Generalized Extreme Value (GEV)-based Error Structures for Multiple Discrete-Continuous Choice Models

Abdul Rawoof Pinjari
University of South Florida
Department of Civil & Environmental Engineering
ENB 118, 4202 E. Fowler Ave., Tampa, FL 33620
Tel: 813-974-9671, Fax: 813-974-2957
E-mail: apinjari@usf.edu

Generalized Extreme Value (GEV)-based Error Structures for Multiple Discrete-Continuous Choice Models

ABSTRACT

This paper formally derives the class of multiple discrete-continuous generalized extreme value (MDCGEV) models, a general class of multiple discrete-continuous choice models based on generalized extreme value (GEV) error specifications. Specifically, the paper proves the existence of, and derives the general form of, closed-form consumption probability expressions for multiple discrete-continuous choice models with GEV-based error structures. In addition to deriving the general form, the paper derives a compact and readily usable form of consumption probability expressions that can be used to estimate multiple discrete-continuous choice models with general cross-nested error structures.

The cross-nested version of the MDCGEV model is applied to analyze household annual expenditure patterns in various transportation-related expenses using data from a consumer expenditure survey in the United States. Model estimation results and predictive log-likelihood based validation tests indicate the superiority of the cross-nested model over the mutually exclusively nested and non-nested model specifications. Further, the cross-nested model was amenable to the accommodation of socio-demographic heterogeneity in inter-alternative covariance across decision-makers through a parameterization of the allocation parameters.

Keywords: discrete-continuous models, Kuhn-Tucker (KT) demand systems, multiple discreteness, MDCEV, GEV, cross-nested error structure

INTRODUCTION

Traditional single discrete choice models are suited to understanding consumer preferences related to the choice of a single, discrete, alternative out of several available alternatives. However, in several situations, consumer behavior may be associated with the choice of multiple alternatives simultaneously, along with a continuous component of choice for the chosen alternatives. Such multiple discrete-continuous choice situations are being increasingly recognized and modeled in the recent literature in transportation, marketing and economics fields.

A variety of econometric model structures have been used to analyze multiple discrete-continuous choice situations (see Hanemann, 1978; Wales and Woodland, 1983; Kim et al., 2002; and von Haefen and Phaneuf, 2005). Among the available modeling frameworks, the recently developed multiple discrete-continuous extreme value (MDCEV) model structure proposed by Bhat (2005 and 2008) is particularly attractive because of at least two features. First, the model offers simple and elegant closed-form consumption probability expressions that simplify to the well-known multinomial logit probabilities when each decision-maker chooses only one alternative (Bhat, 2005). Second, the model employs a utility specification that enables a clear interpretation of the utility parameters and a convenient specification of the alternative attributes while maintaining the property of weak complementarity (Bhat, 2008).

An important limitation of the MDCEV model, however, is the neglect of potential interdependence (or similarity) among choice alternatives. This is due to an assumption that the stochastic components associated with the utility expressions of the alternatives are independently distributed (or uncorrelated). This assumption is somewhat analogous to the independent and identically distributed (IID) error terms assumption in the MNL model, and may not be justified in several empirical specifications. To relax this assumption, empirical applications in the literature used a mixed MDCEV (MMDCEV) model formulation. A problem with this approach, however, is that the resulting consumption probabilities do not have closed-form expressions. This necessitates the use of simulation-based estimation methods that are computationally expensive, and are associated with accuracy and parameter identification issues.

In a recent paper, Pinjari and Bhat (2010) formulated a multiple discrete-continuous nested extreme value (MDCNEV) model by employing a two-level nested extreme value (NEV) error specification instead of the IID extreme value specification in Bhat's MDCEV formulation.

The MDCNEV model accommodates error term correlations among alternatives in mutually exclusive subsets (or nests) of the choice set, while maintaining the closed-form of probability expressions for any (and all) consumption pattern. This extension of the MDCEV model is analogous to the nested logit extension of the multinomial logit model in that a nested extreme value error structure is assumed instead of an IID extreme value error structure. Thus, similar to the nested logit model, a drawback of the MDCNEV model is the restriction of the alternatives to mutually exclusive nests. On the other hand, in several empirical specifications, it is possible that the alternatives belong to more than one nest. That is, a more general, cross-nested structure may better represent the interactions between the unobserved utility components of various alternatives. Further, since the MDCEV model applications so far involve a large number of highly disaggregate choice alternatives, the likelihood of very general, cross-nested type of correlations is higher due to a greater possibility of similarity across alternatives (see, for example, Kapur and Bhat, 2007; Sener et al., 2008; and Spissu et al., 2009, who use a mixed MDCEV formulation to accommodate cross-nested error structures). Thus, it is important to incorporate more general stochastic specifications that can allow for very general patterns of correlation, while retaining the closed-form of probability expressions. Recognizing such a need, Bhat (2008) suggested that generalized extreme value (GEV) error structures could potentially be used to allow inter-alternative error term correlations while retaining the closed-form of multiple discrete-continuous choice probability expressions. However, neither a formal proof was provided on the existence of, nor a general form was derived for, the closed-form probability expressions from such a Multiple Discrete-Continuous Generalized Extreme Value (MDCGEV) model.

In the context of the preceding discussion, the objectives of this paper are three-fold:

(1) To formally derive the MDCGEV model by proving the existence of, and deriving the general form of, closed-form consumption probability expressions for multiple discrete-continuous choice models with GEV-based error terms,

(2) To derive a compact form of analytical expressions for consumption probabilities of multiple discrete-continuous choice models with general cross-nested error structures, which can be readily used to code the likelihoods for empirical model estimations, and

(3) To conduct an empirical study to assess the importance of cross-nesting structures in multiple discrete-continuous choice models.

The first objective is pursued (in Section 2) by building on the GEV-theory developed by McFadden (1978) in the context of traditional single discrete choice models. Specifically, we derive the class of MDCGEV models, a general class of multiple discrete-continuous choice models based on generalized extreme value (GEV) error specifications. The MDCGEV model includes Bhat's MDCEV and Pinjari and Bhat's (2010) MDCNEV models as special cases. The second objective is pursued (in Section 3) by building on the recent MDCNEV model development of Pinjari and Bhat (2010). The third objective is pursued (in Section 4) by applying the cross-nested model to analyze household annual expenditure patterns in various transportation-related expenses using data from a Consumer Expenditure Survey conducted by the Bureau of Labor Statistics of the United States. Section 5 summarizes and concludes the paper.

