitution effects).

$U(\tilde{\mathbf{f}})$ in Equation (1) is a valid utility function if $\psi_k > 0$, and $\gamma_k > 0$ for all k (we will use the terms "good" and "alternative" interchangeably to refer to any good k). As discussed in detail in Bhat (2008), ψ_k represents the baseline marginal utility, and γ_k is the vehicle to introduce corner solutions (that is, zero fractional splits) for the goods, but also serves the role of a satiation parameter (higher values of γ_k imply less satiation). The satiation operates at the fractional split level, so that the marginal utility of a good decreases as the fractional split investment in the good increases.

To ensure the non-negativity of the baseline marginal utility, while also allowing it to vary across individuals based on observed and unobserved characteristics, ψ_k is parameterized as follows:

$$\psi_k = \exp\left(\boldsymbol{\beta}'\mathbf{z}_k - \frac{1}{\sigma}\ln p_k + \varepsilon_k\right) = q_k \exp\left(\varepsilon_k\right);\ q_k = \exp(\boldsymbol{\beta}'\mathbf{z}_k)\left(\frac{1}{p_k^{1/\sigma}}\right),\ k = 1, 2, ..., K, \tag{2}$$

where $\mathbf{z}_k$ is a set of attributes that characterize alternative k and the decision maker (including a constant), p_k is the unit price for good k, the negative inverse of σ (σ>0) is the coefficient on $\ln p_k$, and ε_k is a standardized scale error term that captures the idiosyncratic (unobserved) characteristics that impact the baseline utility of good k. As should be obvious, the parameter σ is estimable only if there is price variation across the inside goods (σ may be arbitrarily normalized to one when there is no price variation). q_k in Equation (2) corresponds to the systematic part of the baseline utility of alternative k. Assume a standard extreme value type 1 (Gumbel) distribution based on the limiting distribution of the minimum of random variables for the ε_k terms:

$$f_{\varepsilon_k}(t) = e^{-e^t}.e^t,\ F_{\varepsilon_k}(t) = \text{Prob}(\varepsilon_k < t) = 1 - e^{-e^t},\ \text{and}\ S_{\varepsilon_k}(t) = \text{Prob}(\varepsilon_k > t) = e^{-e^t}\ \text{for}\ k = 1, 2, 3, ..., K. \tag{3}$$

We can then write the joint multivariate survival distribution function (SDF) for the error terms $\xi_k = \varepsilon_k - \varepsilon_1$ as follows[3]:

$$\boldsymbol{S}_{\xi}(t_2, t_3, ..., t_K) = \text{Prob}(\xi_2 > t_2, \xi_3 > t_3, ..., \xi_K > t_K) = \frac{1}{\left(1 + \sum_{k=2}^{K} e^{t_k}\right)}. \tag{4}$$

The multivariate cumulative distribution function (CDF) of the $\boldsymbol{\xi} = (\xi_2, ..., \xi_K)$ vector can be written as a function of the SDFs corresponding to the random variates as follows:

$$\boldsymbol{F}_{\xi}(t_2, t_3, ..., t_K) = \text{Prob}(\xi_2 < t_2, \xi_3 < t_3, ..., \xi_K < t_K) = 1 + \sum_{D \subset \{2,...K\}, |D| \geq 1} (-1)^{|D|} \boldsymbol{S}_D(\mathbf{t}_D), \tag{5}$$

[3] The ξ_k error terms (k=2,3,…,K) are essentially multivariate logistically distributed with a correlation of 0.5, with the SDF expression as given below (see Appendix A of Bhat et al., 2022).

where $\boldsymbol{S}_D(.)$ is the SDF of dimension *D*, *D* represents a specific combination of the ξ terms (representing a specific sub-vector of the $\boldsymbol{\xi}$ vector; there are a total of $(K-2)+C(K-2,2)+C(K-2,3)+...C(K-2,K-2)=2^{K-2}-1$ possible combinations, |*D*| is the cardinality of the specific combination *D*, and $\mathbf{t}_D$ is a sub-vector of the vector $\mathbf{t}=(t_2,t_3,...,t_K)$ with the appropriate elements corresponding to the combination *D* extracted.

To find the optimal fractional splits of goods, the Lagrangian is constructed and the first order equations are derived based on the Karush-Kuhn-Tucker (KKT) conditions. In particular, designating activity purpose 1 as a purpose to which the individual allocates some non-zero fraction of consumption, the probability expression for the fractional allocation pattern where the first *M* goods are consumed at levels $\tilde{f}_k^*$ ($k=2,3,...,M$), and $\tilde{f}_1^*=1-\sum_{k=2}^{M}\tilde{f}_k^*$ is (see Bhat, 2022):

$$=|J|(M-1)!\left[\frac{\exp\left(\sum_{i=1}^{M}V_k\right)}{\left(\sum_{k=1}^{M}\exp(V_k)\right)^M}+\sum_{D\subset\{M+1,M+2,...,K\},|D|\geq 1}(-1)^{|D|}\frac{\exp\left(\sum_{i=1}^{M}V_k\right)}{\left(\sum_{k=1}^{M}\exp(V_k)+\sum_{k\in D}\exp(V_{k0})\right)^M}\right]$$

$$\text{where } |J|=\left(\prod_{i=1}^{M}c_i\right)\left(\sum_{i=1}^{M}\frac{1}{c_i}\right),\ \text{where}\quad c_i=\left(\frac{1}{\tilde{f}_i^*+\gamma_i}\right),\ \text{and} \tag{6}$$

$V_k=-\boldsymbol{\beta}'\mathbf{z}_k+\frac{1}{\sigma}\ln p_k+\ln\left(\frac{\tilde{f}_k^*}{\gamma_k}+1\right)$ (k = 1, 2, 3,…, K), and $V_{k0}=-\boldsymbol{\beta}'\mathbf{z}_k+\frac{1}{\sigma}\ln p_k$ (k = 1, 2, 3,…, K).

The probability that all the inside goods are consumed at fractional levels $\tilde{f}_2^*,\tilde{f}_3^*,...,\tilde{f}_K^*$ is:

$$P\left(\tilde{f}_2^*,\tilde{f}_3^*,...,\tilde{f}_K^*\right)$$

$$=|J|\,\boldsymbol{f}_{\xi}(V_2,V_3,...,V_K)=|J|(K-1)!\frac{\exp\left(\sum_{i=1}^{K}V_k\right)}{\left(\sum_{k=1}^{K}\exp(V_k)\right)^K} \tag{7}$$

The probability that none of the inside goods are consumed is:

$$P\left(0,...,0\right)=1+\sum_{D\subset\{2,...,K\},|D|\geq 1}(-1)^{|D|}\frac{e^{V_{10}}}{\left(e^{V_{10}}+\sum_{k\in D}e^{V_{k0}}\right)}\ . \tag{8}$$

The parameters to be estimated in the model above include the $\boldsymbol{\beta}$ vector, the $\boldsymbol{\gamma}=(\gamma_1,\gamma_2,...,\gamma_K)$ vector, and the σ scalar. However, note that these parameters from the MDCEV fractional model also appear in the total budget model, and hence we discuss the overall estimation procedure in Section 2.3 after first discussing the total budget (in the product group) and the linking approach between the fractional allocation model and the total budget count model.

An important note here that is relevant to the linking. Conditional on positive consumption in the product category *G*, and conditional on the first *M* inside goods being selected for consumption, Bhat (2022) shows that the optimal fractional allocation to each of the consumed goods is given by the formula below:

$$\tilde{f}_k^* = \frac{\psi_k \gamma_k + \gamma_k \sum_{\substack{j=1 \\ j \neq k}}^{M} \gamma_j \left(\psi_k - \psi_j\right)}{\sum_{j=1}^{M} \psi_j \gamma_j}, \; k = 1, ..., M.^4 \tag{9}$$

2.2. Linking with the Fractional Split Model

The basic notion of two-stage budgeting is to have a first-stage count allocation between the product group of interest (product group *G*) and other product groups not of immediate interest to the analyst, without detailed information about the prices of individual goods within each product group (Strotz, 1957). To allow this in a utility consistent manner, we follow a procedure based on the separability of the indirect utility function (see Blackorby et al., 1978, Rouwendal and Boter, 2009, and Fally, 2022 for detailed discussions of indirect separability).[5] Specifically, we begin by considering the following indirect utility function (IUF) of an individual for the choice of consumption within a product group of interest *G* and other consumption goods:

$$v = v\left[E, \rho, \ln b_G(\boldsymbol{\pi})\right], \tag{10}$$

where *E* denotes total expenditure, ρ refers to the prices of other consumption goods outside the product group of interest (referred to typically as outside goods, with one of the goods being a numeraire good), and $b_G(\boldsymbol{\pi})$ serving as an aggregate group price index for product group *G*. We write $b_G(\boldsymbol{\pi})$ as a function of quality adjusted price indices π_k for the inside goods of product category *G*, where $\boldsymbol{\pi} = (\pi_1, \pi_2, ..., \pi_K)$. For this exposition, assume that ψ_k is deterministic and, conditional on consumption in product group *G*, all inside goods are consumed (this latter

---

[4] The derivation of this expression is provided in Appendix B of Bhat (2022). As shown there, the fractional allocations among the consumed goods will sum to one. The expression also guarantees that the predictions f_k for any consumed good will be positive and less than 1. Note also that, among the consumed goods, the fraction increases as ψ_k or γ_k increases (that is, as the baseline preference of good *k* increases or the satiation for good *k* decreases), though the magnitude of this increase is dependent on the values of the baseline preferences and satiation values of the other consumed goods. Also, as discussed in Bhat (2022), the second term in the numerator of this expression is critical to ensure that the KKT conditions of equal marginal utility at the point of actual fractional consumptions of the consumed goods hold.

[5] We assume standard regularity conditions for the indirect utility function, including that it is continuous, nonincreasing and quasiconvex in all prices and total expenditure, and homogeneous of degree zero in all prices and total expenditure. Separability of the indirect utility function implies essentially that we are able to partition the indirect utility function into subgroups, with the indirect utility contribution of each subgroup dependent only on the price of goods within that subgroup and total expenditures.

assumption is only for presentation purposes, and is relaxed later). Consider the following quality-adjusted price index for inside good *k* within the product group *G* of interest:

$$-\pi_k = \frac{1}{1/\sigma}\left[\ln\left(\psi_k\gamma_k + \gamma_k\sum_{\substack{j=1\\ j\neq k}}^{K}\gamma_j\left(\psi_k - \psi_j\right)\right)\right], \text{ or } \left[\exp(\pi_k)\right]^{-\frac{1}{\sigma}} = \psi_k\gamma_k + \gamma_k\sum_{\substack{j=1\\ j\neq k}}^{K}\gamma_j\left(\psi_k - \psi_j\right). \tag{11}$$

The negative sign in front of π_k on the left side of the first equation above is because π_k is a price index; as the price of any inside good *k* (p_k) increases, or as the non-cost systematic baseline utility element for any inside good *k* ($\boldsymbol{\beta}'\mathbf{z}_k$) decreases, ψ_k decreases and π_k increases. Also, as the γ_k parameter for any inside good increases, satiation decreases; that is, the marginal utility of that good does not decrease substantially with consumption. Equivalently, the quality adjusted price index of a good should decrease as the γ_k parameter increases, as is the case with the specific form of Equation (11). Next, consider a constant elasticity of substitution (CES) formulation for the linking function for product group *G* (the linking function is homogenous-of-degree-one):

$$b_G(\boldsymbol{\pi}) = \left[\sum_k\left(\left[\exp\left(\pi_k\right)\right]^{-\frac{1}{\sigma}}\right)\right]^{-\sigma} = \left[\sum_k\left(\psi_k\gamma_k + \gamma_k\sum_{\substack{j=1\\ j\neq k}}^{K}\gamma_j\left(\psi_k - \psi_j\right)\right)\right]^{-\sigma} = \left[\sum_k\psi_k\gamma_k\right]^{-\sigma}. \tag{12}$$

