A Flexible Multiple Discrete-Continuous Probit (MDCP) Model: Application to Analysis of Expenditure Patterns of Domestic Tourists in India

Shobhit Saxena
Research Scholar
Department of Civil Engineering
Indian Institute of Science, Bengaluru, Karnataka, India, 560012
Email: shobhits@iisc.ac.in

Abdul Rawoof Pinjari (Corresponding author)
Associate Professor
Department of Civil Engineering
Centre for Infrastructure, Sustainable Transportation, and Urban Planning (C*i*STUP)
Indian Institute of Science (IISc)
Bengaluru 560012, India
Tel: +91-80-2293-2043
Email: abdul@iisc.ac.in

Chandra R. Bhat
Department of Civil, Architectural and Environmental Engineering
The University of Texas at Austin
301 E. Dean Keeton St. Stop C1761, Austin TX 78712
Email: bhat@mail.utexas.edu

Aupal Mondal
Ph.D. Student
Department of Civil, Architectural and Environmental Engineering
University of Texas at Austin, Austin, Texas, USA, 78712.
Email: aupal.mondal@utexas.edu

September 2022

ABSTRACT

Traditional multiple discrete-continuous (MDC) choice models impose tight linkages between consumers' discrete choice and the continuous consumption decisions due to the use of a single utility parameter driving both the decision to choose and the extent of choice. Recently, Bhat (2018) proposed a flexible MDCEV model that employs a utility function with separate parameters to determine the discrete choice and continuous consumption values. However, the flexible MDCEV model assumes an independent and identically distributed (IID) error structure across the discrete and continuous baseline utilities. In this paper, we formulate a flexible non-IID multiple discrete-continuous probit (MDCP) model that employs a multivariate normal stochastic distribution to allow for a more general variance-covariance structure. In doing so, we revisit Bhat's (2018) flexible utility functional form and highlight that the stochastic conditions he used to derive the likelihood function are not always consistent with utility maximization. We offer an alternate interpretation of the model as representing a two-step decision-making process, where the consumers first decide which goods to choose and then decide the extent of allocation to each good. We demonstrate an application of the proposed flexible MDCP model to analyze households' expenditure patterns on their domestic tourism trips in India. Our results indicate that, if the analyst is willing to compromise on the strict utility-maximizing aspect of behavior, while also enriching the behavioral dimension through the relaxation of the tie between the discrete and continuous consumption decisions, the preferred model would be the flexible non-IID MDCP model. On the other hand, if the analyst wants the model to be strictly grounded on utility-maximizing behavior (which may also have benefits by way of welfare measure computations), and is willing to assume a very tight tie between the discrete and continuous consumption decision processes, the preferred model would be the non-IID traditional MDCP model.

Keywords: multiple discrete-continuous choice models, flexible utility functions, probit kernel, domestic tourism in India, tourists' expenditures

1. INTRODUCTION

Choice situations characterized by the choice of multiple alternatives at the same time (as opposed to the choice of a single alternative) are ubiquitous in consumer decision behavior. Hendel (1999) coined the term “multiple discreteness” (MD) to refer to such consumer choice situations, while Bhat (2005, 2008) proposed the term “multiple discrete-continuous” (MDC) to refer to situations where the consumer also decides on a continuous dimension (or quantity) of consumption. The consideration of MDC models to analyze consumer choice situations has exploded in recent years, with studies in environmental economics, regional science, transportation, marketing, and many other fields using a utility-maximizing approach to analyze such situations (see, for example, von Haefen and Phaneuf, 2005, Wafa *et al.,* 2015, Satomura *et al.,* 2011, Yonezawa and Richards, 2017, Sobhani *et al.,* 2013, Khan and Machemehl, 2017, Calastri *et al.,* 2017, and Enam *et al.,* 2018).

The basic approach in a utility-maximizing framework for MDC choices utilizes a non-linear (but quasi-concave, increasing and continuously differentiable) utility structure with decreasing marginal utility (or satiation). The approach, originally proposed by Wales and Woodland (1983) (see also Hanemann, 1984, Kim *et al.,* 2002; von Haefen and Phaneuf, 2003; Bhat, 2005), assumes that consumers maximize the utility obtained from the MDC consumption, subject to a budget constraint. The optimal consumption quantities (including possibly zero consumptions of some alternatives, creating the discrete choice dimension) are obtained by writing the first-order optimality conditions for the utility function. To simplify the analysis, it is common practice to consider an additively separable utility structure in which the rate of substitution between any pair of goods is dependent only on the quantities of the two goods in the pair, and independent of the quantity of other goods (see Pollak and Wales, 1992).

Among the many general additively separable utility structures, the Box-Cox utility function form proposed by Bhat (2008) subsumes several other non-linear forms as special cases and allows a clear interpretation of model parameters. But, in Bhat’s (2008) formulation, as in all other MDC formulations until recently, the discrete and continuous choice decisions are tied very tightly together because the same baseline utility preference parameter influences both the choice of making a positive consumption of a good (the discrete choice) as well as the starting point for satiation effects (that impact the continuous choice). [1] Recently, Bhat (2018) provides at least a

[1] The baseline preference parameter of a good is the marginal utility at zero consumption of that good.

couple of reasons why this is generally not likely to be the case, either due to (a) simply a need for variety-seeking that implies the consumption of certain goods, but at very low continuous quantities, or (b) a branding effect (that is, a prestige/image effect) that operates at the pure discrete level but does not necessarily carry over with the same intensity to the continuous consumption decision. Besides, to the extent that these variety-seeking and branding effects may vary across demographic groups, this immediately implies that demographic factors may have differential effects on the underlying preferences for the discrete and continuous choices. Bhat (2018) then proceeds to formulate a new MDC model that introduces separate baseline preference parameters for discrete and continuous dimensions of choice – called the *discrete preference parameters* and *continuous preference parameters* – to break the tight linkage between the discrete and continuous choice dimensions. In terms of the stochastic structures for the baseline preference parameters, he uses extreme-value error distributions. More importantly, he assumes independence and identical distribution (IID) across the error terms of different alternatives in both the discrete and continuous dimensions, as well as across the error terms of the discrete and continuous dimensions of the same alternative. However, there are several reasons why the IID assumption may not hold. It is likely that the error terms of different choice alternatives are correlated due to the influence of common unobserved factors. Also, in the flexible utility structure, the discrete and continuous dimensions of a choice alternative may be influenced by common unobserved factors that cause correlation between the discrete and continuous preference parameters. For example, the decision to consume in an activity category, say shopping, can be correlated with how much to consume in that category because of unobserved factors that make a person more or less likely to enjoy that activity. However, the IID assumption misses out on capturing such correlations. Although multivariate extreme value (MEV) distributions have been used in the past for allowing a general covariance structure in Bhat's (2008) Box-Cox utility function form (Pinjari and Bhat, 2010, and Pinjari, 2011), it becomes difficult to do so in the flexible utility form. This is because it is not easy to derive the likelihood function for a model with MEV distributed error terms in both the discrete preference parameters and continuous preference parameters.

In this paper, we apply a probit-based error kernel (rather than the extreme-value kernel) that allows non-IID stochastic structure to the flexible utility functional form proposed by Bhat (2018). Such a "flexible MDCP" model proposed in this paper relaxes the assumptions of independence among the utility parameters using a multivariate normal (MVN) error kernel. In this process, we revisit Bhat's (2018) flexible utility form and discuss a few nuances regarding the role of the discrete preference parameters in the optimality conditions associated with the utility profile. Specifically, we highlight the point that the discrete preference parameters in his formulation do not appear in the conditions for optimal utility. That is, although the introduction of separate parameters for the discrete and continuous dimensions helps break the tight linkage between the discrete and continuous preferences, the conditions he uses on the discrete preference parameters are not always consistent with utility maximization. As a result, Bhat's (2018) approach can be interpreted as an externally imposed set of conditions (external to utility maximization) on the discrete preference parameters that determine whether an alternative is chosen. Doing so helps in separating the effects of exogenous variables on the discrete decision to consume a good from the decision of how much to consume that good.

We apply the flexible MDCP model developed in this paper (and its simpler versions), as well as the traditional MDCP model that does not include a separate set of discrete preference parameters, to an empirical analysis of tourist's expenditure allocations on domestic recreational travel in India. The model fit statistics of the various formulations are compared (in both estimation and holdout datasets) to highlight the value of: (a) separate parameters for discrete and continuous preferences (even if the formulation does not always conform to utility maximization), and (b) relaxing the IID assumption in the error kernel of the model with flexible utility functions. Further, the empirical analysis sheds light on the determinants of Indian households' expenditure patterns on their recreational/tourism trips.

The rest of the paper is structured as follows. Section 2 discusses the nuances in the formulation of the flexible MDC choice model and why the model is not necessarily consistent with utility maximization. Further, an alternate interpretation of the flexible MDC model is presented. Subsequently, the section presents the formulation of the flexible MDCP model with MVN stochastic distribution. Section 3 illustrates an application of the flexible MDCP model and its variants for analyzing Indian households' expenditures on their domestic tourism trips. The final section offers concluding thoughts and directions for further research.

2. MODEL FORMULATION

2.1. Revisiting Bhat's (2018) Flexible MDC Model Structure

Consider the case of an incomplete demand with an essential, numeraire Hicksian outside good with a linear utility profile and multiple non-essential, inside goods. To model such a system, Bhat (2018) proposes the following utility form (here, we suppress the index for individuals)[2]:

$$U(\boldsymbol{x}) = \psi_1 x_1 + \sum_{k=2}^{K} \gamma_k \left\{ \left(\psi_{kd}\right)^{1[x_k=0]} \times \left(\psi_{kc}\right)^{1[x_k>0]} \right\} \ln\left(\frac{x_k}{\gamma_k} + 1\right) \tag{1}$$

subject to a linear budget constraint and non-negativity constraints on consumptions, as below:

$$\sum_{k=1}^{K} p_k x_k = E,\; x_k \geq 0 \;\forall\; k = 2,3,....,K \tag{2}$$

$U(\boldsymbol{x})$ in Equation (1) is quasi-concave, increasing and continuously differentiable utility function, where $\boldsymbol{x}$ is a $(K \times 1)$ dimension vector of consumption quantities), ψ_1 is the baseline marginal utility for the essential outside good, and ψ_{kd} , ψ_{kc}, and γ_k are parameters associated with inside goods (k =*2,3,…K*). The expression for $U(\boldsymbol{x})$ is a valid utility function if $\psi_1 > 0$, $\psi_{kc} > 0$, and $\gamma_k > 0$, for all k. The equality constraint in Equation (2) is the linear budget constraint, where E is the total expenditure across all goods k (k = 1, 2,…, K) and $p_k > 0$ is the unit price of good k (with $p_1 = 1$ to represent the numeraire nature of the essential outside good). The total expenditure across all goods (*i.e., E*) is assumed to be large relative to the total allocation to all inside goods. This assumption is due to the linear outside-good utility profile employed in the flexible utility form of Equation (1). Specifically, as discussed in Saxena et al. (2022), the use of a linear utility profile for the outside good results in a likelihood function that does not include the outside good value or the total budget value (E). Such models can be interpreted as if the overall budget is very large relative to the total expenditure allocation to all inside goods. In the current empirical context, it is safe to make this assumption because

[2] The discussion here pertains to the $\gamma-$profile utility form since it has been generally the case that the $\gamma-$profile comes out to be superior to the $\alpha-$profile function (see Bhat, 2018; Jian *et al.,* 2017). However, the discussion in the rest of the paper, including the model formulation, is applicable for the $\alpha-$profile utility function as well.

households/individuals allocate only a small portion of their overall incomes (i.e., budget) to a single vacation trip (i.e., total allocation to all inside goods).