2 THE MDCGEV MODEL

Following the traditional utility maximization theory of consumer behavior, assume that consumers maximize the following random utility function proposed by Bhat (2008) for multiple discrete-continuous choice situations:

$$U(\boldsymbol{t}) = \sum_{k=1}^{K} \frac{\gamma_k}{\alpha_k} \exp(\beta' z_k + \varepsilon_k) \left\{ \left(\frac{t_k}{\gamma_k} + 1 \right)^{\alpha_k} - 1 \right\};\ \alpha_k \le 1, \gamma_k > 0$$

subject to the following linear budget constraint and non-negativity constraints on t_k (1)

$$\sum_{k=1}^{K} t_k = T \text{ (where } T \text{ is the total budget) and } t_k \ge 0\ \forall\ k\ (k = 1,2,...,K)$$

In the above consumer demand formulation, $U(\boldsymbol{t})$ is the total utility accrued from consuming $\boldsymbol{t}$ (a Kx1-consumption vector with non-negative consumption quantities t_k) amount of the K alternatives (or goods) available to the decision maker. The term $\exp(\beta' z_k + \varepsilon_k)$, called as the baseline utility for alternative k, represents the random marginal utility of one unit of consumption of alternative k at the point of zero consumption for the alternative (In this term, z_k is a vector of observed alternative-specific and decision-maker attributes, β is the corresponding vector of coefficients, and ε_k is the stochastic term capturing the effect of unobserved factors on baseline utility of alternative k). The α_k terms, called as satiation parameters, serve to capture satiation effects in consumer behavior by reducing the marginal utility accrued from each unit of additional consumption of alternative k (Bhat, 2008). The γ_k terms, called as translation parameters, play a similar role of satiation as that of α_k terms, as well as an additional role of translating the indifference curves associated with the utility function to allow corner solutions (*i.e.,* accommodate the possibility that decision-makers may not consume all alternatives; see Bhat, 2008).

For the above consumer demand problem, assuming that good 1 is always consumed (*i.e.,* $t_1 > 0$), the following stochastic Kuhn-Tucker (KT) conditions can be formed to obtain the optimal consumptions t_k^* (k = 1, 2, 3,…, K) (see Bhat, 2008 for details):

$$V_k + \varepsilon_k = V_1 + \varepsilon_1 \text{ if } t_k^* > 0, \ (k = 2, 3, \ldots, K)$$

$$V_k + \varepsilon_k < V_1 + \varepsilon_1 \text{ if } t_k^* = 0, \ (k = 2, 3, \ldots, K) \tag{2}$$

where, $V_k = \beta' z_k + (\alpha_k - 1)\ln\left(\frac{t_k^*}{\gamma_k} + 1\right)$, ($k$ = 1, 2, 3,…, K). [1]

These stochastic KT conditions can be used to write the joint probability expression of consumption patterns if the density function of the stochastic terms (*i.e.,* the ε_k terms) is known. In the general case, let the joint probability density function (pdf) of the ε_k terms be $g(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_K)$, let M alternatives be chosen out of the available K alternatives, and let the consumptions of these M alternatives be $(t_1^*, t_2^*, t_3^*, \ldots, t_M^*)$. As given in Bhat (2008), the joint probability expression for this consumption pattern is as follows:

$$P(t_1^*, t_2^*, t_3^*, \ldots, t_M^*, 0, 0, \ldots, 0) = |J| \int_{\varepsilon_1=-\infty}^{+\infty} \int_{\varepsilon_{M+1}=-\infty}^{V_1 - V_{M+1} + \varepsilon_1} \int_{\varepsilon_{M+2}=-\infty}^{V_1 - V_{M+2} + \varepsilon_1} \cdots \int_{\varepsilon_{K-1}=-\infty}^{V_1 - V_{K-1} + \varepsilon_1} \int_{\varepsilon_K=-\infty}^{V_1 - V_K + \varepsilon_1}$$
$$g(\varepsilon_1, V_1 - V_2 + \varepsilon_1, V_1 - V_3 + \varepsilon_1, \ldots, V_1 - V_M + \varepsilon_1, \varepsilon_{M+1}, \varepsilon_{M+2}, \ldots, \varepsilon_{K-1}, \varepsilon_K) \tag{3}$$
$$d\varepsilon_K d\varepsilon_{K-1} \ldots d\varepsilon_{M+2} d\varepsilon_{M+1} d\varepsilon_1 \ ,$$

where J is the Jacobian whose elements are given by (see Bhat, 2005) $J_{ih} = \dfrac{\partial[V_1 - V_{i+1} + \varepsilon_1]}{\partial t_{h+1}^*} = \dfrac{\partial[V_1 - V_{i+1}]}{\partial t_{h+1}^*}$; $i, h = 1, 2, \ldots, M-1$.

---

[1] The model formulation presented in the current paper assumes absence of price variation across the choice alternatives. It is straight forward to extend the formulation to the case with price variation. Specifically, as discussed in Bhat (2008), presence of price variation allows the identification of a scale parameter associated with the stochastic terms.

Further, as given in Pinjari and Bhat (2010), the above probability expression can be rewritten as an integral of an M^{th} order partial derivative of the K-dimensional joint cumulative distribution function (CDF) of the error terms $(\varepsilon_1, \varepsilon_2, .., \varepsilon_K)$:

$$P(t_1^*, ..., t_M^*, 0, ..., 0) = |J| \int_{\varepsilon_1=-\infty}^{+\infty} \left[\left\{ \frac{\partial^M}{\partial\varepsilon_1 ... \partial\varepsilon_M} F(\varepsilon_1, \varepsilon_2, ..., \varepsilon_K) \right\} \Bigg|_{\varepsilon_i = V_1 - V_i + \varepsilon_1, \forall i = 1,2,...,K} \right] \mathrm{d}\varepsilon_1 \quad (4)$$

where $F(\varepsilon_1, \varepsilon_2, ..., \varepsilon_K)$ is the joint CDF of the error terms $(\varepsilon_1, \varepsilon_2, .., \varepsilon_K)$. The reader will note here that, in the above expression, the differentials in the partial derivative are with respect to the error terms of the M chosen alternatives.

In Equation (4), the specification of the joint CDF $F(\varepsilon_1, \varepsilon_2, ..., \varepsilon_K)$ of the error terms $(\varepsilon_1, \varepsilon_2, ..., \varepsilon_K)$ determines the form of the consumption probability expressions. In this paper, following McFadden (1978), we assume a GEV form for the CDF as below:

$$F_{GEV}(\varepsilon_1, \varepsilon_2, .., \varepsilon_K) = \exp\left[-G\left(e^{-\varepsilon_1}, e^{-\varepsilon_2}, ..., e^{-\varepsilon_K} \right) \right] \quad (5)$$

where G is a non-negative function with the following properties (McFadden, 1978; Ben-Akiva and Francois, 1983):

1. $G\left(y_1, ...y_i ..., y_K\right) \geq 0, \quad \forall\, y_i > 0 \ (i = 1,2,...,K)$
2. G is homogeneous of degree $\mu > 0$, that is $G\left(ay_1, ...ay_i ..., ay_K\right) = a^{\mu} G(y_1, ...y_i ..., y_K)$,
3. $\lim_{y_i \to +\infty} G\left(y_1, ...y_i ..., y_K\right) = +\infty, \forall i = 1,2,...,K$, and
4. $\left(-1\right)^M \dfrac{\partial^M G(y_1, ...y_K)}{\partial y_1 ... \partial y_M} \leq 0, \quad \forall\, y_i > 0 \ (i = 1,2,...,K)$.