The right side of the above equation results because $\sum_k\left(\gamma_k\sum_{\substack{j=1\\ j\neq k}}^{K}\gamma_j\left(\psi_k - \psi_j\right)\right) = 0$. Also, note that as the quality-adjusted price index for any inside good π_k increases, the group price index $b_G(\boldsymbol{\pi})$ increases, as required. Then, using Q_G and Q_{Gk} for the total consumption in the product group *G* and the consumption of inside good *k* within product group *G*, respectively, we get the following conditional demands using Roy's identity (note that Q_G and Q_{Gk} are purely imaginary quantities and, in general, are not identical to the actual count consumptions y_G and y_{Gk} used later):

$$Q_{Gk} = \frac{\frac{\partial v}{\partial \pi_k}}{\frac{\partial v}{\partial E}} = \frac{\frac{\partial v}{\partial \ln b_G(\boldsymbol{\pi})}}{\frac{\partial v}{\partial E}} \times \frac{\partial \ln b_G(\boldsymbol{\pi})}{\partial \pi_k}, \text{ with } \frac{\partial \ln b_G(\boldsymbol{\pi})}{\partial \pi_k} = \frac{\left[\exp(\pi_k)\right]^{-\frac{1}{\sigma}}}{\sum_k\left(\left[\exp(\pi_k)\right]^{-\frac{1}{\sigma}}\right)}. \tag{13}$$

From above, we immediately get the result that:

$$\sum_k \frac{\partial \ln b_G(\boldsymbol{\pi})}{\partial \pi_k} = 1. \tag{14}$$

Next, we may write the total count Q_G as follows:

$$Q_G = \sum_k Q_{Gk} = \left(\frac{\dfrac{\partial v}{\partial \ln b_G(\boldsymbol{\pi})}}{\dfrac{\partial v}{\partial E}} \right) \left(\sum_k \frac{\partial \ln b_G(\boldsymbol{\pi})}{\partial \pi_k} \right) = \left(\frac{\dfrac{\partial v}{\partial \ln b_G(\boldsymbol{\pi})}}{\dfrac{\partial v}{\partial E}} \right) \tag{15}$$

Finally, substituting Q_G for the right side of the expression above into Equation (13), we get:

$$\begin{aligned} Q_{Gk} &= Q_G \times \frac{\partial \ln b_G(\boldsymbol{\pi})}{\partial \pi_k} = Q_G \times \frac{\left[\exp(\pi_k)\right]^{-\frac{1}{\sigma}}}{\sum_k \left(\left[\exp(\pi_k)\right]^{-\frac{1}{\sigma}} \right)} = Q_G \times \frac{\left[\exp(\pi_k)\right]^{-\frac{1}{\sigma}}}{\sum_k \left(\left[\exp(\pi_k)\right]^{-\frac{1}{\sigma}} \right)} \\ &= Q_G \times \frac{\psi_k \gamma_k + \gamma_k \sum_{\substack{j=1 \\ j \neq k}}^{K} \gamma_j \left(\psi_k - \psi_j \right)}{\sum_k \psi_k \gamma_k} = Q_G \times \tilde{f}_k^*. \end{aligned} \tag{16}$$

Immediately, we note that our proposed model is consistent with two-stage budgeting, with the consumption of inside good k being the product of the first stage total count model multiplied by the second stage fractional MDCEV model. In reality, not all inside goods will be consumed conditional on consumption in product group *G*, in which case the non-zero fractional allocations will be among the consumed *M* inside goods (so the summations in Equation (16) will only be among the consumed first *M* goods). In this situation, $\tilde{f}_k^*$ in Equation (16) takes an identical form as in Equation (9). At the same time, using the concept of non-negative virtual prices, the consistency of the proposed model with two-stage budgeting is still retained even in the case when some inside goods are consumed and other are not (see Bhat, 2022 for a detailed discussion of this point). In summary, our approach is consistent with economic two-stage budgeting. Specifically, in the first stage total count model for consumption in product group *G*, we are able to use an aggregate price index for the product group *G* in the form of the following:

$$\ln b_G(\boldsymbol{\pi}) = -\sigma \ln \left[\sum_k \psi_k \gamma_k \right] \tag{17}$$

Interestingly, and tellingly, if all individuals choose only one inside good to which the total count is allocated (that is, a single discrete choice is at play rather than a multiple discrete situation), the above aggregate price index, in deterministic form, collapses exactly to the consumer surplus (logsum) originating from a simple multinomial logit model, as used by Hausman et al. (1995) as the aggregate price index in their single discrete choice model. Specifically, when there is no satiation, as in a single discrete choice model, γ_k =1 for all inside goods *k* in Equation (17). Then, Equation (17) is exactly Equation (2.1.9) of Hausman et al., again continuing to ignore the stochasticity in ψ_k ($1/\alpha$ in Hausman et al.'s Equation (2.1.9) is the inverse of the price coefficient in the utility function, as is σ in our formulation; see Equation (2)). Thus, our aggregate price index may be viewed as a multiple discrete generalization of the single discrete choice case.

2.3. Total Count in Product Category Model

A key to linking the MDCEV fractional split model to the total count model is our recasting of the count model as a generalized ordered-response model in which a single latent continuous variable spanning the real line is partitioned into mutually exclusive intervals (see Castro et al., 2012 and Bhat et al., 2015). 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 fractional split 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 the recasting of the count model as a special case of the generalized ordered-response model in Section 2.3.1, and then discuss the linkage with the fractional split model in Section 2.3.2. In doing so, we use a customized statistical distributional specification for stochasticity and a specific parameterized form for the threshold values, both of which represent first such introductions in the literature. These customizations are the key to retaining a closed-form model for the resulting combined total count-fractional split model leading up to the proposed multivariate count model.

2.3.1. The Basic Recasting

Define a latent propensity underlying the count variable for consumption in product group G as y_G^* and consider the following structure:

$$y_G^* = -\zeta \ , \ y_G = i \ \text{ if } \Theta_{i-1} < y_G^* < \Theta_i \ , \ \ \Theta_i = f_i(\boldsymbol{\varpi}) + \sum_{l=0}^{i} \alpha_l \tag{18}$$

where ζ is a random error term assumed to be reverse Gumbel distributed with a scale parameter of ϑ. Of course, the scale ϑ is not identified above (as in usual ordered-response models) and may be arbitrary set to the value of 1. The latent count propensity y_G^* is mapped to the observed count variable y_G^* by the thresholds Θ_i, which satisfy the ordering conditions $(i = 0,1,2,...,I; \Theta_{-1} = -\infty; -\infty < \Theta_0 < \Theta_1 < \Theta_2 < ...)$ in the usual ordered-response fashion. $f_i(\boldsymbol{\varpi})$ is a non-linear function of a vector of individual-specific variables $\boldsymbol{\varpi}$ ($\boldsymbol{\varpi}$ includes a constant), and α_l is a scalar similar to the thresholds in a standard ordered-response model ($\alpha_{-1} = -\infty$ and $\alpha_I = 0$ for identification). Defining $S_\zeta^{-1}(\tilde{t}) = \vartheta\left\{\ln\left[-\ln(\tilde{t})\right]\right\}$ as the inverse survival function of the reverse Gumbel with scale ϑ, and $\lambda = e^{\boldsymbol{\mu}'\boldsymbol{\varpi}}$ ($\boldsymbol{\mu}$ is a coefficient vector to be estimated), write $f_i(\boldsymbol{\varpi}) = -S_\zeta^{-1}\left(e^{-\lambda}\sum_{l=0}^{i}\frac{\lambda^l}{l!}\right)$, so that the thresholds in Equation (18) take the following form[6]:

---

[6] The negative sign in $f_i(\boldsymbol{\varpi}) = -S_\zeta^{-1}\left(e^{-\lambda}\sum_{l=0}^{i}\frac{\lambda^l}{l!}\right)$ is because, as the term in the parenthesis increases, $S_\zeta^{-1}\left(e^{-\lambda}\sum_{l=0}^{i}\frac{\lambda^l}{l!}\right)$ decreases. So, to get the thresholds to be in ascending order, we take the negative of the survival inverse. By

$$\Theta_i = -S_\zeta^{-1}\left(e^{-\lambda}\sum_{l=0}^{i}\frac{\lambda^l}{l!}\right)+\sum_{l=0}^{i}\alpha_l \text{ , with } \alpha_l = 0 \text{ if } l > L^* , \tag{19}$$

where L^* is an appropriate count level that may be determined based on the empirical context under consideration and empirical testing. The presence of the α_l terms $\left(\boldsymbol{\alpha} = \alpha_0, \alpha_1, \alpha_2, ..., \alpha_{L^*}\right)$ provides flexibility to accommodate high or low probability masses for specific count outcomes without the need for using hurdle or zero-inflated mechanisms. The probability of observing an individual with a count of *i* may be obtained as follows:

$$\begin{aligned} \text{Prob}[y_G = i] &= \text{Prob}\left[\Theta_{i-1} < \left(-\zeta\right) < \Theta_i\right] \\ &= \text{Prob}\left[-\Theta_{i-1} > \left(\zeta\right) > -\Theta_i\right] \\ &= \exp\left[-\exp(-\Theta_i)\right] - \exp\left[-\exp(-\Theta_{i-1})\right] \end{aligned} \tag{20}$$

The typical Poisson count model may be obtained as a restrictive specification of the generalized ordered-response model. To see this, assumes all α_l terms are zero. Then,

$$\begin{aligned} P\ [y_G = i] &= P\left[\ -S_\zeta^{-1}\left(e^{-\lambda}\sum_{l=0}^{i-1}\frac{\lambda^l}{l!}\right) < \left(y_G^* = -\zeta\right) < -S_\zeta^{-1}\left(e^{-\lambda}\sum_{l=0}^{i}\frac{\lambda^l}{l!}\right)\right] \\ &= P\left[\ S_\zeta^{-1}\left(e^{-\lambda}\sum_{l=0}^{i-1}\frac{\lambda^l}{l!}\right) > \left(\zeta\right) > S_\zeta^{-1}\left(e^{-\lambda}\sum_{l=0}^{i}\frac{\lambda^l}{l!}\right)\right] \\ &= S_\zeta\left(S_\zeta^{-1}\left(e^{-\lambda}\sum_{l=0}^{i}\frac{\lambda^l}{l!}\right)\right) - S_\zeta\left(S_\zeta^{-1}\left(e^{-\lambda}\sum_{l=0}^{i-1}\frac{\lambda^l}{l!}\right)\right) = \frac{e^{-\lambda}\lambda^i}{i!} = \frac{e^{-e^{\boldsymbol{\mu}'\boldsymbol{\varpi}}}\left(e^{\boldsymbol{\mu}'\boldsymbol{\varpi}}\right)^i}{i!} \end{aligned} \tag{21}$$