As can be observed, the utility function is linear with respect to consumption of the outside good, implying no satiation in the consumption of the outside good. Thus, multiple discreteness occurs because of satiation effects on the consumption of inside goods (k = 2,3,…,K). For inside goods, the ψ_{kd} parameter corresponds to the baseline preference that determines whether good k is consumed (this is the discrete preference baseline utility component, or simply the D-preference parameter). The ψ_{kc} parameter, on the other hand, corresponds to the baseline preference if good k is consumed; this is, the continuous preference baseline utility component, or simply the C-preference parameter. The exponents in the utility function $1[x_k = 0]$ and $1[x_k > 0]$ are indicators representing situations when $x_k = 0$ and $x_k > 0$, respectively. Note, however, that the utility function is continuously differentiable at zero consumption as well as all other positive consumption values. Specifically, the derivative of the utility function (marginal utility) with respect to consumption is ψ_{kc} at 0 as well as at 0^+ (i.e., at an infinitesimally small consumption value above zero).[3] Finally, γ_k is the vehicle to introduce corner solutions (aka, zero consumptions) for inside goods (k = 2,3,…,K), and also serves the role of a satiation parameter (higher values of γ_k imply less satiation). There is no γ_1 term for the first good because it is always consumed, and there is no satiation for this good.

Bhat proposes the following as optimality conditions for the utility form in Equation (1):

3

$$\text{At } x_k = 0,\quad \left.\frac{\partial U(\boldsymbol{x})}{\partial x_k}\right|_{x_k=0} = \lim_{h\to 0^+} \left.\frac{U(\boldsymbol{x}+h)-U(\boldsymbol{x})}{h}\right|_{x_k=0} = \lim_{h\to 0^+} \frac{U(\boldsymbol{x}+h)\big|_{x_k=0} - U(\boldsymbol{x})\big|_{x_k=0}}{h} = \lim_{h\to 0^+} \frac{U(\boldsymbol{x}+h)\big|_{x_k=0} - U(\boldsymbol{x})\big|_{x_k=0}}{h}$$

$$= \lim_{h\to 0^+} \frac{\psi_{kc}\gamma_k \ln\left(\dfrac{h}{\gamma_k}+1\right) - 0}{h} = \lim_{h\to 0^+} \frac{\dfrac{\partial}{\partial h}\psi_{kc}\gamma_k \ln\left(\dfrac{h}{\gamma_k}+1\right)}{\dfrac{\partial}{\partial h}h} = \psi_{kc}.\quad \text{Note:} U(\boldsymbol{x}+h) = U(x_1, x_2, ..., (x_k+h), ..., x_K).$$

$$\text{At } x_k > 0,\quad \left.\frac{\partial U(\boldsymbol{x})}{\partial x_k}\right|_{x_k=0^+} = \lim_{x\to 0^+} \psi_{kc}\left(\frac{x_k}{\gamma_k}+1\right)^{-1} = \psi_{kc}.$$

Given the above, the ψ_{kd} parameters do not enter the marginal utility functions of Bhat’s (2018) utility form, which as we discuss later, has implications for whether the formulation is always consistent with utility maximization.

$$\psi_{kd} - \lambda p_k > 0 \text{ and } (\psi_{kc})\left(\frac{x_k^*}{\gamma_k}+1\right)^{-1} = \lambda p_k \text{ if } x_k^* > 0, k = 2,3,...K$$
$$\psi_{kd} - \lambda p_k < 0 \text{ if } x_k^* = 0,\ k = 2,3,...,K$$
$$\psi_1 = \lambda$$

One can observe from the inequalities ($(\psi_{kd} - \lambda p_k > 0, \text{ if } x_k^* > 0)$ and $(\psi_{kd} - \lambda p_k < 0, \text{ if } x_k^* = 0)$) in the above conditions that only the *D*-preference parameter of an inside good is used to determine whether that good is chosen or not (the *C*-preference parameters do not play a role in the discrete choice). The *C*-preference parameters come into the picture to determine the consumed quantity of the chosen goods.[4] The inequality conditions based on *D*-preference parameters to determine whether a good is chosen or not, combined with the equality conditions based on *C*-preference parameters to determine the extent of consumption of chosen goods, help in estimating the *D*-preference parameters separately from the *C*-preference parameters.

It is important to note here that the conditions in Equation (3) proposed by Bhat (2018) do not always ensure optimality of the utility function in Equation (1). This is because, as discussed in Footnote 3, the *D*-preference parameters (ψ_{kd}) do not have any role to play in the marginal utility functions, even at zero consumptions. As a result, the *D*-preference parameters are not relevant to maximization of the utility function in Equation (1). Therefore, the conditions in Equation (3) based on the *D*-preference parameters are not necessarily consistent with maximization of the utility function. For example, according to Equation (3), $\psi_{kd} < \lambda p_k$ implies that good *k* is not chosen. However, if $\psi_{kc} > \lambda p_k$, choosing good k will lead to greater utility (for the utility function in Equation (1)) than not choosing it; regardless of the value of ψ_{kd}. Similarly, when $\psi_{kd} > \lambda p_k$ but $\psi_{kc} < \lambda p_k$ the conditions in Equation (3) can lead to suboptimal consumptions. On the other hand, if the analyst were to work with only the conditions of optimal utility, the resulting model would not have the *D*-preference parameters. Instead, it would be similar that of a traditional MDC choice model, except with a linear utility for the outside good.

[4] Note also that the equality condition for the chosen goods, which is based on the *C*-preference parameters, automatically implies an inequality that $\psi_{kc} - \lambda p_k > 0$. Such an inequality based on the *C*-preference parameters is not explicitly stated in the above conditions since it would be redundant.

To be sure, Bhat's (2018) flexible formulation, while valuable in loosening the tight tie between the discrete and continuous consumption decisions, is not always going to be optimal from a strict utility-maximizing perspective. In this context, one can interpret the conditions in Equation (3) involving ψ_{kd} as a set of heuristics that are assumed to be followed by individuals in making their decisions of which goods to choose and how much to consume of the chosen goods. That is, the model in Equation (3) can be interpreted as a two-step decision-making process somewhat akin to a Tobit model for each inside good. Specifically, for a given inside good *k*, in the first step, the discrete choice decision is made based on its ψ_{kd} value. In the second step, the continuous consumption amount is determined based on the ψ_{kc} and γ_k values of that good. While the continuous consumption decision is still based on an intuitive and behavioral utility satiation concept, the flexible MDC formulation extends the traditional utility-maximizing MDC formulation in ways that allow the statistical stitching of the discrete and continuous consumption decisions, as we discuss in the rest of this paper.

2.2 The Flexible MDCP Model

2.2.1. Model Structure

The conditions of Equation (3) may be rewritten, after substituting $\lambda = \psi_1$, as:

$$\begin{aligned} &\psi_{kd} - \psi_1 p_k > 0 \ \text{ and } \ (\psi_{kc})\left(\frac{x_k^*}{\gamma_k}+1\right)^{-1} = \psi_1 p_k \ \text{ for } k = 2,\ldots,K \text{ when } x_k^* > 0 \\ &\psi_{kd} - \psi_1 p_k < 0, \text{if } \ x_k^* = 0 \ \text{ for } k = 2,\ldots,K \end{aligned} \tag{4}$$

Re-arranging and taking logarithms, the conditions may be rewritten as follows:

$$\begin{aligned} &\ln(\psi_{kd}) - \ln(\psi_1) - \ln p_k > 0 \ \text{ and } \ln(\psi_{kc}) - \ln\left(\frac{x_k^*}{\gamma_k}+1\right) - \ln(\psi_1) - \ln p_k = 0 \ \text{ when } x_k^* > 0, \\ &\ln(\psi_{kd}) - \ln(\psi_1) - \ln p_k < 0 \text{ when } x_k^* = 0 \end{aligned} \tag{5}$$

Next, we specify ψ_1 as $\psi_1 = \exp(\boldsymbol{\beta}'\boldsymbol{z}_1 + \varepsilon_1)$, where $\boldsymbol{z}_1$ is a *D*-dimensional vector of attributes that characterizes good *1*, without the inclusion of a constant. Similarly, the *D*-preference and the *C*-preference terms are specified as follows:

$$\psi_{kd} = \exp(\boldsymbol{\beta}'\boldsymbol{z}_k + \varepsilon_k) \text{ and } \psi_{kc} = \exp(\boldsymbol{\theta}'\boldsymbol{w}_k + \xi_k)\ ,\ k = 2,3,\ldots,K. \tag{6}$$

In the expression for ψ_{kd}, $\boldsymbol{z}_k$ is a D-dimensional vector of observed attributes that influence the choice of good k (including a dummy variable for each inside good), $\boldsymbol{\beta}$ is a corresponding vector of coefficients, and ε_k captures the idiosyncratic (unobserved) characteristics influencing the choice of good k. We assume that the error terms ε_k are multivariate normally distributed across goods k: $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_K)' \sim MVN_K(\boldsymbol{0}_K, \tilde{\boldsymbol{\Lambda}})$, where $MVN_K(\boldsymbol{0}_K, \tilde{\boldsymbol{\Lambda}})$ indicates a K-variate normal distribution with a mean vector of zeros denoted by $\boldsymbol{0}_K$ and a covariance matrix $\tilde{\boldsymbol{\Lambda}}$. In the expression for ψ_{kc}, $\boldsymbol{w}_k$ is a vector of observed attributes that influence the consumption amount of inside good $k\,(k = 2, \ldots, K)$, with $\boldsymbol{\theta}$ being the corresponding vector of coefficients. Note that some variables may appear exclusively in $\boldsymbol{z}_k$ or in $\boldsymbol{w}_k$, indicating that the variable affects one but not the other baseline preference. We assume that the error terms $\boldsymbol{\xi} = (\xi_2, ..., \xi_K)'$ are also multivariate ((K-1)-variate to be precise) normally distributed across goods k: $\boldsymbol{\xi} \sim MVN_{K-1}(\boldsymbol{0}_{K-1}, \boldsymbol{\Omega})$. Next, define the following for the inside goods $k\,(k = 2, \ldots, K)$:

$$\begin{aligned}
&\eta_k = \varepsilon_k - \varepsilon_1 \text{ and } \varsigma_k = \xi_k - \varepsilon_1, \\
&\tilde{V}_{k,1} = \boldsymbol{\beta}'\boldsymbol{z}_1 - \left(\boldsymbol{\beta}'\boldsymbol{z}_k - \ln p_k\right), \text{ and,} \\
&\vec{V}_{k,1} = \boldsymbol{\beta}'\boldsymbol{z}_1 - \left(\boldsymbol{\theta}'\boldsymbol{w}_k - \ln p_k\right) + \ln\left(\frac{x_k^*}{\gamma_k} + 1\right)
\end{aligned} \tag{7}$$