To prove that Equation (5) is a multivariate extreme value distribution, McFadden (1978) expanded the M^{th} order partial derivative of $F_{GEV}(\varepsilon_1, \varepsilon_2, ..., \varepsilon_K)$ as:[2]

$$\frac{\partial^M}{\partial\varepsilon_1 ... \partial\varepsilon_M} F_{GEV}(\varepsilon_1, \varepsilon_2, ..., \varepsilon_K) = e^{-\varepsilon_1} e^{-\varepsilon_2} ... e^{-\varepsilon_M} Q_M F_{GEV} = \prod_{i=1}^{M} e^{-\varepsilon_i} Q_M F_{GEV} \quad (6)$$

[2] In this equation, and from now onwards, for simplicity in notation, $F_{GEV}(\varepsilon_1, \varepsilon_2, ..., \varepsilon_K)$ is represented as F_{GEV}. A formal proof of this expansion (and the GEV-differentiability condition 4) can be found in Daly and Bierlaire (2006). Also see Bierlaire (2006) for a proof for the specific case of cross-nested errors.

In the above equation, Q_M is a recursive function defined as $Q_M = Q_{M-1}G_M - \frac{\partial Q_{M-1}}{\partial y_M}$, where G_M is the partial derivative of G with respect to its M^{th} element and $Q_1 = G_1$. Thus, Q_M, as indicated by McFadden, is a sum of several mixed partial derivative terms, with each term a product of partial derivatives of G of various orders. In each term, the sum of orders of the partial derivatives is equal to *M*. To help the reader understand this better, we expand the terms Q_2, Q_3, and Q_4 below:

$$
\begin{aligned}
&Q_2 = G_1G_2 - G_{12}^2, \\
&Q_3 = G_1G_2G_3 - G_{12}^2G_3 - G_1G_{23}^2 - G_{13}^2G_2 + G_{123}^3, \text{and} \\
&Q_4 = G_1G_2G_3G_4 - G_{12}^2G_3G_4 - G_1G_{23}^2G_4 - G_{13}^2G_2G_4 - G_{14}^2G_2G_3 - G_{24}^2G_1G_3 - G_{34}^2G_1G_2 + G_{12}^2G_{34}^2 + \\
&\qquad G_{13}^2G_{24}^2 + G_{32}^2G_{14}^2 + G_{123}^3G_4 + G_{124}^3G_3 + G_{134}^3G_2 + G_{234}^3G_1 - G_{1234}^4,
\end{aligned}
\tag{7}
$$

where $G_1G_2 = \frac{\partial G}{\partial \varepsilon_1}\frac{\partial G}{\partial \varepsilon_2}$, $G_{12}^2 = \frac{\partial^2 G}{\partial \varepsilon_1 \partial \varepsilon_2}$, and all other derivatives are defined in a similar fashion.

Using Equation (6), the multiple discrete-continuous choice probability of Equation (4) can be expanded as:

$$
P(t_1^*,...,t_M^*,0,...,0) = |J| \int_{\varepsilon_1=-\infty}^{+\infty} \left[\left\{ \prod_{i=1}^{M} e^{-\varepsilon_i} Q_M F_{GEV} \right\} \Bigg|_{\substack{\varepsilon_i = V_1 - V_i + \varepsilon_1, \\ \forall i=1,2,...,K}} \right] \mathrm{d}\varepsilon_1
$$

$$
= |J| \int_{\varepsilon_1=-\infty}^{+\infty} \left[\left\{ \prod_{i=1}^{M} e^{-\varepsilon_i} \begin{pmatrix} \left\{ \pm\left(G_1G_2..G_M\right) \right\} + \\ \left\{ \pm\left(G_{12}^2G_3..G_M\right) \pm \left(G_1G_{23}^2..G_M\right) \pm ... \pm \left(G_1G_2..G_{(M-1)M}^2\right) \right\} + \\ \left\{ \pm\left(G_{123}^3G_4..G_M\right) \pm \left(G_{12}^2G_{34}^2..G_M\right) \pm ... \right\} + \\ ... + \\ \left\{ \pm\left(G_{123..M-1}^{M-1}G_M\right) \pm ... \pm \left(G_1G_{23...M}^{M-1}\right) \right\} + \\ \left\{ \pm\left(G_{123...M}^{M}\right) \right\} \end{pmatrix} F \right\} \Bigg|_{\substack{\varepsilon_i = V_1 - V_i + \varepsilon_1, \\ \forall i=1,2,...,K}} \right] \mathrm{d}\varepsilon_1
\tag{8}
$$

In the above expression, the $\pm$ sign in front of each mixed partial derivative term indicates the possibility that the sign can be either $+$ or $-$ depending on the number of partial derivatives in the term and the number of chosen alternatives. Specifically, the sign before each term is given by $\left(-1\right)^{M+N}$, where N is the number of partial derivatives in the term and M is the number of chosen alternatives. To better understand this, the reader is referred to Equation (7) in which each of the mixed partial derivate terms in Q_2, Q_3, and Q_4 have alternating signs depending on the on the number the partial derivatives in the term.

The probability expression in Equation (8) can be split into several integrals (each of which has a closed-form) and simplified into a closed-form expression as below (see Appendix A for the proof):

$$P(t_1^*,...,t_M^*,0,...,0)=|J|\prod_{i=1}^{M}e^{V_i}\times\begin{bmatrix}\frac{(M-1)!}{H^M}\left\{\pm(H_1H_2..H_M)\right\}+\\ \frac{(M-2)!}{H^{M-1}}\left\{\pm(H_{12}^2H_3..H_M)\pm(H_1H_{23}^2..H_M)\pm...\pm(H_1H_2..H_{(M-1)M}^2)\right\}+\\ \frac{(M-3)!}{H^{M-2}}\left\{\pm(H_{123}^3H_4..H_M)\pm...\pm(H_{12}^2H_{34}^2..H_M)\pm...\right\}+\\ ...\\ \frac{1!}{H^2}\left\{\pm(H_{123...M-1}^{M-1}H_M)\pm...\pm(H_1H_{234...M}^{M-1})\right\}+\\ \frac{0!}{H}\left\{\pm(H_{123...M}^{M})\right\}\end{bmatrix}$$

(9)

where $H_i=\dfrac{\partial H(e^{V_1},..,e^{V_K})}{\partial e^{V_i}}$, $H_{123...n}^n=\dfrac{\partial^n H(e^{V_1},..,e^{V_K})}{\partial e^{V_1}..\partial e^{V_n}}$, and all other terms are defined in a similar fashion[3]. Thus, we formally prove the existence of closed-form consumption probability expressions for multiple discrete-continuous choice models with GEV error structures, and derive a general form for the probability expressions.

One may observe from the general form of probabilities presented in Equation (9) that for a simple MDCEV model with independently distributed error terms, all the terms in the equation with derivatives of order 2 or more will be zero.[4] Thus, Equation (9) simplifies to

[3]Note that G and H are similar functions, but with different arguments; G represents $G(e^{-\varepsilon_1},...,e^{-\varepsilon_n})$, whereas H represents $G(e^{V_1},..e^{V_i}..,e^{V_n})$. Also note from the $\pm$ signs used in the expression that the sign in front of each mixed partial derivative term depends on the number of partial derivatives in the term and the number of chosen alternatives (similar to that in Equation (8)).