2.3.2. Linking with the Fractional Split Model

The linking function refers to the way the utility maximization process underlying the fractional split model gets incorporated within the count model. An improvement in the quality or a reduction in price of any alternative in the MDCEV model gets manifested as an increase in overall utility (or consumer surplus), resulting in a higher propensity for the consumer product under consideration and a shift toward a high total count of units purchased in the product category. For this purpose, we introduce the function $\ln\left[b_G(\boldsymbol{\pi})\right]$ as developed from the fractional split model into the count model of Equation (18) as follows:

$$y_G^* = -\tilde{\vartheta}\ln b_G(\boldsymbol{\pi}) - \zeta \text{ , } y_G = i \text{ if } \Theta_{i-1} < y_G^* < \Theta_i \text{ , } \Theta_i = f_i(\boldsymbol{\varpi}) + \sum_{l=0}^{i}\alpha_l \text{ ,} \tag{22}$$

---

convention, we will designate $\left(e^{-\lambda}\sum_{l=0}^{-1}\frac{\lambda^l}{l!}\right) = 0,$ so that $\Theta_{-1} = -S_\zeta^{-1}\left(e^{-\lambda}\sum_{l=0}^{-1}\frac{\lambda^l}{l!}\right) = -S_\zeta^{-1}\left(0\right) = -\infty.$ Also, note that rather than use a generalized version of the Poisson model as we use here, it is possible to use a generalized version of a negative binomial model too, as considered by Bhat et al. (2016a). We leave this for further exploration in future studies.

where $\tilde{\vartheta}$ is a scalar link parameter ($\tilde{\vartheta}$>0), and ζ is a random variable as earlier capturing the effects of unobserved variables (as required, as the price of any inside good increases, $\ln b_G(\boldsymbol{\pi})$ increases, and the propensity of consumption in the product group G, y_G^*, decreases). We continue to use $f_i(\boldsymbol{\varpi}) = -S_\zeta^{-1}\left(e^{-\lambda}\sum_{l=0}^{i}\frac{\lambda^l}{l!}\right)$. We next re-introduce stochasticity within the aggregate price index, but rather than use the same error terms ε_k of the fractional split model in the ψ_k terms from Equation (2), we introduce independent error terms τ_k because the linking term needs to be strictly upwards from the second stage to the first stage, and needs to be exogenous to the total count equation (see Bhat, 2022 for a discussion). For similar reasons, as also in Hausman et al. (1995), the total count cannot be introduced as a variable in the second stage fractional MDCEV model. So, we use independently distributed standard reverse-Gumbel error terms τ_k (different from, and independent of, ε_k) to moderate the effect of the systematic baseline preferences from the fractional model in the budget model, and write an empirical specification version of Equation (17) as:

$$\ln b_G(\boldsymbol{\pi}) = -\sigma \ln\left[\sum_k \left\{q_k \exp(\tau_k)\right\}\gamma_k\right]. \tag{23}$$

The use of τ_k that is different from ε_k in the linking function also breaks any correlation based on unobserved factors between the total count and the fractional split model, allowing for a closed-form and simple structure for the proposed MDCNTEV model. Similar to our study, earlier two-stage models assume independence between the two stages conditional on observed characteristics, by constructing the linking function as being purely deterministic; but, instead of using the traditional deterministic form, we introduce the error terms τ_k in our linking function, which immediately accommodates heteroscedasticity in the total count model (as discussed later in this section). Then, by retaining independence (conditional on observed characteristics), we are able to write the choice probabilities in our model as a product of the probability from the first stage count model and the probability from the second stage fractional split model, as we now set out to do.

To obtain the probability of observing each count value, one needs the distribution of the underlying latent variable y_G^* after adding the linking function as above. To obtain this distribution, we re-write y_G^* using Equation (18) as follows:

$$y_G^* = \vartheta \ln\left(\sum_{k=1}^{K} q_k \gamma_k \exp(\tau_k)\right) - \zeta = -\left[\zeta - \vartheta \ln\left(\sum_{k=1}^{K} a_k \exp(\tau_k)\right)\right] = -\eta,\ a_k = q_k \gamma_k,$$

with $\eta = \left[\zeta - \vartheta \ln\left(\sum_{k=1}^{K} a_k \exp(\tau_k)\right)\right]$ and $\vartheta = \tilde{\vartheta}\sigma.$ (24)

Of course, once again the parameter ϑ influencing the variance of η is not identifiable, and may be arbitrarily set to one. The implied linking parameter, $\tilde{\vartheta}$, is then $\tilde{\vartheta} = \frac{\vartheta}{\sigma} = \frac{1}{\sigma}$. As importantly, the distribution function of η takes a surprisingly elegant form with the survival distribution as follows:

$$S_\eta(t) = \text{Prob}(\eta > t) = \frac{1}{\prod_{k=1}^{K}\left(1 + a_k e^{t/\vartheta}\right)} \tag{25}$$

Bhat (2022) recently studied the properties of the random variable η, which he labels as having a minLogistic distribution. This distribution is unimodal (which facilitates stable estimation), with mean and variance as follows (with ϑ set to one for identification):

$$E(\eta) = -\left[\sum_{k=1}^{K} \frac{a_k^{K-1} \ln(a_k)}{\prod_{\substack{j=1 \\ j\neq k}}^{K}\left(a_k - a_j\right)}\right] \text{ and}$$

$$Var(\eta) = E(\eta^2) - \left[E(\eta)\right]^2 = \left[\sum_{k=1}^{K}\left(\frac{3a_k^{K-1}\ln^2(a_k) + \pi^2 a_k^{K-1}}{3 \times \prod_{\substack{j=1 \\ j\neq k}}^{K}\left(a_k - a_j\right)}\right)\right] - \left[E(\eta)\right]^2 \tag{26}$$

Thus, introducing the stochastic linking function into the underlying count propensity variable y_G^* leads to a change in the mean of the underlying propensity variable y_G^* as well as heteroscedasticity across individuals (because the value of a_k varies across individuals), while also changing the distributional shape of the y_G^* variable from a reverse-Gumbel to a minLogistic distribution.[7] In particular, as also indicated earlier, as the price of any inside good k (p_k)

[7] If the linking function is purely deterministic (that is, $\ln b_G(\boldsymbol{\pi}) = -\sigma \ln\left[\sum_k q_k \gamma_k\right]$, or $E(\ln b_G(\boldsymbol{\pi}))$ is used, the y_G^* distribution would be the same as without the linking (that is, the y_G^* distribution would remain reverse-Gumbel as in

decreases, or as the non-cost systematic (log) baseline utility element for any inside good k ($\boldsymbol{\beta}'\mathbf{z}_k$) increases, the value of a_k increases, $E(\eta)$ decreases, and the count propensity in group G, y_G^*, increases. The probability of observing an individual with a total count of g, where g is a specific value of count, may be obtained as follows:

$$\begin{aligned}\text{Prob}[y_G = g] &= \text{Prob}\left[\Theta_{g-1} < (-\eta) < \Theta_g\right] \\ &= \text{Prob}\left[-\Theta_{g-1} > (\eta) > -\Theta_g\right] = S_\eta(-\Theta_g) - S_\eta(-\Theta_{g-1}) \\ &= \left[\frac{1}{\prod_{k=1}^{K}\left(1 + a_k e^{-\Theta_g}\right)}\right] - \left[\frac{1}{\prod_{k=1}^{K}\left(1 + a_k e^{-\Theta_{g-1}}\right)}\right].\end{aligned} \tag{27}$$

The expression above for total count is dependent on both the fractional split model as well as count model parameters (the fractional split parameters are embedded in the a_k elements).[8]

2.4. Estimation Technique

As already discussed, our overall model of total count and fractional split is consistent with a two-stage budgeting approach within a utility-theoretic planning framework. The net econometric consequence for estimation purposes is that the total count model can be separately analyzed in a first stage (as long as $\ln b_G(\boldsymbol{\pi})$ is introduced at this first stage), and the fractional split among the

the unlinked model). Thus, a pure deterministic linking function would nest the unlinked model, while our proposed stochastic linking function would not nest the unlinked model.

[8] Note that, with price variation, $(1/\sigma)$ is the ln(price) coefficient in the fractional MDCEV model which is estimable in both our model and an unlinked model. But, in the unlinked model, the term $-\tilde{\vartheta}\ln b_G(\boldsymbol{\pi})$ does not appear in the latent propensity y_G^*, while this term appears in our linked model (see Equations (18) and (22)). As discussed earlier, in our econometric formulation, the addition of the term $-\tilde{\vartheta}\ln b_G(\boldsymbol{\pi})$ in the linked model leads to a different distributional form for y_G^* than in the unlinked form, and contributes to a mean shift (that varies across individuals based on the fractional allocation parameters embedded in $b_G(\boldsymbol{\pi})$) as well as generates heteroscedasticity across individuals in y_G^*. However, the parameter ϑ is not estimable in our stochastically-linked formulation, and is set to one. This has the result of (innocuously) setting the linking parameter $\tilde{\vartheta}$ in Equation (22) to be the same as the price coefficient $(1/\sigma)$ in the fractional MDCEV model. Of course, if, as discussed in footnote (6), a deterministically-linked formulation is considered, then the linking parameter $\tilde{\vartheta}$ does not get entangled with the variance of y_G^*, and can be estimated separately from the price coefficient $(1/\sigma)$ estimated in the fractional MDCEV model. This alternative deterministically-linked formulation would nest the unlinked formulation, and add an additional parameter relative to the unlinked formulation, but without contributing to possible heteroscedasticity across individuals in the total count model. Our stochastically-linked formulation, on the other hand, may be viewed as a parsimonious specification that allows linkage with both a mean and variance shift in the count model without adding any more parameters than in the unlinked model (and still maintains a closed-form probability expression). Of course, this also does not rule out the occasional possibility that the unlinked model or the deterministically-linked model will perform better than our proposed stochastically-linked model in specific empirical situations, in which case the analyst will have to decide whether to pursue a good data fit or pursue a more behaviorally plausible case of heterogeneity in the linking effect.

many alternative k $(k = 1, 2, ..., K)$ can be separately analyzed in a second stage. 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 $\ln b_G(\boldsymbol{\pi})$ variable) and the likelihood of the MDCEV model conditional on a non-zero count for the product group. For the case of zero count to the product group, the likelihood function is:

$$L(\boldsymbol{\alpha}, \boldsymbol{\mu}, \sigma) = \text{Prob}[y_G = 0] = S_\eta(-\Theta_0) - S_\eta(-\Theta_{-1}) = S_\eta(-\Theta_0) - S_\eta(-\infty)$$
$$= \left[\frac{1}{\prod_{k=1}^{K} \left(1 + a_k e^{-\Theta_0}\right)} \right]. \tag{28}$$

The likelihood function for a specific individual having a count value of g ($g > 0$), and consuming the first M goods at levels $\tilde{f}_k^*$ ($k = 2, 3, ..., M$), and $\tilde{f}_1^* = 1 - \sum_{k=2}^{M} \tilde{f}_k^*$ may be obtained from Equations (6) and (27) as follows:

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\mu}, \sigma) &= \text{Prob}[y_G = g, \tilde{f}_2^*, \tilde{f}_3^*, ..., \tilde{f}_M^*, 0, 0..., 0],\ g > 0 \\ &= \left\{ \text{Prob}[y_G = g] \times \text{Prob}[\tilde{f}_2^*, \tilde{f}_3^*, ..., \tilde{f}_M^*, 0, 0..., 0] \right\},\ g > 0 \end{aligned} \tag{29}$$

The second line in the above equation is a result of the error term η in the count model being independent of the error vector $\boldsymbol{\varepsilon} = (\varepsilon_2, \varepsilon_3 ..., \varepsilon_K)'$ in the fractional MDCEV model. However, note that parameters of the fractional MDCEV model also appear in the count model probability. Similar equations as Equation (29) may be written for the case of a total count value of g ($g > 0$) with fractional allocation to all the inside goods (based on Equation (7)) and the case of a total count value of g ($g > 0$) and the fractional allocation to only one inside good (based on Equation (8)).