Then, the conditions in Equation (5) may be rewritten as:

$$\begin{aligned}
&\eta_k > \tilde{V}_{k,1} \text{ and } \varsigma_k = \vec{V}_{k,1} \text{ if } x_k^* > 0, \ \left(k = 2, 3, \ldots, K\right) \\
&\eta_k < \tilde{V}_{k,1} \text{ if } x_k^* = 0, \left(k = 2, 3, \ldots, K\right)
\end{aligned} \tag{8}$$

For identification, we set $Var(\varepsilon_1) = 0.5$, $Var(\varepsilon_2) = 0.5$, and $\text{cov}(\varepsilon_1, \varepsilon_2) = 0$. This normalizes the first element of the covariance matrix of η_k terms to 1, which constitutes a scale normalization for the error terms in the D-preference parameters. In addition, for a straightforward interpretation of the covariance matrices $\tilde{\boldsymbol{\Lambda}}$ and $\boldsymbol{\Omega}$, as well as the covariances between the ε_k and ξ_k terms, we will assume that the stochastic term influencing the baseline utility for the outside good ε_1 is independent of all inside good error terms ε_k $(k = 2, 3, ...K)$ in the D-

preference parameters as well as the inside good error terms ξ_k $(k=2,3,....,K)$ in the *C*-preference parameters. As such, the assumption is innocuous because the full covariance matrix of the ε_k terms and ξ_k terms is unidentifiable. Using the assumption of error independence between the outside good utility and the inside good utility parameters, and using the notation $\mathbf{1}_M$ to refer to a square matrix of dimension *M* with all its elements equal to 1, the covariance matrix for η_k and ς_k takes the following form, where $\boldsymbol{\eta}=(\eta_2,\eta_3,...,\eta_K)'$ and $\boldsymbol{\varsigma}=(\varsigma_2,\varsigma_3,...,\varsigma_K)'$:

$$\boldsymbol{\Xi}=Cov(\boldsymbol{\eta},\boldsymbol{\varsigma})=\ 0.5\times\mathbf{1}_{2(K-1)}\ +\tilde{\boldsymbol{\Xi}}\ ,$$

$$\text{where } \tilde{\boldsymbol{\Xi}}=\begin{bmatrix}\vec{\boldsymbol{\Lambda}} & \boldsymbol{\Sigma}\\ \boldsymbol{\Sigma}' & \boldsymbol{\Omega}\end{bmatrix}=\left[\begin{array}{ccccc|ccccc} 0.5 & \tilde{\lambda}_{23} & \tilde{\lambda}_{24} & \cdots & \tilde{\lambda}_{2K} & \sigma_{22} & \sigma_{23} & \sigma_{24} & \cdots & \sigma_{2K}\\ \tilde{\lambda}_{23} & \tilde{\lambda}_{33} & \tilde{\lambda}_{34} & \cdots & \tilde{\lambda}_{3K} & \sigma_{32} & \sigma_{33} & \sigma_{34} & \cdots & \sigma_{3K}\\ \tilde{\lambda}_{24} & \tilde{\lambda}_{34} & \tilde{\lambda}_{44} & \cdots & \tilde{\lambda}_{4K} & \sigma_{42} & \sigma_{43} & \sigma_{44} & \cdots & \sigma_{4K}\\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ \tilde{\lambda}_{2K} & \tilde{\lambda}_{3K} & \tilde{\lambda}_{4K} & \cdots & \tilde{\lambda}_{KK} & \sigma_{K2} & \sigma_{K3} & \sigma_{K4} & \cdots & \sigma_{KK}\\ \hline \sigma_{22} & \sigma_{32} & \sigma_{42} & \cdots & \sigma_{K2} & \omega_{22} & \omega_{23} & \omega_{24} & \cdots & \omega_{2K}\\ \sigma_{23} & \sigma_{33} & \sigma_{43} & \cdots & \sigma_{K3} & \omega_{23} & \omega_{33} & \omega_{34} & \cdots & \omega_{3K}\\ \sigma_{24} & \sigma_{34} & \sigma_{44} & \cdots & \sigma_{K4} & \omega_{24} & \omega_{34} & \omega_{44} & \cdots & \omega_{4K}\\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots\\ \sigma_{2K} & \sigma_{3K} & \sigma_{4K} & \cdots & \sigma_{KK} & \omega_{2K} & \omega_{3K} & \omega_{4K} & \cdots & \omega_{KK} \end{array}\right] \quad (9)$$

In the above matrix, $\tilde{\boldsymbol{\Xi}}$ is the variance-covariance matrix of all (but ε_1) error terms in the model and $\boldsymbol{\Xi}$ is the variance-covariance matrix of error differences with respect to ε_1. Note that $\vec{\boldsymbol{\Lambda}}$ is the covariance matrix of the *D*-preference error terms for inside alternatives ε_k $(k=2,3,....K)$, with the elements $\tilde{\lambda}_{kk}$ $(k=2,3,....K)$ denoting the variances of these error terms and the elements $\tilde{\lambda}_{jk}$ $(j,k=2,3,....K;\ j\neq k)$ denoting the covariances between ε_j and ε_k $(j,k=2,3,...K;\ j\neq k)$. $\boldsymbol{\Omega}$, as already defined, is the covariance matrix of the *C*-preference error terms for the inside goods ξ_k $(k=2,3,....K)$, with the elements ω_{kk} $(k=2,3,....K)$ denoting the variances of these error terms and the elements ω_{jk} $(j,k=2,3,....K;\ j\neq k)$ denoting the covariances. Finally, $\boldsymbol{\Sigma}$ captures the covariances between the *D*-preference and *C*-preference error terms, with $\sigma_{kk}(k=2,3,...K)$ denoting the covariances between ε_k and ξ_k $(k=2,3,...K)$ and σ_{jk} $(j,k=2,3,....K;\ j\neq k)$ denoting the covariances between ε_j and ξ_k $(j,k=2,3,...K;\ j\neq k)$.

The reader will note that Ξ is positive definite as long as the covariance matrix $\tilde{\Xi}$ is positive definite, because the addition of a non-negative constant to all entries of a positive definite matrix also results in a positive definite matrix. The positive definiteness of $\tilde{\Xi}$ is ensured by applying a Cholesky decomposition of this matrix and estimating the corresponding Cholesky parameters. Once the Cholesky parameters are estimated, the actual covariance parameters can be easily obtained. For later use, we will also write the following equation:

$$\Xi = Cov(\boldsymbol{\eta},\boldsymbol{\varsigma}) = \ 0.5\times\mathbf{1}_{2(K-1)} \ + \tilde{\Xi} = \left[\begin{array}{c|c} 0.5\times\mathbf{1}_{(K-1)}+\tilde{\Lambda} & 0.5\times\mathbf{1}_{(K-1)}+\Sigma \\ \hline 0.5\times\mathbf{1}_{(K-1)}+\Sigma' & 0.5\times\mathbf{1}_{(K-1)}+\Omega \end{array}\right] = \left[\begin{array}{c|c} \Xi_{11} & \Xi_{12} \\ \hline \Xi'_{12} & \Xi_{22} \end{array}\right] \qquad (10)$$

An important implication of the conditions in Equation (8) on the identification of the covariance elements is worth discussing. Specifically, the conditions in Equation (8) imply that the discrete choice probability of consumption of each inside good is essentially a binary probit. That is, for each inside good, the condition for choosing a good for consumption is given by the univariate probit model condition of $\eta_k < \tilde{V}_{k,1}$ if $x_k^* = 0$ and $\eta_k > \tilde{V}_{k,1}$ if $x_k^* > 0$. Since in such binary choice situations, it is not possible to estimate scale parameters of the error terms, the variance elements (*i.e.,* $\tilde{\lambda}_{kk}$ $(k = 2,3,....K)$) are all set to 0.5 in the covariance matrix in Equation (9).

2.2.2. Model Estimation

Consider an individual who chooses a total of *M* inside goods to consume (*M* can take the value of zero, indicating that no inside goods are consumed, and all the budget is invested in the outside good). Define a selection matrix $\mathbf{R}$ of size $(K-1+M)\times[2(K-1)]$. Consider the first (*K*-1) rows. If $M > 0$, in the first of these rows, place a value of '1' in the column corresponding to the first of the inside goods consumed. Place values of zeros everywhere else in this row. Next, in the second of these rows, place a value of '1' in the column corresponding to the second of the inside goods consumed and zeros everywhere else. Continue until the M^{th} row. Next, start with the $(M+1)^{th}$ row, place a value of '1' in the column corresponding to the first non-consumed inside good, and values of '0' everywhere else. Continue this until the first (*K*-1) rows are completely populated. Next, in the K^{th} through $(K-1+M)^{th}$ rows and the K^{th} through $2(K-1)^{th}$ columns, reproduce the sub-matrix in the first M rows and (*K*-1) columns of $\mathbf{R}$. All other columns of the K^{th} through $(K-1+M)^{th}$ rows receive a value of zero. Thus, if there are three inside goods (that is, (K-1) = 3), and assuming that the first and the third of those goods are consumed by an individual, the $\mathbf{R}$ matrix takes the following form:

$$\mathbf{R}=\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right]=\left[\begin{array}{c|c} \mathbf{R}_{D1} & \mathbf{0} \\ \hline \mathbf{R}_{D2} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{R}_C \end{array}\right]=\left[\begin{array}{c|c} \mathbf{R}_D & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{R}_C \end{array}\right]. \tag{11}$$

If none of the inside goods are consumed, the $\mathbf{R}$ matrix takes the following form:

$$\mathbf{R}=\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{array}\right]=\left[\mathbf{R}_{D2} \;\middle|\; \mathbf{0}\right]=\left[\mathbf{R}_D \;\middle|\; \mathbf{0}\right] \tag{12}$$