[4]This is because for the MDCEV model, $H(e^{V_1},..,e^{V_K})=e^{V_1}+e^{V_2}+...+e^{V_K}$.

$P(t_1^*,...,t_M^*,0,...,0) = |J| \prod_{i=1}^{M} e^{V_i} \frac{(H_1 H_2 .. H_M)}{H^M} (M-1)!$, which further simplifies to the MDCEV probability expression derived by Bhat (2005).

Following the above example of MDCEV, one may wish to use Equation (9) to derive the probability expressions with nested and cross-nested error structures. However, except in cases with small number of choice alternatives and simple nesting structures, it is anything but straight-forward to begin with the most general form of the probability expression provided in Equation (9) and simplify it further to a compact probability expression for general nested and cross-nested error structures. This is because the number of terms in the summation of Equation (9) explodes rather quickly as the number of chosen alternatives (*i.e., M*) increases. This number is equal to the number of ways in which the integer *M* (or the set of *M* choice alternatives) can be "split" or partitioned into any number of positive integers (subsets), as given below:

$$\text{No. of terms in the summation of Equation (9)} = \sum_{i=1}^{M} B(M,i), \qquad (10)$$

where $B(M,i)$ is the number of ways in which the integer *M* (or the set of *M* choice alternatives) can be split into "i" number of positive integers (subsets), given by the recursive relationship $B(M,i) = B(M-1,\, i-1) + B(M-1,\, i) \times i,$ with $B(M,M) = B(M,1) = 1$. One may observe from using these formulae that the number of terms in the summation of the Equation (9) is 15 for *M* = 4 (as shown in Equation 7 for Q_4), 52 for *M* = 5, 877 for *M* = 7, and 115975 for *M* = 10, to understand how quickly the number explodes with increase in *M*. Of course, several of these terms will be zero depending on the nesting structure under consideration. Nonetheless, it is still difficult to assemble all the non-zero terms into a compact expression. To further complicate things, with cross-nested error structures, each term in the summation of the general expression of Equation (9) would in turn be another sum of several terms (with the number of terms depending on the number of nests each alternative is allocated to). In summary, the top down approach of starting with the most general form of MDCGEV probability expressions may not yield compact probability expressions for complex nesting structures. Thus, instead of the top down approach (*i.e.,* beginning from the most general form), we follow a bottom up approach (*i.e.,* beginning from simple error structures to more general error structures) by building on the MDCNEV model developed by Pinjari and Bhat (2010) to derive the multiple discrete-continuous choice probability expressions for general cross-nested error structures. The end goal is to derive compact forms of probability expressions that can be readily used to code the likelihoods for empirical model estimations.

3 MDCGEV MODEL WITH CROSS-NESTED ERROR STRUCTURES

In this section, Section 3.1 briefly describes the MDCNEV model developed by Pinjari and Bhat (2010), Section 3.2 builds on the MDCNEV model to derive and describe a multiple discrete continuous-choice model with general cross-nesting structures, and Section 3.3 provides examples of multiple discrete continuous-choice models with specific cross-nested error structures.

3.1 The MDCNEV Model

To derive the MDCNEV model probabilities, Pinjari and Bhat (2010) employed the nested extreme value (NEV) error term specification with the following cumulative distribution function (CDF):

$$F_{NEV}(\varepsilon_1, \varepsilon_2, .., \varepsilon_K) = \exp\left[-\sum_{s=1}^{S_K} \left\{ \sum_{i \in s^{\text{th}}\text{nest}} \left(Y_i\right)^{1/\theta_s} \right\}^{\theta_s} \right] \tag{11}$$

where, $Y_i = e^{-\varepsilon_i}$, $s\,(=1,2,...,S_K)$ is the index to represent a nest of alternatives and S_K is the total number of nests the K alternatives belong to. $\theta_s\,(0 < \theta_s \le 1; s = 1,2,...,S_K)$ is the (dis)similarity parameter (or the nesting parameter) introduced to capture correlations among the stochastic components of the utilities of alternatives belonging to the s^{th} nest.

Using the above CDF, they break down the M^{th} order partial derivative in Equation (4) into a product of various smaller order partial derivatives, one for each nest, as below:

$$\begin{aligned}
\frac{\partial^M F_{NEV}}{\partial\varepsilon_1 \partial\varepsilon_2 ... \partial\varepsilon_M} &= \prod_{i=1}^{M} \left(\frac{\partial Y_i}{\partial \varepsilon_i} \right) \times \frac{\partial^M F_{NEV}}{\partial Y_1 \partial Y_2 ... \partial Y_M} \\
&= \prod_{i=1}^{M} \left(-e^{-\varepsilon_i} \right) \times (-1)^M \, F_{NEV} \prod_{s=1}^{S_M} \left(\frac{1}{F} \frac{\partial^{q_s} F_{NEV}}{\partial Y_{1s} \partial Y_{2s} ... \partial Y_{q_s}} \right) \\
&= F_{NEV} \prod_{s=1}^{S_M} \left[\left(\prod_{\substack{i \in s^{\text{th}}\text{nest, and} \\ i \in \{\text{chosen alternatives}\}}} e^{-\frac{\varepsilon_i}{\theta_s}} \right) \sum_{r_s=1}^{q_s} \left\{ \left(\sum_{i \in s^{\text{th}}\text{nest}} e^{-\frac{\varepsilon_i}{\theta_s}} \right)^{(q_s - r_s + 1)\theta - q_s} sum(X_{r_s}) \right\} \right]
\end{aligned} \tag{12}$$

In the above equation, $s\ (=1,2,...,S_M)$ is the index to represent the nests to which the M chosen alternatives belong to, q_s is the number of chosen alternatives in the s^{th} nest (thus, $q_1 + q_2 + ... + q_{S_M} = M$), $Y_{1s}, Y_{2s}, ..., Y_{qs}$ are negative exponentials of the error terms associated with each of the chosen alternatives in the s^{th} nest, $sum(X_{rs})$ is a sum of elements of a row matrix X_{rs}, whose form is described in Pinjari and Bhat (2010).