The above model formulation does not consider unobserved heterogeneity in the sensitivity to exogenous variables beyond that introduced through the τ_k terms in the linking function. Such additional unobserved heterogeneity may be introduced in a straightforward way by allowing the $\boldsymbol{\beta}$ and $\boldsymbol{\mu}$ parameters to be randomly distributed. For example, assuming a specific continuous parametric distribution (say $\vec{f}$) for $\boldsymbol{\varphi} = (\boldsymbol{\beta}', \boldsymbol{\mu}')'$ with an underlying parameter vector $\tilde{\boldsymbol{\varphi}}$, the likelihood function for the case of a count value of g, and the fractional allocation to the first M inside goods in the product group, is:

$$L(\tilde{\boldsymbol{\varphi}}, \boldsymbol{\alpha}, \sigma) = \int_{\boldsymbol{\varphi}} \left[L(\boldsymbol{\alpha}, \boldsymbol{\varphi}, \sigma) \times 1(g = 0) + L(\boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\varphi}, \sigma) \times 1(g > 0) \right] \left\{ \vec{f}(\boldsymbol{\varphi}; \tilde{\boldsymbol{\varphi}})\, \mathbf{d}\boldsymbol{\varphi} \right\}. \tag{30}$$

where $1(g = 0)$ denotes the occurrence of a zero total count and $1(g > 0)$ denotes the occurrence of a non-zero count. The log-likelihood function may be developed across all individuals, and the parameters may be estimated using a maximum likelihood estimation approach (or a maximum simulated likelihood approach in the case of the model with unobserved heterogeneity effects).

2.5. Forecasting Approach

Once estimated, the model may be used to forecast the total count investment in the product group of interest (including zero count) and the counts of individual good consumptions within the product group of interest for a non-zero total count. The split of the total count is obtained from the fractional MDCEV model. However, simply applying the fractions to a positive integer count would not provide an integer value for the counts for individual goods in the product category. While these non-integer values for the intensity of consumptions of individual goods may be viewed as expected values over the group of individuals with a certain vector set of characteristics, it may be desirable to have a forecasting mechanism that predicts, even at a specific individual level, integer counts for the individual good-level consumptions. An approach to accomplish this is presented below:

- Step 1: Compute the count probability of each count value for the product group G using Equation (27). In doing so, if there is a typical upper limit exhibited by the empirical data, set the upper bound value for this upper limit to ∞. Stack these probabilities consecutively starting from the probability of zero count. Draw a random variable from the uniform distribution. Identify the count value on which this random draw falls in the stacked set of count probabilities. Declare this as the total count value prediction $\hat{y}_G$ for the individual.
- Step 2: If $\hat{y}_G$ =0, declare a zero allocation for product group G for the observation and put $\hat{y}_{Gk} = 0 \ \forall \ k \ (k = 1, 2, ..., M)$. STOP .
- Step 3: Draw K independent realizations of ε_k, one for each good $k \ (k = 1, 2, ..., K)$ from the reverse extreme value distribution with location parameter of 0 and the scale parameter equal to one; label this distribution as REV(0,1) . Compute ψ_k using Equation (2).
- Step 4: Re-order the goods in descending order of ψ_k; set a new index $h \ (h = 1, 2, ..., K)$ for this new ordering of the inside goods. Let $\tilde{\psi}_h$ be the re-ordered vector of values of ψ_k and $\tilde{\gamma}_h$ be the re-ordered vector of values of γ_k.
- Step 5: If $\hat{y}_G$ =1, set $\tilde{f}_h^* = 1$ and $\hat{y}_{Gh} = 1$ for h=1; and $\tilde{f}_h^* = 0$ and $\hat{y}_{Gh} = 0$ for $h = 2, 3, ..., K$. Declare the consumptions as $\hat{y}_{Gh} \ (h = 1, 2, ..., K)$. STOP.
- Step 6: Set $M = 2$ and $\tilde{f}_1^* = 1$.
- Step 7: If $\tilde{\psi}_M < \dfrac{\tilde{\psi}_1}{\left(\dfrac{\tilde{f}_1}{\tilde{\gamma}_1} + 1\right)}$, set $\tilde{f}_h^* = 0 \ (h = M, M+1, ..., K)$ and *go to Step 9*. Else, set $\tilde{f}_h^*$ using the formula below:

$$\tilde{f}_h^* = \frac{\tilde{\psi}_h \tilde{\gamma}_h + \tilde{\gamma}_h \sum_{\substack{j=1 \\ j \neq h}}^{M} \tilde{\gamma}_j \left(\tilde{\psi}_h - \tilde{\psi}_j \right)}{\sum_{j=1}^{M} \tilde{\psi}_j \tilde{\gamma}_j}, \; h = 1, ..., M \tag{31}$$

- Step 8: Set $M = M + 1$. If $M = K$, *go to step 9*. Otherwise, *go to step 7*.
- Step 9: Compute the first-level (not necessarily integer) intensity consumptions for each of the inside goods as the product of the predicted continuous allocation for the product group and the fractional allocation to each inside good: $\hat{y}_{Gh} = \tilde{f}_h^* \times \hat{y}_G$ for $h = 1, 2, ..., M$; $\hat{y}_{Gh} = 0$ for $h = M + 1, M + 2, ..., K$.
- Step 10: Consider only the goods $h = 1, 2, ..., M$ with $\tilde{f}_h^*$>0. Set s=1, $\tilde{L} = \hat{y}_G$, and $\breve{y}_{G1} = \tilde{f}_1^* \hat{y}_G$. Also, set $\tilde{y}_{Gh} = 0$ for $h = M + 1, M + 2, ..., K$.
- Step 11: Partition $\breve{y}_{Gs}$ into an integer part ($\vec{y}_{Gs}$) and a remainder decimal part (θ_{Gs}). Draw a random variable from the standard uniform distribution. Let the result be $\tilde{u}$. If $\theta_{Gs} \leq \tilde{u}$, set $\tilde{y}_{Gs} = \vec{y}_{Gs}$, else $\tilde{y}_{Gs} = \vec{y}_{Gs} + 1$. Compute $\tilde{L} = \tilde{L} - \tilde{y}_{Gs}$.
- Step 12: If $\tilde{L}$=0, set $\tilde{y}_{Gh} = 0$ for $h = s + 1, s + 2, ..., M$. Declare the inside good intensity consumptions as $\tilde{y}_{Gh} = 0 \;\; (h = 1, 2, ..., K)$. STOP.
- Step 13: If s =M–1, set $\tilde{y}_{GM} = \tilde{L}$. STOP. Else, set s = s+1, compute $\breve{y}_{Gs} = \tilde{L}\left(\frac{\tilde{f}_s^*}{\sum_{l=s}^{M} \tilde{f}_l^*} \right)$, and *go to step 11*.

3. AN EMPIRICAL DEMONSTRATION

An application of the MDCNTEV model is demonstrated in the context of individuals' multivariate count of recreational episodes (over a four week period) to each of multiple possible tourism destination locations. The growing amount of recreational tourism trips, mostly undertaken using the personal auto mode, has led to increased attention on this leisure travel market, especially because of the contribution to such trips to traffic on city streets and between cities in close proximity. At the same time, unraveling the "push and pull" factors associated with leisure activity decisions helps cities and regions position themselves as unique and even exotic destinations, with an eye on generating jobs and revenue. This confluence of interest on recreational travel from the transportation and tourism domains has led to many studies in this space in the past decade, particularly those that consider such leisure destination choices as intrinsically arising from a combination of variety-seeking and loyalty behavior based on Iso-Ahola's (1983) theory of vacation participation in which the individual/family balances needs for familiarity and novelty, within budget constraints, to provide an "optimally arousing experience".

Examples of such studies include Van Nostrand et al. (2013), Bhat et al. (2013, 2016b), and Kuriyama et al. (2020). But, unlike these earlier studies in the transportation, tourism, and recreational fields that approximate the counts to each destination location as a continuous quantity, the current paper explicitly acknowledges the discrete nature of the intensity of participation. Also, unlike most of these earlier studies, as well as the study by Kuriyama and Hanemann (2006), we also consider the possibility of a zero count, representing zero long distance personal auto leisure trips made by the individual's household (more generally, the earlier studies assume that the total number of leisure trips across all destinations is known, and focus on how this total number of trips is allocated across destination regions, while the current study models the total number of leisure trips jointly with destination choice). To our knowledge, this is the first application of a horizontal choice multivariate count data model that accounts for the integer nature of the intensity of participation, while also accommodating zero total participation count.

3.1. Data and Sample Description

The empirical analysis is undertaken using data drawn from the 2012 New Zealand Domestic Travel Survey (DTS) (see Ministry of Business, Innovation and Employment, 2013). The DTS was targeted at individuals and not households in that only one randomly selected individual (over the age of 15 years) from each sampled household was interviewed. Respondents were asked to provide information on all long distance leisure trips of over 40 kilometers one-way made four weeks prior to the survey date. Individual and household socio-demographic information was also elicited. For the current analysis, only those individuals residing in the North Island of New Zealand are considered. The travel of these individuals using a personal auto to nine aggregate destination regions in the North Island is considered.[9] The regional spatial classification scheme of the North Island into nine regions is the same as that used by the Department of Tourism of New Zealand for its marketing campaigns, and is also the commonly used geo-political partitioning of the country (see Figure 1).[10] From the survey, the total count of long distance trips is obtained, as is the count of trips to each of the nine destination regions of North Island by residents of the North Island. Survey records are supplemented with a network level of service file that provided information on land travel distance and highway travel time between the residential city of each respondent and the nine destination regions. Travel cost skims are then computed as a function of the respondent's reported household income, the estimated cost of

[9] Personal auto trips comprise around 90% of all leisure domestic trips within New Zealand; see Ministry of Business, Innovation and Employment, 2008, and this has remained stable over time as indicated in Lawson et al., 2021. Also, there is little personal auto vacation travel from the North Island to the South Island or vice versa (only 0.7% of total long personal auto distance domestic leisure trips are from one island to the other, which is possible using a ferry service that transports personal vehicles across the Cook Strait). Further, the North Island of New Zealand is the more populated part of the country, home to 76% of the overall New Zealand population (the North Island's population was about 3.2 million in 2013, while that of the South Island was about 1 million in 2013). In the DTS sample, 74.6 percent reside in the North Island, close to the population representation in 2013. Commensurately, 74% of long distance domestic leisure trips are contained within the North Island.

[10] We acknowledge the possible estimation bias because of using aggregate spatial destination regions (see, for example, Parsons and Needelman (1992)). But, given we are using such spatial aggregation, we have developed many proxy measures that capture the intensity of available activity opportunities, as discussed further in this section.

vehicle fuel on land, and the land travel distance and highway travel time skims, following the standard approach of valuing travel time at a fixed proportion of one-half of the wage rate (see Bhat et al., 2016b for a detailed discussion).