Next, also define a set of dummy variables ϑ_k $(k = 2,3\ldots K)$ that take the value of '1' if inside good k is consumed, and zero otherwise. Let $\boldsymbol{\vartheta}=(\vartheta_2,\vartheta_3...,\vartheta_K)$, $\tilde{\boldsymbol{V}}=(\tilde{V}_{2,1},\tilde{V}_{3,1},...\tilde{V}_{K,1})$, $\tilde{\boldsymbol{V}}_{D1}=\mathbf{R}_{D1}\tilde{\boldsymbol{V}}$, $\tilde{\boldsymbol{V}}_{D2}=\mathbf{R}_{D2}\tilde{\boldsymbol{V}}$, $\vec{\boldsymbol{V}}=(\vec{V}_2,\vec{V}_3,...,\vec{V}_K)'$, $\vec{\boldsymbol{V}}_C=\mathbf{R}_C\vec{\boldsymbol{V}}$, $\tilde{\boldsymbol{\varsigma}}=\mathbf{R}_C\boldsymbol{\varsigma}$, $\boldsymbol{\eta}_{D1}=\mathbf{R}_{D1}\boldsymbol{\eta}$, $\boldsymbol{\eta}_{D2}=\mathbf{R}_{D2}\boldsymbol{\eta}$, and $\boldsymbol{\Theta}=\mathbf{R}\boldsymbol{\Xi}\mathbf{R}'$. Also, let the actual observed consumption vector be $\boldsymbol{x}^*=(x_2^*,x_3^*,...,x_K^*)$, where some or all of the inside good-specific consumption values $(x_k^*;\,k=2,3,...K)$ may be zero (the outside good consumption value is relevant only in situations with finite budgets, where it can be determined immediately from the consumption values of the inside goods, and so is not included explicitly in the consumption vector $\boldsymbol{x}^*$). Then, based on the KKT conditions in Equation (8), we may write the following:

$$P\left(\boldsymbol{x}^*\right)=|J|\int_{\boldsymbol{\eta}_{D1}=\tilde{\boldsymbol{V}}_{D1}}^{\boldsymbol{\eta}_{D1}=\infty}\int_{\boldsymbol{\eta}_{D2}=-\infty}^{\boldsymbol{\eta}_{D2}=\tilde{\boldsymbol{V}}_{D2}} \boldsymbol{f}_{K-1+M}(\boldsymbol{\eta}_{D1},\boldsymbol{\eta}_{D2},\vec{\boldsymbol{V}}_C;\mathbf{0}_{K-1+M},\boldsymbol{\Theta})\,d\boldsymbol{\eta}_{D2}d\boldsymbol{\eta}_{D1}, \tag{13}$$

where $|J| = \left[\prod_{k=2}^{K}(f_k)^{\vartheta_k}\right]$, $f_k=\left(\dfrac{1}{x_k^*+\gamma_k}\right)(k=2,3,....K)$, and $\boldsymbol{f}_{K-1+M}(.,.,...;\mathbf{0}_{K-1+M},\boldsymbol{\Theta})$ represents the (K-1+M)-variate multivariate normal density function (pdf) with a mean vector of $\mathbf{0}_{K-1+M}$ and covariance matrix $\boldsymbol{\Theta}$.[5] Due to the symmetric nature of the mean-centered multivariate normal density function, the expression in Equation (13) may be simplified by constructing a (K-1+M) column vector $\boldsymbol{\delta}$ with entries of '1' in the first M rows and entries of '0' in the remaining rows.[6] Then, define $\tilde{\boldsymbol{\eta}} = (\text{-}1)^{\delta} .* (\boldsymbol{\eta}'_{D1}, \boldsymbol{\eta}'_{D2})'$. Next, define a $(K-1+M)\times(K-1+M)$ matrix $\boldsymbol{\Psi} = (-1)^{\delta+\delta'} .* \boldsymbol{\Theta}$, where the $(-1)^{\delta+\delta'}$ refers to a matrix whose elements correspond to the value of '-1' raised to the elements of the matrix $\boldsymbol{\delta}+\boldsymbol{\delta}'$, and '.*' refers to the element by element multiplication of two matrices. Then, by construction and symmetry, $\boldsymbol{\Psi} = Cov(\tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\varsigma}})$. Let $\boldsymbol{S} = (-\tilde{\boldsymbol{V}}'_{D1}, \tilde{\boldsymbol{V}}'_{D2})'$. Then, we may write Equation (13) compactly as:

$$P\left(\boldsymbol{x}^*\right) = |J| \text{ Prob}[\tilde{\boldsymbol{\eta}} < \boldsymbol{S}, \tilde{\boldsymbol{\varsigma}} = \vec{\boldsymbol{V}}_C] = |J| \int_{\tilde{\boldsymbol{\eta}}=-\infty}^{\tilde{\boldsymbol{\eta}}=\boldsymbol{S}} \boldsymbol{f}_{K-1+M}(\tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\varsigma}} = \vec{\boldsymbol{V}}_C; \mathbf{0}_{K-1+M}, \boldsymbol{\Psi})\, d\tilde{\boldsymbol{\eta}}. \tag{14}$$

To further simplify the expression above, partition $\boldsymbol{\Psi}$ as follows:

$$\boldsymbol{\Psi} = Cov(\tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\varsigma}}) = \begin{bmatrix} \boldsymbol{\Psi}_{11} & \boldsymbol{\Psi}_{12} \\ \boldsymbol{\Psi}'_{12} & \boldsymbol{\Psi}_{22} \end{bmatrix}, \tag{15}$$

where $\boldsymbol{\Psi}_{11}$ is a $(K-1)\times(K-1)$ sub-matrix comprising the first (K-1) rows and (K-1) columns of $\boldsymbol{\Psi}$ (corresponding to the covariance matrix of $\tilde{\boldsymbol{\eta}}$), $\boldsymbol{\Psi}_{22}$ is a $M\times M$ sub-matrix comprising the last M rows and M columns of $\boldsymbol{\Psi}$ (corresponding to the covariance matrix of $\tilde{\boldsymbol{\varsigma}}$), and $\boldsymbol{\Psi}_{12}$ is a

---

[5] In the special case that M=0 (that is, none of the inside goods is consumed), Equation (13) collapses to the following:

$$P\left(\boldsymbol{x}^*\right) = P(0,0,0,...0) = \int_{\eta_{D2}=-\infty}^{\eta_{D2}=\tilde{V}_{D2}} \boldsymbol{f}_{K-1+M}(\boldsymbol{\eta}_{D2}; \mathbf{0}_{K-1+M}, \boldsymbol{\Theta})\, d\boldsymbol{\eta}_{D2},$$

which takes the form of a simple multivariate cumulative normal distribution function.

[6] The symmetric nature of the multivariate normal density function is a distinct advantage over the asymmetric multivariate logistic density function used in Bhat (2018). While the multivariate logistic has a closed form expression for the cumulative distribution function, computing the integral of the multivariate logistic density function with a combination of upper and lower limits cannot be collapsed to the evaluation of a single cumulative distribution function. However, while the multivariate normal cumulative distribution (MVNCD) function is not available in closed-form, the integral of the multivariate normal density function with a combination of upper and lower limits can be collapsed to the evaluation of a single MVNCD function. This is a particularly useful result for the proposed flexible MDCP model.

$(K-1)\times M$ sub-matrix comprising the first (K-1) rows and last M columns of $\boldsymbol{\Psi}$ (corresponding to the covariance between the $\tilde{\boldsymbol{\eta}}$ and $\tilde{\boldsymbol{\varsigma}}$ vectors). Define the following: $\boldsymbol{W}=\boldsymbol{\Psi}_{12}\boldsymbol{\Psi}_{22}^{-1}\vec{\boldsymbol{V}}_C$ $[(K-1)\times 1$ vector$]$, $\boldsymbol{\Delta}=\boldsymbol{\Psi}_{11}-\boldsymbol{\Psi}_{12}\boldsymbol{\Psi}_{22}^{-1}\boldsymbol{\Psi}'_{12}$ $[(K-1)\times(K-1)]$ matrix and let $\boldsymbol{\omega}_{\boldsymbol{\Psi}_{22}}$ be the diagonal matrix of standard errors corresponding to matrix $\boldsymbol{\Psi}_{22}$ and $\boldsymbol{\omega}_{\boldsymbol{\Delta}}$ be another diagonal matrix of standard errors of $\boldsymbol{\Delta}$. Finally, let ϖ represent the product of the diagonal entries of the matrix $\boldsymbol{\omega}_{\boldsymbol{\Psi}_{22}}$. Using the marginal and conditional distribution properties of the multivariate normal distribution, the likelihood function in Equation (14) for the individual can be written as:

$$\begin{aligned} P\left(\boldsymbol{x}^*\right) &= |J| \times \text{Prob}[\tilde{\boldsymbol{\varsigma}}=\vec{\boldsymbol{V}}_C]\times \Pr ob[\tilde{\boldsymbol{\eta}}<\boldsymbol{S}\,|\,\tilde{\boldsymbol{\varsigma}}=\vec{\boldsymbol{V}}_C] \\ &= |J| \times \varpi^{-1}\left[\phi_M(\boldsymbol{\omega}_{\boldsymbol{\Psi}_{22}}^{-1}\vec{\boldsymbol{V}}_C,\ \boldsymbol{\omega}_{\boldsymbol{\Psi}_{22}}^{-1}\boldsymbol{\Psi}_{22}\boldsymbol{\omega}_{\boldsymbol{\Psi}_{22}}^{-1})\right]\times\Phi_{K-1}\left[\boldsymbol{\omega}_{\boldsymbol{\Delta}}^{-1}(\boldsymbol{S}-\boldsymbol{W}),\ \boldsymbol{\omega}_{\boldsymbol{\Delta}}^{-1}\boldsymbol{\Delta}\boldsymbol{\omega}_{\boldsymbol{\Delta}}^{-1}\right], \end{aligned} \tag{16}$$

where $\phi_M(\boldsymbol{a},\boldsymbol{\Sigma})$ represents the multivariate standard normal density function of dimension M with a correlation matrix $\boldsymbol{\Sigma}$ and evaluated at the abscissae value vector $\boldsymbol{a}$, and $\Phi_{K-1}(\boldsymbol{b},\boldsymbol{\Sigma})$ represents the multivariate normal standard cumulative distribution (MVNCD) function with a correlation matrix $\boldsymbol{\Sigma}$ and evaluated at vector $\boldsymbol{b}$.

The likelihood function, which is the same as the expression of Equation (16) written as a function of the parameter vector ($\left(\boldsymbol{\beta}',\boldsymbol{\theta}',\boldsymbol{\gamma}',\bar{\tilde{\boldsymbol{\Xi}}}'\right)'$) can be maximized in the usual fashion to estimate the parameters ($\boldsymbol{\gamma}$ is a column vector collecting all the satiation parameters for the inside goods, and $\bar{\tilde{\boldsymbol{\Xi}}}$ is a column vector collecting all the upper/lower triangular covariance elements of the matrix $\tilde{\boldsymbol{\Xi}}$). However, one has to compute the multivariate normal cumulative distribution (MVNCD) function of (K-1) dimensions. For this, we use the analytic approximations proposed by Bhat (2018), which have been shown to be more accurate than traditional frequentist and Bayesian simulators.

2.3. Restricted Versions of the Flexible MDCP Model

Several restricted versions of the proposed model are possible. A few restricted versions are discussed here.