They use the above expression in Equation (4) and follow a series of algebraic rearrangement and integration steps to derive the following form of consumption probability expressions for the MDCNEV model:

$$P(t_1^*, t_2^*, ...t_M^*,\ 0,...,0) =$$

$$|J| \prod_{s=1}^{S_M} \left\{ \frac{\prod\limits_{\substack{i \in s^{\text{th}} \text{nest, and} \\ i \in \{\text{chosen alternatives}\}}} e^{\frac{V_i}{\theta_i}}}{\left(\sum\limits_{i \in s^{\text{th}} \text{nest}} e^{\frac{V_i}{\theta_s}} \right)^{q_s}} \right\} \sum_{r_1=1}^{q_1} ... \sum_{r_s=1}^{q_s} ... \sum_{r_{S_M}=1}^{q_{S_M}} \left\{ \prod_{s=1}^{S_M} \left[\frac{\left(\sum\limits_{i \in s^{\text{th}} \text{nest}} e^{\frac{V_i}{\theta_s}} \right)^{\theta_s}}{\sum\limits_{s=1}^{S_k} \left\{ \left(\sum\limits_{i \in s^{\text{th}} \text{nest}} e^{\frac{V_i}{\theta_s}} \right)^{\theta_s} \right\}} \right]^{q_s - r_s + 1} \left(\prod_{s=1}^{S_M} sum(X_{rs}) \right) \left(\sum_{s=1}^{S_M} (q_s - r_s + 1) - 1 \right)! \right\}$$

(13)

3.2 The MDCGEV Model with General Cross-Nested Errors

In this paper, following Wen and Koppelman (2001), we consider a general cross-nested extreme value (CNEV) distributed error term structure that has the following joint CDF:

$$F_{CNEV}(\varepsilon_1, \varepsilon_2, .., \varepsilon_K) = \exp\left[-\sum_{s=1}^{S_K} \left\{ \sum_{i \in s^{\text{th}} \text{nest}} \left(\alpha_{si} Y_i \right)^{1/\theta_s} \right\}^{\theta_s} \right]; \text{ where } Y_i = e^{-\varepsilon_i}; 0 \le \alpha_{si} \le 1; \sum_{s=1}^{S_K} \alpha_{si} = 1. \quad (14)$$

In the above CDF for cross-nested errors, α_{si} is the allocation parameter corresponding to alternative i and nest s, and all other terms have the same definitions as in Equation (11) for NEV errors. Using this CDF for cross-nested errors, however, unlike in the case of the NEV error structure as in Equation (12), one cannot directly break down the M^{th} order partial derivative in Equation (4) into a product of various smaller order partial derivatives. This is due to the presence of cross-nesting.

Now, consider $\dfrac{\partial F_{CNEV}}{\partial Y_i}$, a first order differential of the cross-nested CDF. Since Y_i can be expanded as: $Y_i = \sum_{s=1}^{S_K}\left(\alpha_{si} Y_i\right)$, $\dfrac{\partial F_{CNEV}}{\partial Y_i}$ can be expanded as:

$$
\begin{aligned}
&\frac{\partial F_{CNEV}}{\partial Y_i} = \sum_{s=1}^{S_K}\left(\frac{\partial\left(\alpha_{si} Y_i\right)}{\partial Y_i}\frac{\partial F_{CNEV}}{\partial\left(\alpha_{si} Y_i\right)}\right) = \sum_{s=1}^{S_K}\left(\alpha_{si}\frac{\partial F_{CNEV}}{\partial\left(\alpha_{si} Y_i\right)}\right)\\
&\text{Thus, } \frac{\partial F_{CNEV}}{\partial \varepsilon_i} = -e^{-\varepsilon_i}\sum_{s=1}^{S_K}\left(\alpha_{si}\frac{\partial F_{CNEV}}{\partial\left(\alpha_{si} Y_i\right)}\right)
\end{aligned}
\tag{15}
$$

Intuitively speaking, the first order partial derivative of the cross-nested CDF with respect to an error term can be expressed as a summation of partial derivatives (multiplied by the corresponding allocation parameter) over all the nests to which the error term belongs. Similarly, higher order partial derivatives can be expressed as a summation of partial derivatives (multiplied by the corresponding allocation parameters) over all the mutually exclusive nesting combinations to which the error terms belong, as below:

$$
\begin{aligned}
&\frac{\partial^2 F_{CNEV}}{\partial Y_1 \partial Y_2} = \sum_{s_1=1}^{S_K}\ \sum_{s_2=1}^{S_K}\left(\alpha_{s_1 1}\alpha_{s_2 2}\frac{\partial^2 F_{CNEV}}{\partial\left(\alpha_{s_1 1} Y_1\right)\partial\left(\alpha_{s_2 2} Y_2\right)}\right)\\
&\text{hence} \frac{\partial^2 F_{CNEV}}{\partial \varepsilon_1 \partial \varepsilon_2} = \left(-e^{-\varepsilon_1}\right)\times\left(-e^{-\varepsilon_2}\right)\times\sum_{s_1=1}^{S_K}\ \sum_{s_2=1}^{S_K}\ \left(\alpha_{s_1 1}\alpha_{s_2 2}\frac{\partial^2 F_{CNEV}}{\partial\left(\alpha_{s_1 1} Y_{1i}\right)\partial\left(\alpha_{s_2 2} Y\right)}\right), \text{ and}
\end{aligned}
\tag{16}
$$

$$
\begin{aligned}
&\frac{\partial^M F_{CNEV}}{\partial Y_1 \partial Y_2 \ldots \partial Y_M} = \sum_{s_1=1}^{S_K}\ \sum_{s_2=1}^{S_K} \cdots \sum_{s_M=1}^{S_K}\left(\prod_{i=1}^{M}\left(\alpha_{s_i i}\right)\frac{\partial^M F_{CNEV}}{\partial\left(\alpha_{s_1 1} Y_1\right)\partial\left(\alpha_{s_2 2} Y_2\right)..\partial\left(\alpha_{s_M M} Y_M\right)}\right)\\
&\text{hence} \frac{\partial^M F_{CNEV}}{\partial \varepsilon_1 \partial \varepsilon_2 \ldots \partial \varepsilon_M} = \prod_{i=1}^{M}\left(-e^{-\varepsilon_i}\right)\times\sum_{s_1=1}^{S_K}\ \sum_{s_2=1}^{S_K} \cdots \sum_{s_M=1}^{S_K}\left(\prod_{i=1}^{M}\left(\alpha_{s_i i}\right)\frac{\partial^M F_{CNEV}}{\partial\left(\alpha_{s_1 1} Y_1\right)\partial\left(\alpha_{s_2 2} Y_2\right)..\partial\left(\alpha_{s_M M} Y_M\right)}\right)
\end{aligned}
\tag{17}
$$

In the above Equations (16) and (17), $s_1 (= 1, 2 \ldots S_K),\ s_2 (= 1, 2 \ldots S_K), \ldots, s_M (= 1, 2 \ldots S_K)$ are indices denoting the nests to which each of the *M* chosen alternatives is allocated. In one extreme case when all indices take the same value, all alternatives are allocated to only one of the S_K nests. In the other extreme case when all indices take different values, then no two alternatives are allocated to the same nest. Thus, each set of values of $s_1, s_2, .., s_M$ represents a particular mutually exclusive nested structure (or a mutually exclusive partitioning) of the M chosen alternatives that can be formed from the given cross-nested structure.