Additionally, a disaggregate spatial land-cover characteristics data obtained from the 2012 Land Cover Database (LCDB) of the Land Resource Information System (LRIS) of New Zealand was used to compute information related to the total land area of each destination, as well as the land-cover in six broadly defined categories: urban area (including central business districts, commercial and industrial areas, urban parklands, urban dumps, and housing and transportation-related land cover), water area (including rivers, land/ponds, freshwater, and estuarine open water), wetland area (context-dependent combinations of areas such as herbaceous freshwater vegetation, flaxland, and saline vegetation), agricultural area (including vineyards and orchards, perennial crops, short rotation cropland, and grasslands), bare-land area, and forest area (pine forests, mangroves, deciduous hardwoods and other exotic/indigenous forest areas). Using these land-cover areas, two measures are computed for use in the empirical analysis. The first measure is a land-cover accessibility measure of the Hansen-type (Fotheringham, 1983), computed for individual q and land-cover type i as presented by destination region k as $AC_{qki} = LC_{ki} / [f(TT_{qk})]$, where LC_{ki} is the percentage area in land-cover category i (i = urban, water, wetland, agricultural, bare-land, and forest) in destination region k, TT_{qk} is the travel time (in hours) from individual q's residence city to the centroid of destination region k, and $f(.)$ is a function. In our specifications, we considered both a linear form, $f(TT_{qk}) = TT_{qk}$, as well as a logarithmic form, $f(TT_{qk}) = \ln(TT_{qk})$. The logarithmic form penalizes destination regions less for being far away from the residential location of the individual. Based on our specification tests, the linear form turned out to be the preferred functional form for $f(TT_{qk})$. The second measure computed using the land cover areas for each destination is an overall land cover diversity index for each destination region computed from the land-cover percentage by category as follows (see Bhat et al., 2016b):

$$\text{Land-cover diversity index } D_k = 1 - \left\{ \frac{\sum_{i=1}^{I} \left| LC_{ki} - \frac{100}{I} \right|}{\left(\frac{200(I-1)}{I} \right)} \right\}, \tag{32}$$

where LC_{ki} is the percentage area in land-cover category i in destination region k (as earlier) and I=6 (that is, we have six land cover categories) in our empirical context. The functional form would assign the value of zero if a region's land-cover is only in one category, and would assign a value of 1 if a region's land-cover is equally split among the different land cover categories. While the first accessibility measure captures any preferences individuals have for specific types of activities that may be featured in each destination region, the second measure recognizes that some individuals may be drawn to destination regions that have a good diversity of activity participation opportunities.

The final data sample used in the estimation included 5,622 individuals. Among these, 3051 individuals do not make any trips. This constitutes 54.3% of individuals. Earlier studies typically do not consider this high fraction of individuals in a multiple discrete allocation model, because the "budget" (total number of trips made) for the multiple discrete model is considered a given and positive. This exogeneity of the total count of trips from the allocation model delinks how characteristics of individual regions can themselves influence whether or not an individual makes a trip and how many trips. Our two-stage budgeting model accommodates for this linkage.

Table 1 provides the distribution of the 2571 individuals who make one or more trips by the number of leisure trips made during the four week period before they were surveyed and by the number of distinct leisure destination regions visited. A sizeable fraction of individuals make one trip (71.8% of individuals). However, there is a non-insignificant percentage (29.2%) of individuals making more than one leisure trip. As expected, most individuals who undertake more than one trip during the survey period prefer to travel to multiple destinations (see the second row and beyond in Table 1). For example, 61% of individuals making two trips during the survey period visit more than one distinct destination region, while close to 70% of individuals making three trips visit more than one distinct region.

Table 2 provides descriptive statistics for each of the nine destination regions. The second broad column presents the mean and standard deviations for the travel impedance skim measures of total travel time, travel distance, and travel cost for each destination region. Not surprisingly, given the spatial layout of the North Island (see Figure 1), the travel impedance measures are the highest for the Northland region (the northernmost region of the island), the Wellington region (the southernmost region of the island), and the Gisborne region (the easternmost point of the island). As expected, the impedance measures are the least for the Auckland and the Waikato regions, not only because of the spatial layout of the Island, but also because these two regions are two of the most populous regions of the Island from where vacation leisure trips originate. In fact, 51% of all individuals in the sample reside in the Auckland and Waikato regions, which is close to the 2013 population representation of 55% for these two regions as a percentage of the North Island population.[11] Auckland, in particular, is home to 37% of all individuals in the sample, similar to its dominant representation in the North Island population (the only other region that comes close to the populations of Auckland and Waikato is the Wellington region, home to 14.6% of the North Island population in 2013, and home to 14.9% of the North Island individuals represented in our sample; but the Wellington region is at the southern tip of the North Island, and so has a high travel impedance on average in Table 2). Consistent with the population representation, the Auckland and Waikato regions contribute 49.2% of all trip origins in the sample.

The third main column in Table 2 provides the land diversity index for each region. Again, Auckland and Waikato come out on top as the most diverse land cover regions in the island, with Northland, Bay of Plenty, and Wellington also showing good land cover diversities relative to other regions. Auckland, in particular, has a high diversity index, thanks to the famous urban tourist

[11] See https://figure.nz/chart/3OjxZbBpv5fOwovb (accessed May 5, 2022).

attraction of the City of Auckland as well as such attractions as the Tiritiri Matangi Islands, a haven for nature hikers who want to experience the rich flora and fauna of the region up close (especially of a host of endangered species of birds, each with a unique bird call pattern). Of course, while activity diversity opportunities of a destination may be important for some individuals, accessibility to specific activities may be important to others, which is captured in our analysis using the six land cover-specific accessibility measures (not shown in Table 2 to conserve on space).

The remaining columns of Table 2 provide descriptive statistics related to leisure trip patterns, starting with the number of individuals visiting each region (see the fourth main column of Table 2). The Waikato region is clearly the one patronized by the most number of individuals, but Auckland and Bay of Plenty also draw quite a few individuals. However, to get a better picture of attractiveness, the fifth main column provides a size measure for each region, and the sixth main column normalizes the number of people visiting by the area of each region. This column shows that, on a per unit area basis, Auckland is by far the most popular destination, followed by Wellington, Bay of Plenty, and Waikato. Interestingly, as indicated earlier, these are also destinations with high land diversity index values, suggesting a generally positive relationship between land cover diversity and tourist draw on a per unit area basis. To capture the volume of activity opportunities effect, we consider the logarithm of the area of each region in the empirical analysis.[12] The seventh broad column presents statistics on the number of visits to a destination region among those who visited the destination region at least once. The mean and range from this column suggest that Auckland, Waikato, Bay of Plenty, and Wellington have the most loyal following.

3.2. Model Estimation Results

The number of destination region alternatives in the fractional MDCEV model is nine. Thus, rather than including nine alternative-specific constants in the baseline preference and nine region-specific satiation parameters (in addition to other explanatory variables) in the MDCEV fractional destination allocation model, we adopted an "unlabeled" specification in which the baseline preferences and satiations are captured through attributes of the individual regions and/or interactions of individual demographics with the regional attributes. Also, while the focus here is on demonstrating the application of the proposed model rather than necessarily on substantive interpretations and policy implications, we did undertake a rigorous specification analysis with the data available to arrive at the best possible specification (for example, considering alternative functional forms for continuous independent variables such as household income, including a linear form, piecewise linear forms in the form of spline functions, and dummy variable specifications for different groupings).

[12] The coefficient on this size variable may be viewed as an inclusive value characterizing the presence of common unobserved destination region attributes affecting the utility of elemental alternatives within each region. If less than one, as would be the expectation, the implication is that there is an inelastic influence of increasing region size on the region's baseline utility.

Several types of variables were considered in the first stage total count model as well as the second stage fractional MDCEV model. These included household sociodemographics (household size, presence and number of children, age of individual as an indicator of lifecycle stage, number of adults in household, traveling party size, household income, family structure, and destination attributes). Further, to accommodate heterogeneity across individuals in the effect of observed variables not only in the baseline preference function (the ψ_k function as in Equation (1)), but also in the satiation parameters (the γ_k parameters), we parameterized the satiation parameters as $\exp(\delta_k' \omega_k)$, where ω_k is a vector of decision maker-related characteristics and δ_k is a vector to be estimated (note that $\gamma_k > 0 \ \forall\, k$). This allows the discrete choice decision of traveling to a destination) less closely tied to the count choice of the number of times of traveling to that destination (see Bhat, 2008).

3.2.1. Baseline Utility Parameters

The parameter estimates in the top row panel of Table 3 relate to the impact of variables on the logarithm of the baseline preference. As expected, the effect of the destination size variable in the baseline utility function is positive and less than one, indicating the inelastic and positive effect of area size on the baseline utility.

The land cover accessibility measures suggest that regions with high wetland land cover tend to "pull" leisure trips, while there is an opposite effect for regions with high urban and forest cover. Wetland areas are likely associated with nature tourism, with the opportunity for activities such as fishing and hunting, and enjoying the flora and fauna associated with such wetlands (for example, bird-watching, as the bays and estuaries typically associated with wetlands also serve as seabird colonies and can also be transit points for migrating birds). On the other hand, urban areas, typically dominated by commercial and industrial areas, urban dumps, and housing and transportation-related land cover may not be the most appealing for leisure trips taken away from home. Similarly, forest land cover with pine forests and deciduous hardwoods, while useful for preserving vegetation and aiding in tempering climate change concerns, may not provide much opportunity for leisure activity. Agricultural land-cover accessibility also impacts destination visitation preference, though only for those with low incomes. This is perhaps a reflection of a conscious effort to narrow down destination possibilities to those that have little uncertainty in terms of pleasure value as well as financial investment, such as visiting vineyards for a relaxed wine-tasting escapade.

The land-cover diversity index effect on the baseline function indicates a generally high preference for visitation to regions with a good diversity of activities for all individuals, as also clearly evidenced in the substantial "pull" of leisure trips to Auckland in the descriptive statistics. This "pull" is tempered for those with low incomes (annual household income less than NZ $50,000), indicating that higher income individuals, in part because of their financial ability to pursue multiple activities, appear to be more predisposed than their lower income peers to visit regions with a high diversity mix of activities.

The σ coefficient is effectively the negative of the inverse coefficient on the logarithm of price in the baseline utility preference (see Equation (2)). This value is estimated as 2.667 in the proposed linked model, with the effective ln(price) coefficient $(-1/\sigma)$ being –0.375 (with a t-stat of –18.65 against a test of zero). In the unlinked model, this value is estimated as 1.854, implying an effective ln(price) coefficient of –0.539 (with a t-statistic of –14.98 against a test of zero). As expected, destinations that incur a high travel cost are less likely to be patronized in both the unlinked and linked models, though our proposed model indicates a much more tempered effect of cost on destination region choice relative to the unlinked model (we are able to compare these two coefficients directly because the MDCEV model structure is the same in both the unlinked and linked models, with the error scale being normalized to one). Of course, the linked model also impacts the total count of trips through an income effect, which can then have an added effect on the consumption of other inside alternatives because of a price change in one alternative (see discussion in Section 3.2.4).

3.2.2. Satiation Effects

These effects are presented in the bottom panel of Table 3. As indicated earlier, the satiation parameter is parameterized as $\gamma_k = \exp(\delta_k' \omega_k)$, and the satiation coefficients in Table 3 are the δ_k parameters. A positive parameter on a variable implies that an increase in the variable has the effect of increasing the γ_k parameter and decreasing satiation (that is, increasing repeat trips of the individual to a destination region), while a negative parameter has the effect of decreasing the γ_k parameter and increasing satiation (that is, decreasing repeat trips of the same individual to a destination region).