2.3.1. Flexible MDCP with IID D-preferences and non-IID C-preferences

To estimate the parameters of the general covariance structure of the model proposed above, the analyst can work with a Cholesky decomposition of the error covariance matrix ($\tilde{\boldsymbol{\Xi}}$). Note that the covariance matrix of the error differences ($\boldsymbol{\Xi}$) (which is an input to the likelihood function) can be easily obtained using the following relationship while ensuring its positive definiteness:

$$\boldsymbol{\Xi} = Cov(\boldsymbol{\eta}, \boldsymbol{\varsigma}) = 0.5 \times \mathbf{1}_{2(K-1)} + \tilde{\boldsymbol{\Xi}} \tag{17}$$

Attempts to directly estimate the parameters of the covariance matrix might lead to estimation breakdowns when it is not ensured that the covariance matrix is positive definite. To work with the Cholesky decomposition, however, it is not straightforward to fix only the diagonal elements $\tilde{\lambda}_{kk}$ $(k = 2,3,....K)$ and estimate the off-diagonal elements $\tilde{\lambda}_{jk}$ $(j,k = 2,3,....K;\ j \neq k)$ of the matrix of *D*-preference error terms. This is because it is not easy to identify restrictions on the Cholesky matrix that results in such a covariance matrix. Therefore, it is convenient to assume the *D*-preference error terms to be IID (i.e., fix $\tilde{\lambda}_{kk} = 0.5\ \forall\ k = 2,3,....K$ and fix $\tilde{\lambda}_{jk} = 0\ \forall\ j,k = 2,3,....K;\ j \neq k$), while allowing for correlations between the error terms of the *D*-preference and *C*-preference terms and those among the *C*-preference error terms (i.e., allow $\sigma_{jk}\ \forall\ j,k = 2,3,....K$ and $\omega_{jk}\ \forall\ j,k = 2,3,....K$ to be non-zero). This is the approach we take in the empirical analysis later in this paper. It is important to note, however, that the above discussed restriction is imposed to avoid any estimation issues. Behaviorally, this restriction translates to an assumption of independence across goods at the discrete level of preference but allows correlations between their continuous preferences. There are practical instances where such correlation patterns are likely to exist. For example, the discrete decision to eat out may be purely need driven, and therefore may not involve significant correlation with participation in other activities, say socialization. However, the extent of time spent in eating is likely to be influenced if it is undertaken together with a social activity, which may result in positive correlations between the continuous preference functions of eating out and social activities.

2.3.2. Flexible MDCP with IID and uncorrelated D- preferences and C-preferences

If we assume the following: (a) the error terms ε_k $(k = 2,3,...,K)$ on the *D*-preference terms for all the inside goods are IID (i.e., $\tilde{\lambda}_{kk} = 0.5\ \forall\ k = 2,3,....,K$ and $\tilde{\lambda}_{jk} = 0\ \forall j,k = 2,3,....K; j \neq k$),

(b) the error terms ξ_k $(k=2,3,...,K)$ on the *C*-preference terms for all the inside goods are IID (i.e., $\omega_{kk}=0.5\mu^2$ $\forall k=2,3,...,K$ and $\omega_{jk}=0$ $\forall j,k=2,3,....,K; j\neq k$), and (c) ε_k $(k=2,3,...,K)$ and ξ_k $(k=2,3,...,K)$ are all pairwise uncorrelated (i.e., $\sigma_{jk}=0$ $\forall$ $j,k=2,3,...K$), the resulting model is IID across all *D*-preferences and *C*-preferences. With all these restrictions, $\Xi = Cov(\boldsymbol{\eta},\boldsymbol{\varsigma})$ has values of '1' in the first (K-1) diagonal entries, values of $0.5(1+\mu^2)$ in the last (K-1) diagonal entries, and all other entries take the value of 0.5. This covariance structure is similar to that in Bhat's (2018) flexible MDCEV model[7]. While one can use the likelihood function of Equation (15) with the restrictions on the covariance matrix, there is an easier way to estimate this model. Specifically, using the IID assumption across all the error terms, one can develop the likelihood of the observed consumption pattern conditional on ε_1 and then integrate over the distribution of ε_1. Doing so results in the following likelihood expression for the flexible MDCP model with IID and uncorrelated *D*- and *C*-preferences:

$$\begin{aligned}
&P\left(x_1^*, x_2^*, x_3^*, \,...,\, x_{M+1}^*, \,0,\, 0,\, ...,\, 0\right) \\
&= |J| \times \left(\frac{1}{\left(\sqrt{0.5}\right)\sigma} \right)^M \times \frac{1}{\sqrt{0.5}} \times \\
&\qquad \int_{\varepsilon_1=-\infty}^{\varepsilon_1=+\infty} \left(\prod_{k=2}^{M+1} \left[\Phi\left(\frac{-\left(\tilde{V}_k+\varepsilon_1\right)}{\sqrt{0.5}} \right) \times \phi\left(\frac{\vec{V}_k+\varepsilon_1}{\left(\sqrt{0.5}\right)\sigma} \right) \right] \times \prod_{k=M+2}^{K} \left[\Phi\left(\frac{\tilde{V}_k+\varepsilon_1}{\sqrt{0.5}} \right) \right] \right) \phi\left(\frac{\varepsilon_1}{\sqrt{0.5}} \right) d\varepsilon_1
\end{aligned} \tag{18}$$

Note that the scale σ in the above expression is equivalent to $0.5(1+\mu^2)$. The above likelihood function entails only a single dimensional integral, regardless of the number of alternatives, and considerably simplifies the estimation.

---

[7] A small nuance is in order here. Bhat's (2018) flexible MDCEV uses the same variance for $\tilde{\lambda}_{kk}$ and ω_{kk}, which leads to a closed-form likelihood expression. However, other than the distribution form, the IID *D-C* MDCP model is similar to the MDCEV in Bhat's paper, because the scale does not matter in the discrete part of the IID *D-C* MDCP, and we can as well normalize $\tilde{\lambda}_{kk}$ to $0.5\mu^2$ (*k*=1,2,…*K*). However, just to be consistent with the model development in Section 2.1, we will keep to the assumption that $\tilde{\lambda}_{kk}=0.5$, because it more clearly shows that the IID *D-C* MDCP model is a restricted version of the full flexible MDCP model proposed in Section 2.1.

3. EMPIRICAL APPLICATION

3.1. Sample Data

To demonstrate an application of the proposed flexible MDCP model, we consider the case of tourism expenditures of domestic tourists in India. The empirical data for this analysis comes from the domestic tourism expenditure survey carried out by the National Sample Survey Office (NSSO) under the Ministry of Statistics and Program Implementation (MOSPI) of the Government of India (NSS 72nd round survey carried out from July 2014 to June 2015). The data contains information on domestic trips (undertaken by Indian households) that involved at least one overnight stay away from the households' usual place of residence (UPR) for any of the following purposes: leisure and recreation, shopping, or health and medicine. For each surveyed household, all such trips made by the household members over a period of 365 days before the survey day were recorded. For each domestic trip, the survey recorded expenditures across six expenditure classes: (a) transportation, (b) accommodation, (c) food and beverages, (d) shopping, (e) recreation and leisure, and (f) health and medicine. Among these trips, the most recent trip of the household with the primary purpose of leisure and recreation was considered. Only those trips that were not reimbursed by an employer or other sources were considered; package-deal trips were removed from consideration since these trips did not contain information on expenditure by category. The final sample had information on the expenditure patterns of 4981 households, of which 3500 were used in model estimation, and the other 1481 were kept aside as a holdout sample.

Table 1 reports the aggregate expenditure patterns of the above-described sample of trips. Note that all trips involve some expenditure on transportation, and therefore, in this analysis, we consider transportation expenditures as part of the essential Hicksian outside good. Since the model framework employs a linear-utility structure on the outside good (Bhat, 2018), the information on the total budget and the allocation to the outside good does not become a part of the analysis (and hence, its expenditure details are not provided in Table 1). As discussed in Bhat (2018) and Saxena *et al.* (2021), the linear outside good utility form is helpful for analyzing empirical situations when the total budget is not known or clearly defined but can be assumed to be large compared to the allocations to inside goods. In the current empirical context, since the allocation to the expenditure classes of interest is likely to be small relative to the total budget (which can be, for instance, the total annual household expenditure) of the household, a linear outside good utility profile is suitable. However, if the information on the budget is available, one can use the information for forecasting with the model.

In addition to the information on household expenditures across different categories during their trip, the sample has information on socio-demographic factors, such as household location type (urban or rural), household income level, and composition of the group of members making the trip (gender and age composition). In addition, trip level information such as travel mode and duration of the trip was collected. All these variables were explored as exogenous variables in the empirical model. It is worth noting here that this empirical data has been used earlier by Saxena *et al.* (2021) to analyze households' tourism expenditures using MDC choice models that do not employ the flexible utility form nor the non-IID error distributions. The focus of the current study, however, is to assess the benefits of the flexible utility form along with the non-IID error structure.

3.2. Estimation Results

The empirical specification was carefully built by systematically adding exogenous variables and dropping the statistically insignificant parameters from the specification. In the final specification, parameters that provided an intuitive interpretation were retained if their t-statistic value was greater than 1. However, constants in the baseline preferences (both *D*- and *C*-preferences) and the satiation parameters were retained in the final specification irrespective of their statistical significance.

The estimation results of the proposed flexible MDCP model are reported in Table 2. The parameter estimates under the heading *"Discrete preferences"* are for the effects of exogenous variables on the *D*-preference baseline utility components. The parameter estimates under the column heading "*Continuous preferences*" report the effects of exogenous variables on the *C*-preference baseline utility components. Finally, the parameter estimates in the satiation functions are reported in the set of rows labeled "*Satiation function*". These results are discussed below.

3.2.1. Effects of exogenous variables on the D-preference baseline utilities

In the context of household characteristics, the negative coefficient of the "urban household" dummy variable in the *D*-preference parameter for shopping suggests that households residing in urban areas are less likely (than those in rural areas) to spend on shopping during their travel outside their usual place of residence (UPR). This may be because urban households tend to have better access to shopping facilities at their UPR than rural households (Venugopal, 2012; Mulky, 2013). Urban households, on the other hand, are more likely to spend on accommodation as compared to their rural counterparts.

Next, income effects were explored using a surrogate variable associated with the household's usual monthly consumer expenditure (UMCE) by categorizing the households into low-income (UMCE < INR 10,000), medium-income (INR 10,000 to 20,000), and high-income (UMCE > INR 20,000) classes. The corresponding parameter estimates in the *D*-preferences suggest that low- and medium-income households are less likely (than high-income households) to spend in most of the expenditure categories considered in this study.

In the context of the variables describing the composition of the travel group, as the group size increases, people are less likely to spend on accommodation, as implied by the negative sign on its coefficient in the corresponding *D*-preference function for accommodation. This may be because large travel groups prefer to stay with friends or relatives to save on accommodation costs. In addition to the size of the group, the effects of group composition were considered through three variables – the proportion of women in the group, the proportion of elderly (> 60 years of age), and the proportion of children in the group. While the proportion of children in the group did not show a significant influence on the expenditure patterns of travelers, the proportion of women in the group had an interesting influence. Travel groups with a higher proportion of women are less likely to opt for paid accommodation, possibly because such groups may prefer staying with friends or relatives for safety reasons. Such groups are also less likely to spend on food and beverages since they might prefer hygienic, home-cooked meals than eating out at unfamiliar locations. On the other hand, groups constituting more women are more likely to spend on shopping, possibly because women derive greater satisfaction from shopping than men (Herter *et al.,* 2014). Interestingly, female travelers are associated with greater spending than male travelers (Brida and Scuderi, 2013; Wang and Davidson, 2010). However, our finding indicates such spending patterns are not consistent across different expenditure categories. Another intuitive finding was that travel groups with a higher proportion of elderly individuals are more likely to spend in the health and medicine category.