Now, for a mutually exclusive nesting structure (or partitioning) given by $\eth_1, \eth_2, .., \eth_M$, denote $S(\eth_1, \eth_2, .., \eth_M)$ as the set of nests to which the M chosen alternatives are allocated, and $N\left(S(\eth_1, \eth_2, .., \eth_M)\right)$ as the number of nests in the set. Using this notation, following Equation (12), for a mutually exclusive nesting structure given by $\eth_1, \eth_2, .., \eth_M$, the M^{th} order partial derivative $\frac{\partial^M F_{CNEV}}{\partial\left(\alpha_{\eth_1 1}Y_1\right)\partial\left(\alpha_{\eth_2 2}Y_2\right)\ldots\partial\left(\alpha_{\eth_M M}Y_M\right)}$ (which is in the right side of Equation 17) can be broken down into a product of $N\left(S(\eth_1, \eth_2, .., \eth_M)\right)$ number of smaller order partial derivatives, as below:

$$\frac{\partial^M F_{CNEV}}{\partial\left(\alpha_{\eth_1 1}Y_1\right)\partial\left(\alpha_{\eth_2 2}Y_2\right)\ldots\partial\left(\alpha_{\eth_M M}Y_M\right)} =$$

$$(-1)^M F_{CNEV} \times \prod_{\eth_i \in S(\eth_1,\eth_2,..,\eth_M)} \left\{ \left(\prod_{\substack{i\in \eth_i^{\text{th}}\text{nest, and}\\ i\in\{\text{chosen alternatives}\}}} \left(\alpha_{\eth_i i}Y_i\right)^{\frac{1-\theta_{\eth_i}}{\theta_{\eth_i}}} \right) \sum_{r_{\eth_i}=1}^{q_{\eth_i}} \left(\left(\sum_{i\in\eth_i^{\text{th}}\text{nest}} \left(\alpha_{\eth_i i}Y_i\right)^{\frac{1}{\theta_{\eth_i}}} \right)^{(q_{\eth_i}-r_{\eth_i}+1)\theta_{\eth_i}-q_{\eth_i}} sum(X_{r_{\eth_i}}) \right) \right\} \tag{18}$$

Using the above expression in Equation (17), the M^{th} order partial derivative in Equation (4) for cross-nested CDF can be expressed as:

$$\frac{\partial^M F_{CNEV}}{\partial\varepsilon_1\partial\varepsilon_2\ldots\partial\varepsilon_M} = \prod_{i=1}^{M} e^{-\varepsilon_i} \times$$

$$\sum_{\eth_M=1}^{S_K}\sum_{\eth_2=1}^{S_K}..\sum_{\eth_1=1}^{S_K}\left[\prod_{i=1}^{M}\alpha_{\eth_i i}\, F_{CNEV} \prod_{\eth_i \in S(\eth_1,\eth_2,..,\eth_M)} \left\{ \left(\prod_{\substack{i\in \eth_i^{\text{th}}\text{nest, and}\\ i\in\{\text{chosen alts}\}}} \left(\alpha_{\eth_i i}Y_i\right)^{\frac{1-\theta_{\eth_i}}{\theta_{\eth_i}}} \right) \sum_{r_{\eth_i}=1}^{q_{\eth_i}} \left(\left(\sum_{i\in\eth_i^{\text{th}}\text{nest}} \left(\alpha_{\eth_i i}Y_i\right)^{\frac{1}{\theta_{\eth_i}}} \right)^{(q_{\eth_i}-r_{\eth_i}+1)\theta_{\eth_i}-q_{\eth_i}} sum(X_{r_{\eth_i}}) \right) \right\} \right]$$

$$= F_{CNEV} \times \sum_{\eth_M=1}^{S_K}\sum_{\eth_2=1}^{S_K}..\sum_{\eth_1=1}^{S_K}\left[\prod_{\eth_i \in S(\eth_1,\eth_2,..,\eth_M)} \left\{ \left(\prod_{\substack{i\in \eth_i^{\text{th}}\text{nest, and}\\ i\in\{\text{chosen alternatives}\}}} \left(\alpha_{\eth_i i}Y_i\right)^{\frac{1}{\theta_{\eth_i}}} \right) \sum_{r_{\eth_i}=1}^{q_{\eth_i}} \left(\left(\sum_{i\in\eth_i^{\text{th}}\text{nest}} \left(\alpha_{\eth_i i}Y_i\right)^{\frac{1}{\theta_{\eth_i}}} \right)^{(q_{\eth_i}-r_{\eth_i}+1)\theta_{\eth_i}-q_{\eth_i}} sum(X_{r_{\eth_i}}) \right) \right\} \right] \tag{19}$$

Using Equation (19) in Equation (4) and following the MDCNEV derivations of Pinjari and Bhat (2010), one can derive the following MDCGEV consumption probabilities with cross-nested errors (derivation details are available with the author):

$P(t_1^*, t_2^*, ... t_M^*, 0, ..., 0) =$

$$\sum_{\partial_1=1}^{S_K}\sum_{\partial_2=1}^{S_K}..\sum_{\partial_M=1}^{S_K}\left[\begin{array}{l} |J| \prod_{\partial_i \in S(\partial_1,\partial_2,..,\partial_M)} \left\{ \dfrac{\prod_{\substack{i\in\partial_i^{\text{th}}\text{nest, and} \\ i\in\{\text{chosen alternatives}\}}} \left(\alpha_{\partial_i i} e^{V_i}\right)^{\frac{1}{\theta_{\partial_i}}}}{\left(\sum_{i\in\partial_i^{\text{th}}\text{nest}} \left(\alpha_{\partial_i i} e^{V_i}\right)^{\frac{1}{\theta_{\partial_i}}}\right)^{q_{\partial_i}}} \right\} \times \\ \sum..\sum_{r_{\partial_i}=1}^{q_{\partial_i}}..\sum_{\forall \partial_i \in S(\partial_1,\partial_2,..,\partial_M)} \left\{ \prod_{\partial_i \in S(\partial_1,\partial_2,..,\partial_M)} \left[\left(\dfrac{\left(\sum_{i\in\partial_i^{\text{th}}\text{nest}} \left(\alpha_{\partial_i i} e^{V_i}\right)^{\frac{1}{\theta_{\partial_i}}}\right)^{\theta_{\partial_i}}}{H} \right)^{q_{\partial_i} - r_{\partial_i} + 1} sum(X_{r_{\partial_i}}) \right] \left(\sum_{\partial_i \in S(\partial_1,\partial_2,..,\partial_M)} (q_{\partial_i} - r_{\partial_i} + 1) - 1 \right)! \right\} \end{array}\right] \tag{20}$$

One may note from the probability expression in Equation (20) that the multiple discrete-continuous consumption probabilities with cross-nested error terms can be expressed as a summation of several terms. The number of terms in the summation is equal to the number of different mutually exclusive nesting combinations (or partitions) of the *M* chosen alternatives $(\partial_1, \partial_2, .., \partial_M)$ that can be formed from the given cross-nested structure.[5] Each of these terms in the summation in turn contains a summation of $\prod_{\forall \partial_i \in S(\partial_1,\partial_2,..,\partial_M)} q_{\partial_i}$ terms, where q_{∂_i} is the number of chosen alternatives in nest ∂_i of the specific mutually exclusive nesting structure or partition $(\partial_1, \partial_2, .., \partial_M)$ of the *M* chosen alternatives. These nuances can be best explained using specific examples, as in the next section.