The wetland land-cover accessibility measure has a positive effect on the satiation parameter, i.e., destinations with higher wetland land cover not only spur a visit to the destination, but also lead to more repeat visits to such regions. This indicates that leisure travelers interested in nature tourism appear to place a high priority for repeat travel, suggesting the strong draw for some peaceful connection time with nature. Regions with a high land cover diversity index, on the other hand, generally draw less repeat visits than those with a low land cover diversity index, particularly so for single adults, high income individuals, and adults younger than 48 years of age (relative to non-single adults, low income individuals, and adults 48 years of age or older, respectively).[13] The only segment of the population in which there is a slight increase in repeat-visit tendency to destinations with a high land cover diversity is non-single, low income individuals 48 years of age or older. Overall, it appears that individuals prefer a first visit to a new destination that has a good

[13] Note that the coefficient on land cover diversity index (multiplied by 10) has a value of –0.067 for a non-single, low income adult younger than 48 years of age; a value of –0.993 for a single, low income adult younger than 48 years of age adults; a value of +0.057 for a non-single, low income adult 48 years of age or older; a value of –0.869 for a single, low income adult 48 years of age or older; a value of –0.765 for a non-single, high income adult younger than 48 years of age; a value of –1.691 for a single, high income adult younger than 48 years of age adults; a value of –0.641 for a non-single, high income adult 48 years of age or older; a value of –1.567 for a single, low income adult 48 years of age or older).

diversity of activities, but then tune in to specific activities in repeat visits, and this is particularly so for nature-connecting activities. In this sense, diversity of activities at a destination may actually hamper repeat visits if that destination has already been visited once. This is consistent with the results from Lawson et al. (2021) in their analysis of the motivating factors for domestic tourist travel to destinations in New Zealand. These authors found that "discovery and experiencing a place" (which is likely to be associated with diversity of activities) was important in early visits to a destination, but faded in importance relative to specific activity pursuits (such as outdoor exercise) in later visits. From a marketing strategy perspective, this suggests that, if regions are able to collect an inventory of visitors with contact information, an effective marketing strategy might be to emphasize activities within each broad leisure category more so than necessarily speak to the diversity of activities.

The constant coefficient has no substantive interpretation, and simply controls the overall level of satiation for destination visits.

3.2.3. Count Data Model Component

The count model parameters in Table 4 represent the elements of the threshold-specific vector $\boldsymbol{\alpha}$ and the $\boldsymbol{\mu}$ vector that appears in the threshold functions of the recast count model (see Equation 19). The first three elements present the threshold-specific constants (α_k values) and the constant in the $\boldsymbol{\mu}$ vector. The threshold specific constants (α_k) provide flexibility in the count model to accommodate high or low probability masses for specific outcomes. In the current empirical analysis, the best specification was reached with two threshold specific constants, one for α_0 (that gets added to the upper bound of the zero count range) and another for α_1 (that gets added to the upper bound of the one count range). The positive values for both of these, and the higher positive value of α_0 over α_1, reflect the inflated number of individuals who make zero counts for leisure trips and a less exaggerated, but still inflated number of individuals who make exactly one leisure trip during the four week survey period (see Table 1). The constant in the $\boldsymbol{\mu}$ vector does not have any substantive interpretation, except for mapping the latent propensity optimally to the observed counts, given the coefficients on other variables embedded in the threshold function.

For the other variable coefficients in the $\boldsymbol{\mu}$ vector, 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).[14] 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). Thus, the results indicate that single individuals, in particular, but also couple

[14] Note that $f_i(\boldsymbol{\varpi}) = -S_{\zeta}^{-1}\left(e^{-\lambda}\sum_{l=0}^{i}\frac{\lambda^l}{l!}\right)$, with $\lambda = \exp(\boldsymbol{\mu}'\boldsymbol{\varpi})$. Thus, if an element of $\boldsymbol{\mu}$ is positive, it decreases $e^{-\lambda}\sum_{l=0}^{i}\frac{\lambda^l}{l!}$, increases $S_{\zeta}^{-1}\left(e^{-\lambda}\sum_{l=0}^{i}\frac{\lambda^l}{l!}\right)$, and therefore decreases $f_i(\boldsymbol{\varpi})$.

families, are more likely to make non-zero leisure trips relative to nuclear family households, perhaps because of fewer child care responsibilities (while the base category includes both nuclear and joint families, nuclear family households dominate this base category). Single parents are also more likely to make non-zero leisure trips compared to nuclear family households (with the additional propensity being the same as that for couples). The differing effects of the presence of children between nuclear families and single parent families is intriguing, and certainly deserves additional investigation. Perhaps single parents find travel as a way to better cope (through a change of their environment) with the sole responsibility for their children. Besides, single parents appear to view tourism travel as a way to have one-on-one interactions with their children, treat such travel as an adventure, and appear to be particularly interested in exposing their children to new things (see Vartan, 2018 and Camargo et al., 2021). Overall, the household structure results indicate that nuclear family households and joint family households are the most likely to make zero leisure trips. Finally, older individuals (48 years or older) are less likely to have leisure trips relative to their younger peers, presumably a reflection of stay-at-home tendencies of middle-aged and older individuals.

3.2.4. Linking Parameter

The parameter that links the fractional MDCEV model with the count model in our final model specification is equal to $1/\sigma$ (see the text surrounding Equation (24)). This value is the same as the negative of the coefficient on the price variable, and is equal to 0.375 (and highly statistically significant with a t-statistic of 18.65). The implication is that factors that increase the attractiveness of any destination region or a decrease in the cost of travel to any destination region will lead to more non-zero counts of leisure tourism trips, while the unlinked model would completely ignore such destination choice effects on trip generation. To demonstrate this, we consider a 15% reduction in cost to the northmost region (Northland) and the southmost region (Wellington), say through an improvement in the road transportation system that enables faster access to these regions. The net effect is that the mean cost to Northland falls from NZ$171.3 to NZ$145.6, and the mean cost to Wellington falls from NZ$198.1 to NZ$168.4. These drops are quite reasonable, as they still exceed the travel costs to the next inner region of the island at both the north and south ends. We then compute the net predicted percentage shifts on total visit counts to the nine regions after such a price change. This prediction is done using the procedure laid out in Section 2.5 in which we ensure that the predictions for each individual are counts. However, we run the procedure using 1000 draws per individual, predicting integer counts for the visits of every individual to each destination for each of the draw occasions. We next aggregate the visit counts to each destination, and take a mean across the 1000 draws. The last step does not guarantee a positive integer count at the aggregate level, but these aggregate values are rounded off.

The results of the scenario described above are shown in Table 5. Point estimates are presented, with the 95% confidence bounds shown in parenthesis. In the following discussion, we will focus on the point estimates (while the confidence bounds show a good bit of overlap between the linked model and the unlinked model results for the 15% travel cost reduction to the northmost and southmost regions of the North Island, the overlap is context-specific and is a function of the specific policy being considered; also the overlap reduces as the price "shock" is increased). Both models show a relatively substantial increase in trip visits to Northland and Wellington after the price drop. However, the effect on other regions is either tempered (in terms of the draw away) or even positive relative to the predictions from the unlinked model (except for the Auckland region, which sees a higher drop as predicted by our model relative to the unlinked model, perhaps because of the compact nature of Auckland that encourages tourists to go beyond Auckland to Northland for their visits). The tempered negative effects or even positive effects for other regions is because the price decreases for travel to Northland and Wellington also lead to an overall trip visit count increase due to an income effect, which then tempers any reductions in visitations to other regions of the island due to the substitution effect (the total count of leisure trips across all regions increases by 1.7% as predicted by the linked model, while there is no change in total leisure trips in the unlinked model). This is clear from Table 5, where the reductions (of trips) to other regions as predicted by our model is lower than that predicted by our model for Bay of Plenty, Gisborne, and Taranaki. And the trips drawn to Waikato, Manawatu-Wanganui, and Hawke's Bay actually increase. This latter effect can be attributed to the fact that, with large areas translating to more activity destination opportunities, these three regions are the beneficiaries of much of the newly generated trips caused by the income effect (note from the "area" column of Table 2 that Waikato, Manawatu-Wanganui and Hawke's Bay are the three largest regions of the island). Overall, the implication is clear. Improvements in travel costs or attraction measures to one or more specific regions may lead to but small reductions in tourist visits to some other regions, and even increased visits to some other regions. Thus, rather than fear competition as would be predicted by the unlinked model, the whole island is likely to benefit because of improvements in access and attractiveness of any region of the country, as long as the improvements across regions are not that asymmetric to make one region substantially more attractive.

In addition to ignoring income effects, the substantive effects of variables in both the destination allocation and the total count model components from the unlinked model are also different from that of the linked model. For instance, while the linked model indicates that low income individuals are less attracted (relative to high income individuals) for their first visits toward regions with high land cover diversity (and this effect is very highly statistically significant), the unlinked model suggests the opposite and statistically insignificant effect at the 0.05 level of significance (the coefficient estimate is +0.6911 with a t-statistic of 1.48). Similarly, while single parents are more likely to make non-zero leisure trips (relative to nuclear family household) according to our model, the unlinked model finds no statistically significant difference in leisure trip-making between single parent families and nuclear families. This difference between the unlinked and linked models is interesting. While the tourism and hotel industry has traditionally focused on "wooing" nuclear families because of the high fraction of such families, our results support the more recent efforts of the industry to pay attention to the rising population segment of single parents, through such incentives as affordable child care facilities and single occupancy hotel room charges (see Camargo et al., 2021).

3.3. Data Fit Measures

Data fit measures are presented in two forms – likelihood-based data fit measures and non-likelihood based data fit measures.

3.3.1. Likelihood-Based Data Fit Measures

As a base naïve model, we compute the log-likelihood of a model with only the following parameters: (a) logarithm of area in the fractional split MDCEV baseline preference, (b) constant in the $\boldsymbol{\varpi}$ vector embedded in the λ parameter of the count thresholds, and (c) the two α_l terms in the threshold. We set the satiation parameters to one for all alternatives, and the scale for the count model error to one. The price variable is also set to one for all regions (so there are four parameters estimated in this base naïve model: the coefficient on ln(area) in the baseline preference, the two count thresholds, and the constant in the count model). Note that because of the inclusion of the two α_l terms, which appropriately adjust based on linkage or not, the log-likelihoods of the naïve model is exactly the same regardless of linkage or not (the log-likelihood value for this base naïve model is –15958.1). Thus, for all likelihood-based data fit measures, we can use the log-likelihood of the naïve model as the base model. We also compute the Bayesian Information Criterion (BIC) values [$= -\log L_{ML}$ + 0.5 (# of model parameters) log (sample size)] for the linked and unlinked models ($\mathscr{L}(\hat{\theta})$ is the log-likelihood at convergence).[15] Finally, the two models are compared using a non-nested likelihood ratio test. The adjusted likelihood ratio index of each model is first computed with respect to the log-likelihood of the base naive model:

$$\bar{\rho}^2 = 1 - \frac{L(\hat{\boldsymbol{\theta}}) - M}{L(c)}, \tag{33}$$

where $L(\hat{\boldsymbol{\theta}})$ and $L(c)$ are the log-likelihood functions at convergence and the base naïve model, respectively, and *M* is the number of parameters estimated in the model (excluding the number of parameters included in the base naïve model). If the difference in the indices is $(\bar{\rho}_2^{\,2} - \bar{\rho}_1^{\,2}) = \tau$, then the probability that this difference could have occurred by chance is no larger than $\Phi\{-[-2\tau L(c) + (M_2 - M_1)]^{0.5}\}$, with a small value for the probability of chance occurrence suggesting that the difference is statistically significant and the model with the higher value for the adjusted likelihood ratio index is preferred.