Next, two trip specific variables were considered in the model: (a) duration of the trip (*i.e.,* the number of nights spent outside the UPR) and (b) the type of trip destination (*i.e.,* within the same district as the UPR, outside the district of the UPR but within the same state, and outside the state of the UPR). Trip duration shows a positive association with the likelihood of spending on health and medicine. This result is unlikely due to a causal effect of the trip duration on health and medical needs. Instead, this result is likely due to an endogenous effect where trips involving health and medical activities tend to be longer than other trips. Interestingly, the trip duration did not affect the discrete choice to spend in other expense categories, except accommodation. The effect of trip duration on a household's likelihood of spending on accommodation was captured through a non-linear (category variable) specification. Specifically, travelers are more inclined to opt for paid accommodation for short- and medium-duration trips (i.e., trips of a duration of 10 days or shorter) than trips longer than ten days. This is not surprising, since shorter duration trips are associated with higher daily expenditures (Yang et al., 2021). Also, on long-duration trips, travelers are likely to stay with friends or relatives to save on high accommodation costs (see Pellegrini *et al.,* 2014 for a similar finding). Finally, the trip destination had an expected impact on the decision of spending in all expenditure classes. Specifically, trips farther from the household's UPR (i.e., those with a destination outside the district of the household's UPR or outside the state of the UPR) are associated with a greater likelihood of expenditure in all expenditure categories as compared to trips to a destination within the same district of the UPR.

3.2.2. Effects of exogenous variables on the C-preference and satiation parameters

Note that both the *C*-preference parameters and the satiation parameters influence the extent of expenditure allocation to each of the inside alternatives. This is a reason why the satiation parameters are estimated with a lower statistical significance as compared to those in the traditional MDC models. For the same reason, not many covariates have a significant influence on the satiation functions (as they already influence the *C*-preference functions). In this context, although the results of a traditional MDCP model are not reported in the paper, the satiation parameter estimates of such a model had higher t-statistic values than those in the flexible MDCP model reported in Table 2.

In the context of *C*-preferences, households from urban areas show higher expenditures on accommodation and shopping than those from rural areas. Recall from the earlier discussion that households from urban areas were found to be less likely to spend on shopping. However, if they spend on shopping or accommodation, urban households are likely to spend more than rural households. The flexible MDCP model, due to separating the influence of covariates on discrete and continuous choices, makes it easier than the traditional MDC choice models to estimate such opposite effects on discrete and continuous choices.

The negative coefficients on the income dummy variables in the *C*-preference specification suggest that low- and medium-income households tend to spend less than high-income households on their recreational and leisure trips. This is presumably because high-income households tend to possess a greater spending capability owing to a larger disposable income than lower-income households. These findings are consistent with the findings from other countries. For instance, a study by Asgary et al. (1997) revealed that high incomes in Mexican travelers boosted their tourism expenditures. In a similar study in Spain, Nicolau and Mas (2005) reported that higher income were associated with greater tourism expenditures.

Next, recall from the discussion on *D*-preferences that the travel group size variable was found to be negatively associated with the likelihood of spending in the accommodation category. In the context of *C*-preferences, however, this variable has a positive coefficient in the shopping and recreation categories. Further, this variable has a positive coefficient in the satiation function for accommodation and food & beverage categories. These results suggest that larger travel groups are likely to spend more on accommodation, food & beverages, shopping, and recreation categories – if they choose to spend in those categories.

Variables describing the composition of the travel group (such as the proportion of women and the proportion of elderly) do not show statistically significant effects in *C*-preferences. However, a larger proportion of women in the travel group is associated with a positive effect on the satiation function for the accommodation category. This result implies that travel groups with more women are likely to spend more on accommodation than those with fewer women. This again supports our conjecture that travel groups with more women tend to prioritize safety and therefore prefer accommodations that are considered safer (and are possibly more expensive) (see Zemke *et al.,* 2015 for similar findings). Finally, as can be observed from the positive coefficient in the corresponding satiation function, travel groups with a larger proportion of elderly show a greater extent of expenditures in the health and medicine category.

3.2.3. Structure on the variance-covariance matrix

In theory, the covariance structure as implied in Equation (9) is identifiable with restrictions on the covariance elements specific to the *D*-preference utilities. As discussed in Section 2.3.1, we estimated the flexible MDCP model with IID *D*-preferences and non-IID *C*-preferences, where

$Var(\varepsilon_k) = \tilde{\lambda}_{kk} = 0.5\ (k = 1, 2, ..., K)$ and $\lambda_{jk} = 0\ \forall (j, k = 2, ..., K\ \&\ j \neq k)$. All other covariance elements were freely estimated using a Cholesky decomposition of the covariance matrix of error terms (i.e., $\tilde{\boldsymbol{\Xi}}$). Since we estimated the Cholesky parameters that correspond to the covariance matrix of the error terms (as opposed to the covariance matrix of error differences), the elements of the resulting covariance matrix are interpretable. The resulting covariance elements, after appropriate conversion from the estimated parameters in the Cholesky matrix, are reported in Table 3.

A few points are worth discussing regarding the error covariance matrix reported in Table 3. First, the model estimated with a general covariance structure (*i.e.,* with IID *D*-preferences and non-IID *C*-preferences) provided a significantly better fit than models with restricted covariance structures (more on this later). Second, all covariance elements corresponding to the *C*-preference utility parameters (*i.e.,* $\omega_{jk}\ \forall\ j, k = 2, 3, ...K\ \&\ j \neq k$) were insignificant across all expenditure classes, except between accommodation and food & beverage categories, where the correlation between the two categories is positive. This highlights that those travelers who prefer staying at expensive accommodations are also likely to patronize expensive eat-out options. Third, most of the covariances between the *D*-preference of an expenditure class *j* and *C*-preference of an expenditure class *k* ($\sigma_{jk}\ \forall\ j, k = 2, 3, ...K\ \&\ j \neq k$) were close to zero. While it may be useful to fix such covariances to zero, as discussed in section 2.3.1, it is not easy to do so with Cholesky decomposition. Therefore, we retained them in the empirical model. Fourth, the covariances between the *D*- and *C*-preferences of an expenditure class (*i.e.,* σ_{kk}) were positive across all expenditure classes, except for the food and beverage category. The positive covariances are intuitive, as one would expect a positive correlation between the *D*- and *C*-preferences of an expenditure class. The negative covariance between the *D*- and *C*-preferences of the food and beverages category is only marginally significant.

3.2.4. Likelihood-based goodness-of-fit measures

Table 4 reports the likelihood-based data fit measures for the following models estimated in this study: (a) flexible MDCP model with IID *D*-preferences and non-IID *C*-preferences (as discussed in Section 2.3.1), (b) flexible MDCP model with IID and uncorrelated *D*-preferences and *C*-preferences (i.e., IID uncorrelated *D*-*C* MDCP model), (c) a traditional MDCP model with a full covariance matrix (or non-IID MDCP model), (d) a traditional MDCP model with IID error terms in the stochastic specification (or IID MDCP model), and (e) a flexible MDCEV model as in Bhat (2018) with IID type I extreme value error terms. The traditional MDCP model was estimated to assess the benefits of the additional *D*-preference parameters in the flexible MDCP model. The flexible MDCEV model was estimated for comparison with its IID probit counterpart (i.e., the flexible MDCP with IID and uncorrelated *D*-preferences and *C*-preferences) to assess which error distribution provides a better fit in the current empirical context.

The likelihood-based goodness-of-fit measures for all the above models were computed on an estimation sample of 3,500 households and a holdout sample of 1,481 households. These measures are reported in Table 4. Note that the three models with the label "flexible" in Table 4 relax the tight tie between the discrete and continuous consumption decisions, but are not always consistent with utility-maximizing behavior. On the other hand, the two models with the label "traditional" tie the discrete and continuous consumptions very tightly, but are strictly consistent with utility-maximizing behavior. Several observations can be observed from the table. First, the flexible MDCP model with IID *D*-preferences and non-IID *C*-preferences (see second column of Table 4) provides a superior goodness-of-fit than all other models estimated in the study. This result highlights the benefits of relaxing the IID assumptions in flexible MDCP models. Second, note that both the IID uncorrelated *D-C* MDCP (third column of Table 4) and the flexible MDCEV model from Bhat (2018) (last column of Table 4) impose a similar structure on the covariance matrix. However, the probit version of the flexible MDC model provides a better fit in the current empirical context. Further, the advantage of the flexible MDCP model structure lies in its ability to easily relax the IID structure imposed in the flexible MDCEV model. Third, between the flexible model with an IID MDCP structure (in the third column) and the traditional IID MDCP (in the fifth column), the former provides a better fit in both the estimation sample and the validation sample (albeit, in the validation sample, the BIC value for the IID *D-C* MDCP model was higher than that for the IID MDCP model, possibly because BIC over-penalizes the flexible MDCP model due to more number of parameters). Similarly, between the flexible MDCP with a non-IID structure (second column) and the corresponding traditional MDCP with a non-IID structure (fourth column), once again, the flexible structure provides a better fit both in estimation and holdout samples. Finally, it is interesting to focus on the comparison of the IID flexible MDCP (second column) and the non-IID but traditional MDCP (fourth column). Here, the non-IID traditional MDCP comes out to be the clear winner.

3.2.5. Comparison of predictive accuracy

The likelihood-based goodness-of-fit measures indicate better performance of the flexible MDCP model with IID D-preferences and non-IID C-preferences than other models evaluated in this study. However, better performance based on likelihood based fit measures does not always translate to improved accuracy in predictions on outside samples. Therefore, we evaluate the predictive performance of these models on the holdout sample (i.e., the sample of 1481 data points not used for estimation). Specifically, we compare the predictions from the proposed non-IID flexible MDCP model with those obtained from (a) the IID D-C flexible MDCP model – to highlight the importance of relaxing the IID assumption in the prior formulation of flexible MDC model, and (b) the non-IID traditional MDC model – to highlight the importance of separating the discrete and continuous preferences in MDC model systems. The results of the predictive assessments are presented in Table 5, where the predictions from the respective models are aggregated over 100 sets of simulation draws.

As is evident from the table, all three models do a decent job in predicting the discrete shares of whether to spend in an expense category, with the IID D-C flexible MDCP model predicting closest to the observed shares, as indicated by the weighted Mean Absolute Percentage Error (MAPE) values corresponding to the discrete shares (see the fourth, sixth and eight columns in Table 5). However, when it comes to the extent of expenditure in these categories, the flexible MDCP model with IID D- and non-IID C-preferences provides more accurate predictions, with the weighted MAPE value of around 16% (as compared to the weighted MAPE of 27.8% corresponding to the IID D-C flexible MDCP model, and 17% for the non-IID traditional MDCP model. At the level of expenditure class, it is evident from the results that all the models predict fairly well across all categories, except accommodation. Interestingly, the non-IID traditional MDCP model is quite off from the observed expenditure in accommodation. A possible reason for this error could be the fact that accommodation category has a relatively lower discrete share but a significantly high expenditure. Since the same baseline preference parameter controls both the discrete as well as continuous outcomes in the traditional MDCP model, the errors are reflected in the continuous predictions. However, in all other categories, both the flexible MDCP model (with IID D- and non-IID C-preferences) and the traditional MDCP model predict with fair accuracy. This is also reflected in the comparable overall weighted MAPE values for the two models, with the flexible model structure predicting slightly better than the traditional model structure.