---

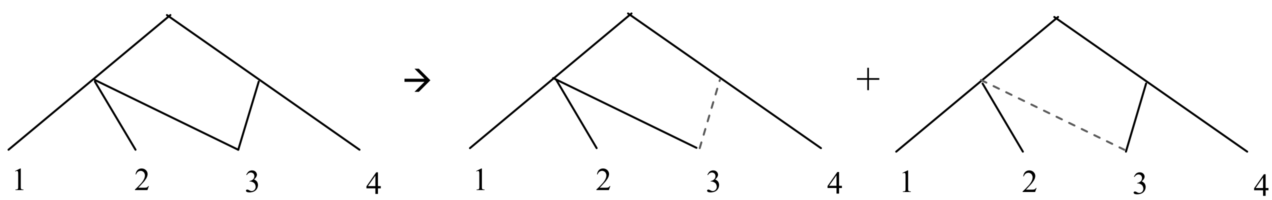
[5]This number is equal to the number of terms in the summation $\sum_{\partial_1=1}^{S_K}\sum_{\partial_2=1}^{S_K}..\sum_{\partial_M=1}^{S_K}[.]$ in Equation (20). It should be noted, however, that the terms in this summation do not necessarily represent the consumption probabilities as such from the mutually exclusive nested structure given by $(\partial_1, \partial_2, .., \partial_M)$. This is because these terms include $\sum_{i\in\partial_i^{\text{th}}\text{nest}} \left(\alpha_{\partial_i i} Y_i\right)^{\frac{1}{\theta_{\partial_i}}}$ and $H = \sum_{\partial=1}^{S_K} \left\{ \sum_{i\in\partial^{\text{th}}\text{nest}} \left(\alpha_{\partial i} e^{V_i}\right)^{1/\theta_\partial} \right\}^{\theta_\partial}$ which are obtained from the primary cross-nested structure, not from the mutually exclusive nested structure given by $(\partial_1, \partial_2, .., \partial_M)$. The term "mutually exclusive nested structure" is used here for the convenience of counting the number of terms in the summation $\sum_{\partial_1=1}^{S_K}\sum_{\partial_2=1}^{S_K}..\sum_{\partial_M=1}^{S_K}[.]$.

3.3 Examples of MDCGEV Models with General Cross-Nested Errors

The probability expression derived in Equation (20) can be used with different types of cross-nested error structures, including simple cross-nested errors (*e.g.* Vovsha, 1997), pair-wise correlated errors (Chu, 1989; Wen and Koppelman, 2001), and Ordered GEV type of errors (Small, 1987). In this section, we illustrate the form of probability expressions for a simple cross-nested structure (Section 3.3.1) and describe the form of probability expressions for the pair-wise correlated error structure (Section 3.3.2).

3.3.1 A simple Cross-nested Error Structure

Consider a simple cross-nested structure $\{\{1,2,3\},\{3,4\}\}$ that has 4 elemental alternatives in two overlapping nests labeled as A and B, respectively. Specifically, alternative 3 belongs to two nests A = $\{1,2,3\}$ and B = $\{3,4\}$. From this cross-nested structure, one can form two mutually exclusive nested structures (or partitions): $\{\{1,2,3\},\{4\}\}$, and $\{\{1,2\},\{3,4\}\}$, as below:



Following Wen and Koppelan (2001), define the CDF of this error structure as:

$$F_{CNEV}(\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4)=\exp-\left[\left\{Y_1^{1/\theta_A}+Y_2^{1/\theta_A}+\left(\alpha_{A3}Y_3\right)^{1/\theta_A}\right\}^{\theta_A}+\left\{\left(\alpha_{B3}Y_3\right)^{1/\theta_B}+Y_4^{1/\theta_B}\right\}^{\theta_B}\right] \quad (21)$$

where $Y_i=e^{-\varepsilon_i}$, and $(\alpha_{A3}+\alpha_{B3})=1$

Now, define $\left(\left(e^{V_1}\right)^{\frac{1}{\theta_A}}+\left(e^{V_2}\right)^{\frac{1}{\theta_A}}+\left(\alpha_{A3}e^{V_3}\right)^{\frac{1}{\theta_A}}\right)$ as h_A, $\left(\left(\alpha_{B3}e^{V_3}\right)^{\frac{1}{\theta_B}}+\left(e^{V_4}\right)^{\frac{1}{\theta_B}}\right)$ as h_B, and $\left(\left(e^{V_1}\right)^{\frac{1}{\theta_A}}+\left(e^{V_2}\right)^{\frac{1}{\theta_A}}+\left(\alpha_{A3}e^{V_3}\right)^{\frac{1}{\theta_A}}\right)^{\theta_A}+\left(\left(\alpha_{B3}e^{V_3}\right)^{\frac{1}{\theta_B}}+\left(e^{V_4}\right)^{\frac{1}{\theta_B}}\right)^{\theta_B}$ as h. Using this notation, for the above nesting structure, the consumption probabilities for some selected consumption patterns are given by the following expressions:

$$P(t_1^*,0,0,0)=|J|\left\{\frac{\left(e^{V_1}\right)^{\frac{1}{\theta_A}}}{\left(h_A\right)^1}\right\}\times\left\{\left(\frac{\left(h_A\right)^{\theta_A}}{h}\right)^1 sum(X_{1_A})\ 0!\right\} \tag{22}$$

$$P(t_1^*,t_2^*,0,0)=|J|\left\{\frac{\left(e^{V_1}\right)^{\frac{1}{\theta_A}}\left(e^{V_2}\right)^{\frac{1}{\theta_A}}}{\left(h_A\right)^2}\right\}\times\left\{\left(\frac{\left(h_A\right)^{\theta_A}}{h}\right)^2 sum(X_{1_A})\ 1!+\left(\frac{\left(h_A\right)^{\theta_A}}{h}\right)^1 sum(X_{2_A})\ 0!\right\} \tag{23}$$

$$P(0,0,t_3^*,0)=|J|\left\{\frac{\left(\alpha_{A3}e^{V_3}\right)^{\frac{1}{\theta_A}}}{\left(h_A\right)^1}\right\}\times\left\{\left(\frac{\left(h_A\right)^{\theta_A}}{h}\right)^1 sum(X_{1_A})\ 0!\right\}+|J|\left\{\frac{\left(\alpha_{B3}e^{V_3}\right)^{\frac{1}{\theta_B}}}{\left(h_B\right)^1}\right\}\times\left\{\left(\frac{\left(h_B\right)^{\theta_B}}{h}\right)^1 sum(X_{1_B})\ 0!\right\} \tag{24}$$