The likelihood based data fit measures are provided in Table 6. Both the linked and unlinked model log-likelihood values are clearly superior to the base naïve model, as can be observed from the nested likelihood ratio test (fourth row). A further comparison between the

[15] Log L_{ML} refers to the log-likelihood value at convergence. As indicated earlier, the linked and non-linked models are non-nested, because the kernel error term distributions are different between the two models. While many measures have been suggested in the literature to evaluate model fit among non-nested models (see Dziak et al., 2020), the BIC-based measures demand a higher strength of evidence to add complexity than do the other measures, and thus the BIC-based measure favors more parsimonious models.

linked and unlinked models using the BIC and the non-nested likelihood ratio index tests (shown in the last two rows) indicates the clear data fit superiority of the proposed linked model over the unlinked model in the current empirical context.

3.3.2. Non-Likelihood Fit Measures

To further supplement the disaggregate likelihood-based performance, we evaluate the performance of the models intuitively and informally at a disaggregate and aggregate level. At the disaggregate level, we estimate an average probability of correct prediction for the observed multivariate count outcome. At the aggregate level, to keep the presentation manageable, we focus on the univariate count of visits to each destination region, rather than consider the multivariate visit combinations to destinations. Specifically, we compare (a) the predicted total number of visits to each destination (which constitutes the multiple discrete-count dimension that considers the combination of participation propensity and satiation behavior) to the actual number of visits to each destination, and (b) the predicted number of distinct individuals visiting each destination (that is, the number of individuals visiting each destination at least once, which constitutes the discrete dimension of participation propensity) to the actual number of individuals visiting each destination at least once. This prediction is done using the same procedure as laid out in Section 3.2.4. We then design an informal heuristic diagnostic check of model fit for each destination region using an absolute percentage error (MAPE) statistic, and then compute a weighted mean absolute percentage error (MAPE) statistic (which is the MAPE for each destination region weighted by the actual percentage draw to that destination region).

In terms of the results, the average probability of correct prediction at the multivariate count level is 0.3154 for the unlinked model and 0.3715 for our proposed linked model. This average probability may also be computed separately for the zero counts and for the non-zero counts. The average probability of correct prediction for zero counts is 0.5440 for the unlinked model and is 0.5478 for the linked model, indicating that the linked count model does marginally better in predicting zero counts at the individual level. The corresponding average probability of correct prediction for non-zero counts is 0.2570 for the unlinked model and 0.4890 for the linked model, showing the substantial improvement for non-zero counts. This is because of the linkage from the destination choice stage to the trip generation count stage in our proposed model, demonstrating the importance of linkage in predicting the multivariate counts.

Moving on to the aggregate predictions, Table 7 provides the actual numbers, the predictions, and the error statistics for the total number of visits to each of the nine destination regions, while Table 8 provides the corresponding information for the distinct number of individuals visiting each destination. Both the linked and unlinked models do reasonably well in the predictions, especially given the unlabeled nature of the model. Of course, the models are not able to predict the very low number of counts to the Gisborne destination region, which suggests there are other regional factors unique to Gisborne that substantially reduce the draw to this eastern region of the island (though the high APEs for Gisborne are also because of the very low actual counts for this region; from an absolute error perspective, the predictions for "Bay of Plenty" are the most out of line, both for total visits and for distinct individuals visiting, while the counts for Wellington are also quite substantially underpredicted). Overall, the linked model performs better than or at about the same level of accuracy as the unlinked model for both total visits and individuals visiting each of the nine regions, demonstrating not only the utility-theoretic value of the proposed model but also its empirical value. As can be observed, the weighted MAPE for total visits is 24.8% for our proposed linked model compared to almost 30% for the unlinked model. The corresponding weighted MAPE for the distinct number of individuals visiting each destination is 21.3% for our proposed linked model relative to 25.1% for the unlinked model.

In summary, the performance of the proposed model at the aggregate as well as disaggregate levels reinforces the value of expressly acknowledging the linkage between the destination region and trip generation decisions of individuals, as well as recognizing the multiple discreteness associated with leisure travel.

4. CONCLUSIONS

Multivariate count data models are applied routinely in many fields, including tourism, environmental economics, transportation, and biostatistics, just to name a few. Such models expressly recognize the positive integer nature of count data. This makes it difficult to apply models based on KKT first order conditions, which are suitable for continuous dependent variables, but not integer count variables. As a result, the literature has employed models based on discrete distributions or based on a vertical accumulation approach. However, such methods do not adequately consider the possibility that the multivariate count observations may originate from a "horizontal" choice of multiple alternatives that are viewed as being imperfect substitutes of one another. That is, there may be an intentional and purposeful "horizontal at-once" choice process at play right from the outset in many product group consumption patterns that reflects satiation behavior, based on the recognition that different alternatives serve different functionalities and provide different kinds of utility "arousal".

In this paper, we have proposed a new utility-theoretic two-stage budgeting-based econometric multiple discrete-count model based on the linking of a fractional split MDCEV model component with a total count model component. Through the specification of a reverse-Gumbel error specification for the random terms in the MDCEV component and the GOR framework of the total count model, we obtain a multiple discrete-count extreme value (MDCNTEV) model that has a closed-form probability expression for the multivariate counts and that is estimable using straightforward maximum likelihood estimation. An application of the proposed model is demonstrated in the context of individuals' multivariate count of recreational episodes to each of multiple possible tourism destination locations. The results highlight the promise of the proposed model for a variety of multivariate count consumer choice settings. The model can also serve as a base model over which random heterogeneity may be superimposed to specify more advanced models. More generally, it would be interesting to undertake a deeper analysis on when a vertical choice approach may be more appropriate and when a horizontal choice approach may be more appropriate for multivariate count modeling. The author and his students are currently undertaking further investigations along these lines.

ACKNOWLEDGMENTS
The author is grateful to Lisa Macias for her help in formatting this document. Sebastian Astroza helped with data compilation efforts. The author would like to thank two anonymous reviewers for valuable comments.

REFERENCES

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

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

Bhat, C.R., 2022. A new closed-form two-stage budgeting-based multiple discrete-continuous model. Technical paper, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin.

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

Bhat, C.R., Paleti, R., and Castro, M., 2015. A new utility-consistent econometric approach to multivariate count data modeling. *Journal of Applied Econometrics*, 30(5), 806-825.

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

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

Bhat, C.R., Mondal, A., Pinjari, A.R., Saxena, S., and Pendyala, R.M., 2022. A multiple discrete continuous extreme value choice (MDCEV) model with a linear utility profile for the outside good recognizing positive consumption constraints. *Transportation Research Part B*, 156, 28-49.

Blackorby, C., Primont, D. and Russell, R., 1978. *Duality, Separability and Functional Structure; Theory and Applications*, American Elsevier, New York.

Camargo, B., Iberri, L., Llano, R., and Lozano, M., 2021. New perspectives on family tourism: Motivations, travel behavior and experiences of single-parent families. Proceedings of the 2021 TTRA Canada Virtual Conference.

Castro, M., Paleti, R., and 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.

Dubé, J.-P., 2004. Multiple discreteness and product differentiation: Demand for carbonated soft drinks. *Marketing Science*, 23(1), 66-81.

Dziak, J.J., Coffman, D.L., Lanza, S.T., Li, R., and Jermiin, L.S., 2020. Sensitivity and specificity of information criteria. *Briefings in Bioinformatics*, 21(2), 553-565.

Fally, T., 2022. Generalized separability and integrability: Consumer demand with a price aggregator. NBER Working Paper No. w29997, Available at SSRN: https://ssrn.com/abstract=4098319

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

Hausman, J.A., Leonard, G.K., and 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.

Hendel, I., 1999. Estimating multiple-discrete choice models: An application to computerization returns. *Review of Economic Studies*, 66, 423-446.

Inouye, D., Yang, E., Allen, G., and Ravikumar, P., 2017. A review of multivariate distributions for count data derived from the Poisson distribution. *Wiley Interdisciplinary Reviews (WIREs) Computational Statistics*, 9(3), e1398.

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

Kuriyama, K., and Hanemann, W.M., 2006. The integer programming approach to a generalized corner-solution model: An application to recreation demand. *Working paper*, Waseda University, Tokyo.

Kuriyama, K., Shoji, Y., and Tsuge, T., 2020. The value of leisure time of weekends and long holidays: The multiple discrete-continuous extreme value (MDCEV) choice model with triple constraints. *Journal of Choice Modelling*, 37, 100238.

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

Lawson, G., Dean, D., He, Y., Huang, X. 2021. Motivations and satisfaction of New Zealand domestic tourists to inform landscape design in a nature-based setting. *Sustainability*, 13(22), 12415.

Lee, S., and Allenby, G.M., 2014. Modeling indivisible demand. *Marketing Science*, 33(3), 364-381. https://doi.org/10.1287/mksc.2013.0829

Ma, J., and Ye, X., 2019. Modeling household vehicle ownership in emerging economies. *Journal of the Indian Institute of Science*, 99(4), 647-671.

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

McCoy, B.M. and Wu, T.T., 1973. *The Two-Dimensional Ising Model*, Harvard University Press, Cambridge, MA.

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

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

Morey, E.R., Rowe, R.D., and Watson, M., 1993. A repeated nested-logit model of Atlantic salmon fishing. *American Journal of Agricultural Economics*, 75(3), 578-592.

Mouter, N., Koster, P., and Dekker, T., 2021. Contrasting the recommendations of participatory value evaluation and cost-benefit analysis in the context of urban mobility investments. *Transportation Research Part A*, 144, 54-73.

Murteira, J.M.R., and Ramalho, J.J.S., 2016. Regression analysis of multivariate fractional data. *Econometric Reviews*, 35(4), 515-552.

Paleti, R., Bhat, C.R., Pendyala, R.M., Goulias, K.G., Adler, T.J., and Bahreinian A., 2014. Assessing the impact of transportation policies on fuel consumption and greenhouse gas emissions using a household vehicle fleet simulator. *Transportation Research Record: Journal of the Transportation Research Board*, 2430, 182-190.

Parsons, G.R., and Needelman, M.S., 1992. Site aggregation in a random utility model of recreation. *Land Economics*, 68, 418-433.

Peyhardi, J., Fernique, P., Durand, J.-B., 2021. Splitting models for multivariate count data. *Journal of Multivariate Analysis*, 181, 104677.

Potts, R.B., 1952. Some generalized order-disorder transformations. *Mathematical Proceedings of the Cambridge Philosophical Society*, 48(1), 106-109.

Razaee, Z.S., and Amini, A.A., 2020. The Potts-Ising model for discrete multivariate data. Proceedings of the 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

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

Saxena, S., Pinjari, A.R., and Bhat, C.R., 2022. Multiple discrete-continuous choice models with additively separable utility functions and linear utility on outside good: Model properties and characterization of demand functions. *Transportation Research Part B*, 155, 526-557.

Shin, J., Hwang, W., and Choi, H., 2019. Can hydrogen fuel vehicles be a sustainable alternative on vehicle market?: Comparison of electric and hydrogen fuel cell vehicles. *Technological Forecasting and Social Change*, 143, 239-248.

Strotz, R.H., 1957. The empirical implications of a utility tree. *Econometrica*, 25(2), 269-280. https://doi.org/10.2307/1910254.