Overall, in our empirical context, the predictive assessment indicates similar trends as those observed form the likelihood-based goodness-of-fit measures, where the flexible (but not always consistent with utility-maximizing-behavior) model structure is the preferred structure relative to the traditional MDC model that is strictly consistent with utility-maximizing behavior. We expect this to be the case in general too, because of the relaxing of the tight tie between the discrete and continuous consumptions. The caveat though is that the IID flexible structure need not always be better than the non-IID traditional structure, as is also observed from our result presented above. The takeaway is that if strictly utility-maximizing behavior is not necessarily desired, the preferred model would, in general, be the flexible non-IID MDCP model proposed and implemented in this paper. However, a potential downside to the flexible MDC model structure is that it cannot be used to calculate welfare measures. In such situations, where consistency with the utility-maximizing behavior is important, the traditional non-IID MDCP with a full covariance matrix would be a good model to consider. This traditional non-IID MDCP will, in general, provide a much superior fit to IID (MDCP or MDCEV) traditional models and, as in our empirical case, may even provide a better data fit and accurate predictions than the flexible IID MDCP model.

4. CONCLUSIONS

In this paper, we formulate a flexible multiple discrete-continuous probit (MDCP) model with a multivariate normal distributed error kernel that allows non-IID error structure in the utility functions. In doing so, we revisit the formulation of Bhat's (2018) flexible utility form and highlight that the discrete preference parameters used in the flexible utility form do not have a role in the optimality conditions associated with the utility function. As a result, the stochastic conditions used by Bhat (2018) to derive the model likelihood function are not always consistent with utility maximization (hence the flexible MDCEV model is not necessarily consistent with utility maximization). Therefore, we provide an alternate interpretation of Bhat's (2018) model as a representation of a two-step decision-making heuristic. Specifically, for a given inside good k, in the first step, the discrete choice decision is made based on its discrete choice preference function. In the second step, the continuous consumption amount is determined based on the continuous choice preference function and the satiation function of that good.

The proposed flexible MDCP model is applied to analyze monetary expenditures incurred on recreational trips undertaken by Indian households that involved staying at a destination outside the household's usual place of residence (UPR) for at least one night. The sample data comes from a domestic tourism survey by the National Sample Survey Office of the Ministry of Statistics and Program Implementation (Domestic Tourism survey, NSS 72nd Round survey conducted from July 2014 to June 15). The expenditures incurred on these trips in five expenditure classes – namely (a) accommodation, (b) food and beverages, (c) shopping, (d) recreation and leisure, and (e) health and medicine – were analyzed. The empirical model sheds light on the determinants of Indian households' expenditure patterns on their recreational/tourism trips.

Likelihood-based goodness-of-fit measures of the proposed flexible MDCP model were compared with a simpler version of it (with IID error terms) and the traditional MDCP, both in the estimation and holdout samples. For completeness, Bhat's (2018) flexible MDCEV model was also estimated. Our results indicate that, if the analyst is willing to compromise on the strict utility-maximizing aspect of behavior, while also enriching the behavioral dimension through the relaxation of the tie between the discrete and continuous consumption decisions, the preferred model would be the flexible non-IID MDCP model. On the other hand, if the analyst wants the model to be strictly grounded on utility-maximizing behavior (which may also have benefits by way of welfare measure computations), but is willing to assume a tight (and oftentimes difficult to justify) tie between the discrete and continuous consumption decision processes, the preferred model would be the non-IID traditional MDCP model.

Future research should consider flexible utility formulations that are also consistent with utility maximization. One approach to do so would be to truncate the space of the *D*-preference and *C*-preference parameters such that the model becomes utility optimal for all possible values of the truncated parameter space, a direction of research that the authors are currently pursuing. Equally useful would be to derive a method to estimate welfare measures from such models.

Acknowledgments

This research was undertaken as part of a SPARC project funded by the Indian Ministry of Education for encouraging international collaborations. Empirical data used in this study were collected by the National Sample Survey Office of the Indian Ministry of Statistics and Program Implementation. The authors are grateful for the support provided by both ministries. The second author would like to thank Tarun Rambha for helpful discussions on non-linear optimization.

Conflict of interest statement

On behalf of all authors, the corresponding author states that there is no known conflict of interest.

Author Contributions

Author contributions to the paper are as follows: study conception and design: CB, SS, AP; analysis and interpretation of results: SS, AP, CB, AM; draft manuscript preparation: SS, AP, CB, AM. All authors reviewed the results and approved the final manuscript.

References

Asgary, N., De Los Santos, G., Vincent, V. and Davila, V. (1997). The determinants of expenditures by Mexican visitors to the border cities of Texas. *Tourism Economics*, 3(4), 319-328.

Bhat, C.R. (2005). A Multiple Discrete-Continuous Extreme Value Model: Formulation and Application to Discretionary Time-Use Decisions. *Transportation Research Part B: Methodological* 39: 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: Methodological*, 42(3), 274–303.

Bhat, C.R. (2018). New Matrix-Based Methods for the Analytic Evaluation of the Multivariate Cumulative Normal Distribution Function. *Transportation Research Part B: Methodological,* 109: 238–256.

Bhat, C. R. (2018). A New Flexible Multiple Discrete–Continuous Extreme Value (MDCEV) Choice Model. *Transportation Research Part B: Methodological*, 2018, 110, 261–279.

Brida, J.G. and Scuderi, R., (2013). Determinants of tourist expenditure: A review of microeconometric models. *Tourism Management Perspectives*, *6*, 28-40.

Calastri, C., Hess, S., Daly, A., Maness, M., Kowald, M., and Axhausen, K. (2017). Modelling

Contact Mode and Frequency of Interactions with Social Network Members Using the Multiple Discrete-Continuous Extreme Value Model. *Transportation Research Part C*, 76, 16–34.

Enam, A., Konduri, K.C., Pinjari, A.R., & Eluru, N. (2018). An Integrated Choice and Latent Variable Model for Multiple Discrete-continuous Choice Kernels: Application Exploring the Association Between Day Level Moods and Discretionary Activity Engagement Choices. *Journal of Choice Modelling*, 26, 80–100.

Hanemann, W.M. (1984). Discrete/Continuous Models of Consumer Demand. *Econometrica*, 52(3) 541–561.

Herter, M.M., Pizzutti, C., & Pinto, D.C. (2014). 'Man, I Shop like a Woman!' The Effects of Gender and Emotions on Consumer Shopping Behaviour Outcomes. *International Journal of Retail & Distribution Management*, 42(9), 780–804.

Hendel, I. (1999). Estimating multiple-discrete choice models: An application to Computerization Returns. *Review of Economic Studies*, 66(2), 423–446.

Jian, S., Rashidi, T.H., & Dixit, V. (2017). An Analysis of Carsharing Vehicle Choice and Utilization Patterns Using Multiple Discrete-Continuous Extreme Value (MDCEV) Models. *Transportation Research Part A: Policy and Practice,* 103, 362–376.

Kim, J., Allenby, G.M., & Rossi, P.E. (2002). Modeling Consumer Demand for Variety. *Marketing Science*, 21(3), 229–250.

Khan, M. & Machemehl, R. (2017) Analyzing Tour Chaining Patterns of Urban Commercial Vehicles. *Transportation Research Part A: Policy and Practice*, 102, 84–97.

Mulky, A.G. (2013). Distribution Challenges and Workable Solutions. *IIMB Management Review,* 25 (3): 179–195.

National Sample Survey Office (NSSO). (2016). Domestic tourism expenditure, NSS 72nd round (July 2014-June 2015). *Ministry of Statistics and Programme Implementations (MOSPI), Government of India*. http://microdata.gov.in/nada43/index.php/catalog/138/. Accessed in June 2020.

Nicolau, J.L. and Mas, F.J. (2005). Stochastic modeling: a three-stage tourist choice process. *Annals of Tourism Research*, *32*(1), pp.49-69.

Pellegrini, A., Sarman, I., & Maggi, R. (2021). Understanding Tourists' Expenditure Patterns: A Stochastic Frontier Approach within the Framework of Multiple Discrete–Continuous Choices. *Transportation*, 48, 931–951.

Pinjari, A.R. (2011). Generalized Extreme Value (GEV)-Based Error Structures for Multiple Discrete-Continuous Choice Models. *Transportation Research Part B: Methodological,* 45(3): 474–489.

Pinjari, A.R. & Bhat, C.R. (2009). A Multiple Discrete-Continuous Nested Extreme Value (MDCNEV) Model: Formulation and Application to Non-Worker Activity Time-Use and Timing Behavior on Weekdays. *Transportation Research Part B: Methodological*, 44(4): 562–583.

Pollak, R. A. & Wales, T.J. (1992). Price-Augmenting Returns to Scale: An Application to Nonseparable Two-Stage Technologies. *The Review of Economics and Statistics*, 74(2), 213–220.

Saxena, S., Pinjari, A.R. & 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 Resarch Part B*, 155, 526-557.

Sobhani, A., Eluru, N., & Faghih-Imani, A. (2013). A Latent Segmentation Based Multiple Discrete Continuous Extreme Value Model. *Transportation Research Part B: Methodological*, 58, 154–169.

Venugopal, P. (2012). Urban Orientation of Rural Consumers: Implication for Consumer Goods Distribution. *International Journal of Rural Management*, 8 (1–2), 107–119.

von Haefen, R.H., Phaneuf, D.J. (2003). Estimating Preferences for Outdoor Recreation: A Comparison of Continuous and Count Data Demand System Frameworks. *Journal of Environmental Economics and Management*, 2003, 45(3) 612–630.

von Haefen, R.H., & Phaneuf, D.J. (2005). Kuhn-Tucker Demand System Approaches to Non-Market Valuation. In: Scarpa, R. & Alberini, A. (eds): Applications of Simulation Methods in Environmental and Resource Economics. The Economics of Non-Market Goods and Resources, 6, Spronger Dordrecht, 135–157.

Wales, T. J., and Woodland, A.D. (1983). Estimation of Consumer Demand Systems with Binding Non-Negativity Constraints. *Journal of Econometrics*, 21(3), 263–285.

Wang, Y. and Davidson, M.C., (2010). A review of micro-analyses of tourist expenditure. *Current issues in Tourism*, *13*(6), 507-524.

Wafa, Z., Bhat, C.R., Pendyala, R.M., & Garikapati, V.M. (2015). Latent-Segmentation-Based Approach to Investigating Spatial Transferability of Activity-Travel Models. *Transportation Research Record*, 2493 (1), 136–144.

Yang, C.W., Wu, C.L. and Lu, J.L. (2021). Exploring the interdependency and determinants of tourism participation, expenditure, and duration: An analysis of Taiwanese citizens traveling abroad. *Tourism Economics*, *27*(4), 649-669.