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

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

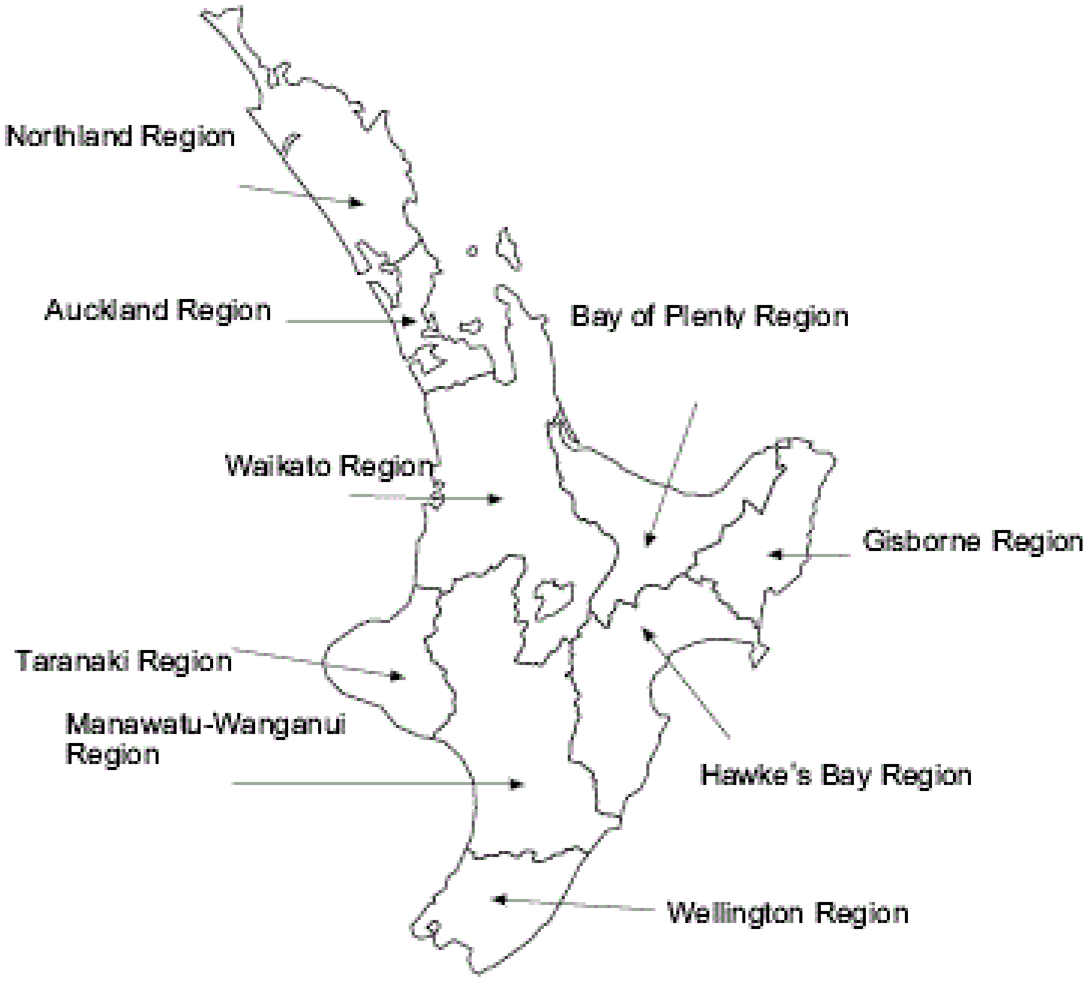
Varghese, V., and Jana, A., 2019. Multitasking during travel in Mumbai, India: Effect of satiation in heterogenous urban settings. *Journal of Urban Planning and Development*, 145(2), 04019002.

Vartan, S., 2018. How single-parent families are changing the way we travel. *CNN Travel*, https://www.cnn.com/travel/article/single-parent-family-travel/index.html

Wiśniewska, A., Budziński, W., and Czajkowski, M., 2020. An economic valuation of access to cultural institutions: Museums, theatres, and cinemas. *Journal of Cultural Economics*, 44, 563-587.

Woodside, A.G., and Lysonski., S., 1989. A general model of traveler destination choice. *Journal of Travel Research*, 27(4), 8-14.

Zhang, Y., Zhou, H., Zhou, J., and Sun, W., 2017. Regression models for multivariate count data. *Journal of Computational and Graphical Statistics*, 26(1), 1-13.



Source: www.stats.govt.nz

Figure 1. Boundaries of New Zealand Regions in the North Island

Table 1. Recreational Travel Number of Trips

Number of trips	Number of individuals	Number (%) of individuals visiting[a]				
		1 region	2 regions	3 regions	4 regions	5 regions
1	1,847 (71.8%)	1,847 (100%)	0	0	0	0
2	539 (21.0%)	241 (39.0%)	298 (61.0%)	0	0	0
3	133 (5.2%)	41 (30.8%)	67 (50.4%)	25 (18.8%)	0	0
4	40 (1.5%)	13 (32.5%)	18 (45.0%)	6 (15.0%)	3 (7.5%)	0
5	7 (0.3%)	1 (14.3%)	2 (28.6%)	3 (42.8%)	0	1 (14.3%)
6	1 (0.0%)	0 (0.0%)	1 (100.0%)	0	0	0
7	2 (0.1%)	0	1 (50%)	1 (50%)	0	0
10	2 (0.1%)	0	0	2 (100%)	0	0

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

Table 2. Destination Region Characteristics

Destination Region	Travel Impedance Measures (Std. Dev.)			Land Coverage Diversity Index	Total number (%) of individuals visiting	Area (miles^2)	Number of visiting individuals per unit area (per mile^2)	Mean and Range of number of trips
	Travel Time (hours)	Travel Distance (miles)	Cost (NZ$)					
Northland	4.64 (2.99)	242.8 (160.5)	171.3 (183.8)	0.264	289 (5.1%)	5,383	0.0537	1.16 (1-4)
Auckland	2.92 (2.76)	152.8 (151.1)	132.3 (155.9)	0.381	574 (10.2%)	2,162	0.2655	1.17 (1-6)
Waikato	2.76 (1.95)	141.6 (103.5)	120.9 (120.3)	0.275	786 (14.0%)	9,883	0.0795	1.19 (1-7)
Bay of Plenty	3.77 (1.69)	195.4 (188.9)	154.0 (122.5)	0.251	454 (8.0%)	4,806	0.0945	1.20 (1-8)
Gisborne	5.24 (1.62)	266.9 (84.4)	205.6 (130.6)	0.232	42 (0.7%)	3,224	0.0130	1.17 (1-4)
Taranaki	4.11 (1.25)	213.3 (68.2)	160.8 (99.4)	0.227	104 (1.8%)	2,808	0.0370	1.12 (1-3)
Manawatu-Wanganui	4.36 (2.13)	230.0 (118.8)	163.8 (122.3)	0.228	288 (5.1%)	8,577	0.0336	1.13 (1-4)
Hawke's Bay	3.99 (0.38)	208.1 (75.9)	156.6 (100.2)	0.223	183 (3.3%)	5,469	0.0335	1.09 (1-3)
Wellington	5.44 (2.85)	285.7 (156.1)	198.1 (160.1)	0.258	325 (5.8%)	3,137	0.1036	1.18 (1-4)

Table 3. Fractional MDCEV Model Component Results

Variable	Estimate	t-stat.
Baseline utilities		
Logarithm of the area (miles^2) – mean	0.442	16.52
Land cover accessibility measure specific to		
Wetland (/10^3)	0.137	1.12
Urban (/10^4)	–1.690	–4.92
Forest (/10^4)	–0.159	–6.68
Agricultural (/10^4) x Income less than NZ \$50,000	0.043	2.89
Land-cover diversity index (LDI) × 10		
Constant	0.662	21.73
Interacted with Income less than NZ \$50,000	–0.343	–11.28
Logarithm of Price ($-1/\sigma$)	–0.375	–18.65
Satiation parameters (δ_k parameters)		
Land cover accessibility measure specific to		
Wetland (/10^3)	2.569	19.05
Land-cover diversity index (LDI) × 10		
Constant	–0.765	–6.95
Single person household	–0.926	–5.33
Income less then NZ \$50,000	0.698	4.07
Age ≥ 48 years	0.124	8.63
Constant	3.957	12.60

The 95% confidence level corresponds to an absolute t-statistic value of 1.96.

Table 4. Count Data Model Component Results

Variable	Estimate	t-stat.
α_0	9.675	32.62
α_1	0.516	8.80
Constant	0.016	0.28
Family Structure (nuclear/joint family is base)		
Single person household	0.549	8.70
Couple family household	0.222	4.42
Single parent household	0.222	3.17
Age ≥ 48 years	–0.559	–9.58
Linkage parameter	0.375	18.65

Table 5. Percentage Shifts in Total Visit Counts Due to a 15% Travel Cost Reduction to Northland and Wellington (numbers in parenthesis are 95% confidence bounds)

Destination Region	Linked Model Percentage Shift	Unlinked Model Percentage Shift
Northland	+10.0 (+6.6;+13.4)	+12.4 (+8.8;+16.0)
Auckland	–4.0 (+2.5;+5.2)	–2.5 (–1.3;+3.7)
Waikato	+0.7 (–0.4;+1.8)	–2.7 (–3.9; –1.5)
Bay of Plenty	–0.6 (–2.6;+1.4)	–3.5 (–1.3;+5.7)
Gisborne	–2.4 (–6.1;+1.3)	–4.0 (–7.5; –0.5)
Taranaki	–0.8 (–3.2;+1.6)	–4.7 (–7.0; –2.4)
Manawatu-Wanganui	+2.9 (–0.7,+6.5)	–3.4 (–7.0;+0.2)
Hawke’s Bay	+1.5 (–0.3,+3.3)	–3.5 (–5.2; –1.8)
Wellington	+13.7 (+9.2,+18.2)	+12.9 (+8.0;+17.8)

Table 6. Likelihood Based Data Fit Measures

Statistic	Estimation Sample (N = 5622)	
	Linked Model	Unlinked Model
Log-likelihood at Convergence	–11,854.38	–12,322.80
Log-likelihood of the base naïve model	–15958.10	
Number of parameters, not including the ln(area) coefficient in the baseline, the two α_l terms, and the constant in the $\boldsymbol{\varpi}$ vector	17	17
Nested Likelihood Ratio Test w.r.t the base naïve model	8,207	7,271
Adjusted Likelihood Ratio Index	0.256	0.227
Bayesian Information Criteria (BIC)	11,945	12,413
Non-Nested Likelihood Ratio Test between the Unlinked and Linked Models (informal test in hold-out sample)	Φ(-30.42) << 0.0001	

Table 7. Total Number of Visits to Each Destination Region

Destination Region	Actual	Linked Model Predictions and APE*		Unlinked Model Prediction and APE	
		Predicted Count	APE	Predicted Count	APE
Northland	335	417	24.5	435	29.9
Auckland	672	776	15.5	839	24.8
Waikato	936	847	9.5	801	14.5
Bay of Plenty	543	319	41.2	286	47.3
Gisborne	49	127	158.5	126	156.4
Taranaki	116	127	9.2	128	10.3
Manawatu-Wanganui	325	417	28.4	445	36.9
Hawke's Bay	200	252	25.8	258	29.0
Wellington	383	233	39.2	241	37.0
Weighted MAPE**		24.8%		29.9%	

Table 8. Total Number of Individuals Visiting Each Destination Region At Least Once

Destination Region	Actual Count	Linked Model Predictions and APE*		Unlinked Model Prediction and APE	
		Predicted Count	APE	Predicted Count	APE
Northland	289	305	5.4	318	10.0
Auckland	574	564	1.6	617	7.5
Waikato	786	622	20.9	591	24.8
Bay of Plenty	454	238	47.6	214	52.8
Gisborne	42	97	130.3	94	124.6
Taranaki	104	97	6.7	96	7.3
Manawatu-Wanganui	288	317	10.0	331	15.1
Hawke's Bay	183	191	4.5	192	5.2
Wellington	325	179	45.0	181	44.2
Weighted MAPE**		21.3%		25.1%	

*APE – Absolute Percentage Error ** Weighted MAPE – Weighted Mean APE