Yonezawa, K., & Richards, T. (2017). Consumer Risk-Reduction Behavior and New Product Purchases. *Managerial and Decision Economics*,38, 1003–1016.

Table 1. Summary of the sample data (Sample size = 4,981 households)

Expenditure category	**Percentage of households spending in the category**	**Details of expenditure in the row category**	
		Avergae expenditure (across household who spend in the category) (Indian Rupess)	**Standard deviation of the expenditure** (across household who spend in the category) (Indian Rupess)
Transportation	100.0	3,252	7,078
Accommodation	46.0	2,983	3,395
Food and beverages	93.0	1,508	2,401
Shopping	85.3	2,626	3,387
Recreation and leisure	35.4	613	594
Health and medicine	16.6	485	1,827

Table 2. Estimation results of the flexible MDCP model with full covariance matrix

	Discrete preferences					Continuous preferences				
	Accommo-dation	Food & beverages	Shopping	Recreation & leisure	Health & medicine	Accommo dation	Food & beverages	Shopping	Recreation & leisure	Health & medicine
Baseline preference function										
Constants	-1.35 *(-10.7)*	1.42 *(8.9)*	0.63 *(7.4)*	-0.98 *(-13.0)*	-1.37 *(-18.3)*	1.61 *(5.6)*	3.46 *(11.0)*	3.89 *(5.7)*	2.78 *(2.8)*	3.00 *(0.5)*
Household specific variables										
Urban household (Base category: Rural)	0.08 *(1.7)*	-	-0.22 *(-4.1)*	-	-	0.23 *(4.3)*	0.17 *(4.0)*	-	-	-
Income (Base category: UMCE > ₹20K)										
Low-income (UMCE < ₹10K)	-0.68 *(-10.9)*	-0.43 *(-3.1)*	-0.07 *(-1.4)*	-0.18 *(-4.2)*	-	-0.89 *(-12.6)*	-0.86 *(-14.4)*	-0.51 *(-7.1)*	-0.57 *(-7.2)*	-0.12 *(-1.1)*
Medium-income (UMCE: ₹10K-₹20K)	-0.32 *(-5.8)*	-0.24 *(-1.7)*	-	-	-	-0.51 *(-8.9)*	-0.47 *(-8.6)*	-0.25 *(-3.6)*	-0.18 *(-2.5)*	-
Travel group and trip specific variables										
Size of the travel group	-0.04 *(-2.5)*	-	-	-	-	-	-	0.14 *(8.4)*	0.11 *(5.2)*	-
Proportion of women in the group	-0.22 *(-3.4)*	-0.50 *(-5.4)*	0.19 *(2.7)*			-	-	-	-	-
Proportion of elderly in the group	-	-	-	-	0.42 *(7.0)*	-	-	-	-	-
Number of nights	-	-	-	-	0.02 (9.7)	-	0.02 *(11.0)*	0.02 *(8.9)*	0.01 *(2.2)*	0.04 *(7.6)*
Trip duration (Base: duration > 10 nights)										
Trip duration is 1-3 nights	1.07 *(13.3)*	*	*	*	*	-0.31 *(-3.3)*	*	*	*	*
Trip duration is 4-10 nights	0.57 *(7.2)*	*	*	*	*	-0.15 *(-2.0)*	*	*	*	*
Trip destination (Base: Same as UPR)										
Same state of UPR (not same district)	0.81 *(10.2)*	0.57 *(6.9)*	0.37 *(5.0)*	0.64 *(8.1)*	0.28 *(3.4)*	0.74 *(6,1)*	0.77 *(11.0)*	0.51 *(6.0)*	0.64 *(4.8)*	0.07 *(1.0)*
Outside the state of UPR	1.43 *(17.5)*	1.09 *(10.5)*	0.68 *(8.5)*	1.0 *(13.1)*	0.43 *(5.4)*	1.45 *(10.5)*	1.50 *(19.7)*	1.26 *(14.6)*	1.23 *(9.2)*	0.57 *(3.1)*
Satiation function										
Constants						0.0 *(0.0)*	-2.53 *(-8.1)*	-2.94 *(-4.3)*	-3.24 *(-3.2)*	-5.6 *(-1.0)*
Size of the travel group						0.07 *(4.3)*	0.12 *(9.4)*	-	-	-
Proportion of women in the group						0.21 *(2.8)*	-	-	-	-
Proportion of elderly in the group						-	-	-	-	1.26 *(6.3)*
Goodness of fit measures										
Number of cases	3500									
Number of parameters	102									
Log-likelihood for constant only model	-43,106.10									
Log-likelihood of the final specification	-40,161.80									
Akaike information criterion (AIC)	80,527.60									
Bayesian information criterion (BIC)	81,155.97									

-: Coefficient was dropped from the specification as it was not statistically significant.
*: Number of nights was specified as a categorical variable in the accommodation utility function and continuous variable in other utility functions.
t-statistics are reported in parentheses next to the parameter estimates.

Table 3. Covariance matrix of error terms in the model

		Discrete preferences					Continuous preferences				
		Acc.	F&B	Shop	R&L	H&M	Acc.	F&B	Shop	R&L	H&M
Discrete preferences	Acc.	0.5 *(fixed)*									
	F&B	0 *(fixed)*	0.5 *(fixed)*								
	Shop	0 *(fixed)*	0 *(fixed)*	0.5 *(fixed)*							
	R&L	0 *(fixed)*	0 *(fixed)*	0 *(fixed)*	0.5 *(fixed)*						
	H&M	0 *(fixed)*	0 *(fixed)*	0 *(fixed)*	0 *(fixed)*	0.5 *(fixed)*					
Continuous preferences	Acc.	0.191 *(3.35)*	-0.266 *(-2.94)*	-0.203 (-6.50)	-0.125 *(-5.48)*	-0.205 *(-7.76)*	0.621 *(25.48)*				
	F&B	0.320 *(11.49)*	-0.248 *(-1.10)*	-0.149 (-5.80)	0.000[#] *(0.00)*	-0.110 *(-4.54)*	0.451 *(2.76)*	0.822 *(29.61)*			
	Shop	-0.077 (-6.27)	0.000[#] *(0.00)*	0.090[#] (0.54)	0.000[#] *(0.00)*	-0.138 *(-3.61)*	0.142[#] *(0.13)*	0.200[#] *(0.51)*	1.39 *(19.99)*		
	R&L	0.000[#] *(0.00)*	0.000[#] *(0.00)*	-0.44 *(-8.40)*	0.208 *(2.15)*	-0.176 *(-4.04)*	0.126[#] *(0.09)*	0.262 *(0.29)*	0.289[#] *(0.334)*	1.023 *(14.99)*	
	H&M	0.152 (2.10)	0.000[#] *(0.00)*	0.37 (3.92)	0.000[#] *(0.00)*	0.904 *(5.60)*	0.039[#] *(0.002)*	0.208[#] *(0.09)*	0.055[#] *(0.01)*	0.226[#] *(0.11)*	2.705 *(9.31)*

[#] t-statistic < 1

Table 4. Goodness-of-fit measures in estimation and holdout samples

Goodness of fit measures in the estimation sample (N = 3,500)					
	Flexible MDCP (with IID *D*-preferences and non-IID *C*-preferences)	Flexible MDCP# (IID *D-C* MDCP)	Traditional MDCP (non-IID covariance structure)	Traditional MDCP (IID MDCP)	Flexible MDCEV#
Log-likelihood at convergence	-40,161.80	-41,071.45	-40,432.60	-41,198.50	-41,431.60
Number of parameters	102	75	71	58	75
Akaike information criterion (AIC)	80,527.60	82,292.90	81,007.20	82,513.00	83,013.20
Bayesian information criterion (BIC)	81,155.97	82,754.94	81,444.60	82,870.31	83,475.24
$ Likelihood ratio test	Test statistic: $-2(LL_{IID\ D-C\ Flex\ MDCP} - LL_{Flex\ MDCP\ (IID\ D\text{-}\ and\ non\ IID\ C\text{-}preferences)}) = 1819.30$ > Chi-squared statistic for 27 degrees of freedom at any reasonable degree of freedom, thus implying that the flexible MDCP model (with IID *D*-preferences and non-IID *C*-preferences is the preferred model.				
Goodness of fit measures in the holdout sample (N = 1,481)					
	Flexible MDCP (with IID *D*-preferences and non-IID *C*-preferences)	Flexible MDCP# (IID *D-C* MDCP)	Traditional MDCP (non-IID covariance structure)	Traditional MDCP (IID MDCP)	Flexible MDCEV#
Predictive log-likelihood	-17,404.12	-17,812.54	-17,520.77	-17,839.42	-17,884.93
Akaike information criterion (AIC)	35,012.24	35,775.08	35,163.54	35,794.84	35,919.86
Bayesian information criterion (BIC)	35,552.89	36,172.61	35,620.94	36,102.27	36,381.90
$ Likelihood ratio test	Test statistic: $-2(LL_{IID\ D-C\ Flex\ MDCP} - LL_{Flex\ MDCP\ (IID\ D\text{-}\ and\ non\ IID\ C\text{-}preferences)}) = 816.84$ > Chi-squared statistic for 27 degrees of freedom at any reasonable degree of freedom, thus implying that the flexible MDCP model (with IID *D*-preferences and non-IID *C*-preferences is the preferred model.				

$Likelihood ratio test was performed between the Flexible MDCP model (with IID *D*-preferences and non-IID *C*-preferences) and the IID D-C MDCP version of the Flexible MDCP model.
#Both Flexible MDCP and Flexible MDCEV models impose similar structure on the covariance matrix.

Table 5. Weighted MAPE in predictions from (a) Flexible MDCP model with IID D- and non-IID C-preferences, (b) IID *D-C* Flexible MDCP model, and (c) non-IID traditional MDCP model

Expenditure classes	Observed patterns in Holdout sample (N =1481)		Predictions from flexible MDCP (with IID *D*-preferences and non-IID *C*-preferences)		Predictions from flexible MDCP (with IID *D*- and IID *C*-preferences)		Predictions from traditional MDCP (with non-IID covariance structure)	
	Discrete shares	Aggregate Expenses (100s of ₹)	Discrete shares	Aggregate Expenses (100s of ₹)	Discrete shares	Aggregate Expenses (100s of ₹)	Discrete shares	Aggregate Expenses (100s of ₹)
Accommodation	49.8	32.8	45.3	42.3	46.3	49.8	46.2	52.1
Food and beverages	93.7	18.5	91.0	17.0	93	26.0	93.8	17.4
Shopping	84.2	21.8	80.8	22.7	82.4	26.6	83.7	23.9
Recreation and leisure	37.9	6.4	38.6	6.6	38.1	9.6	42.2	9.0
Health and medicine	16.5	5.0	19.4	3.5	18.4	3.6	10.3	2.8
Weighted MAPE (percentage)	--	--	5.03	16.09	2.16	27.80	5.10	17.04
Overall weighted MAPE (percentage)	--		7.58		8.10		8.00	