$$\begin{aligned}P(t_1^*,0,t_3^*,0)=&|J|\left\{\frac{\left(e^{V_1}\right)^{\frac{1}{\theta_A}}\left(\alpha_{A3}e^{V_3}\right)^{\frac{1}{\theta_A}}}{\left(h_A\right)^2}\right\}\times\left\{\left(\frac{\left(h_A\right)^{\theta_A}}{h}\right)^2 sum(X_{1_A})\ 1!+\left(\frac{\left(h_A\right)^{\theta_A}}{h}\right)^1 sum(X_{2_A})\ 0!\right\}+\\&|J|\left\{\frac{\left(e^{V_1}\right)^{\frac{1}{\theta_A}}\left(\alpha_{B3}e^{V_3}\right)^{\frac{1}{\theta_B}}}{\left(h_A\right)^1\left(h_B\right)^1}\right\}\times\left\{\left(\frac{\left(h_A\right)^{\theta_A}}{h}\right)^1\left(\frac{\left(h_B\right)^{\theta_B}}{h}\right)^1 sum(X_{1_A})sum(X_{1_B})\ 1!\right\}\end{aligned} \tag{25}$$

$$\begin{aligned}&P(t_1^*,t_2^*,t_3^*,0)=\\&|J|\left\{\frac{\left(e^{V_1}\right)^{\frac{1}{\theta_A}}\left(e^{V_2}\right)^{\frac{1}{\theta_A}}\left(\alpha_{A3}e^{V_3}\right)^{\frac{1}{\theta_A}}}{\left(h_A\right)^3}\right\}\times\left\{\left(\frac{\left(h_A\right)^{\theta_A}}{h}\right)^3 sum(X_{1_A})\ 2!+\left(\frac{\left(h_A\right)^{\theta_A}}{h}\right)^2 sum(X_{2_A})\ 1!+\left(\frac{\left(h_A\right)^{\theta_A}}{h}\right)^1 sum(X_{2_A})\ 0!\right\}+\\&|J|\left\{\frac{\left(e^{V_1}\right)^{\frac{1}{\theta_A}}\left(e^{V_2}\right)^{\frac{1}{\theta_A}}\left(\alpha_{B3}e^{V_3}\right)^{\frac{1}{\theta_B}}}{\left(h_A\right)^2\left(h_B\right)^1}\right\}\times\left\{\left(\frac{\left(h_A\right)^{\theta_A}}{h}\right)^2\left(\frac{\left(h_B\right)^{\theta_B}}{h}\right)^1 sum(X_{1_A})sum(X_{1_B})\ 2!+\left(\frac{\left(h_A\right)^{\theta_A}}{h}\right)^1\left(\frac{\left(h_B\right)^{\theta_B}}{h}\right)^1 sum(X_{1_A})sum(X_{2_B})\ 1!\right\}\end{aligned} \tag{26}$$

where, $sum(X_{1_A})=1$, $sum(X_{1_B})=1$, $sum(X_{2_A})=(1-\theta_A)/\theta_A$, and $sum(X_{2_B})=(1-\theta_B)/\theta_B$.

As one may observe from the above expressions, the probability expressions that involve a non-zero consumption of alternative three (that belongs to more than one nest) are a sum of two terms, each corresponding to a mutually exclusive nesting structure (or partitioning) of the chosen alternatives that can be formed from the given cross nesting structure. Each of these terms in turn contains a summation of $\prod_{\forall\, \delta_i \in S(\delta_1,\delta_2,..,\delta_M)} q_{\delta_i}$ terms. For example, the first lines of Equations (25) and (26) are based on the mutually exclusive nesting structure $\{\{1,2,3\},\{4\}\}$, whereas the second lines of these Equations are based on the mutually exclusive nesting structure $\{\{1,2\},\{3,4\}\}$. The terms h_A, h_B, and h in these equations, however, are obtained from the primary cross-nesting structure $\{\{1,2,3\},\{3,4\}\}$.

3.3.2 *The Pair-wise Correlated Error Structure*

Wen and Koppelman (2001) used the following pair-wise correlated error structure (a more general form of the paired combinatorial logit (PCL) structure used by Chu, 1989 and Koppelman and Wen, 2000) in the context of single discrete choice analysis:

$$F_{PGEV}(\varepsilon_1, \varepsilon_2, .., \varepsilon_K) = \exp\left[-\sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \left\{ \left(\alpha_{i,ij} Y_i\right)^{1/\theta_{ij}} + \left(\alpha_{j,ij} Y_j\right)^{1/\theta_{ij}} \right\}^{\theta_{ij}} \right]; \quad Y_i = e^{-\varepsilon_i} \text{ and } \sum_{i=1}^{k-1} \alpha_{i,ij} = 1 \quad (27)$$

In the above error structure, each alternative pair (denoted by *ij*) represents a nest. The number of such alternative pairs (or nests) for a total of *K* choice alternatives = *K(K-1)/2,* and the number of nests to which each alternative may be allocated = *K-1.* The same error structure can be used for multiple discrete-continuous choice situations, and Equation (20) can be used to write the consumption probability expressions. The resulting expression will be a summation of terms over a total of $(K-1)^M$ number of mutually exclusive nesting combinations (or partitioning combinations) of the *M* chosen alternatives. The number of summations within each of these terms (*i.e.,* the value of $\prod_{\forall\, \delta_i \in S(\delta_1, \delta_2, .., \delta_M)} q_{\delta_i}$) can range anywhere from 1 to $2^{\lfloor M/2 \rfloor}$, because q_{δ_i} can take either a value of 1 or a value of 2.

4 EMPIRICAL ANALYSIS

4.1 Empirical Context

We employed the MDCGEV model with cross-nested errors to an empirical case of household transportation expenditures, using the 2002 Consumer Expenditure (CEX) Survey (conducted by the Bureau of Labor Statistics; BLS, 2003) data obtained from the archives of the National Bureau of Economic Research (NBER, 2003). The final sample used in this analysis consists of data from 4101 households, of which data from 4000 households were used for model estimation and data from the remaining 101 households were kept aside for model validation. The details of the data and sample extraction are available in Ferdous *et al.,* (2010) who originally extracted and processed the data for an analysis of household expenditures in 17 categories that range from housing, food, and utilities, to a variety of transportation alternatives. The current study limits the analysis to transportation expenditures only. Specifically, the focus is on the proportion of annual income spent in the following transportation categories: (1) Vehicle purchase, (2) Gasoline and motor oil (termed as gasoline in the rest of the document), (3) Vehicle insurance, (4) Vehicle maintenance, (5) Air travel, and (6) Public transportation. Along with these 6 transportation expenditure categories, the analysis includes an "other" category that includes all other expenses and the remaining portion of the annual income (if any). The proportions of expenditures in all these 7 alternatives add up to 100. While the expenditure proportions in the transportation categories can be zero for some households, those in the "other" category are greater than zero for all households. That is, the "other" category acts as an "outside" good (or essential Hicksian good) in the model specification.

4.2 Model Specification

Several model specifications were explored in this empirical analysis. The model specification process started with the determination of the best MDCEV specification. Subsequently, several mutually exclusive nesting (*i.e.,* MDCNEV) specifications were estimated – with two-alternative nests, three-alternative nests, a four-alternative nest, a five alternative-nest, and a six-alternative nest. Among these MDCNEV specifications, a four-alternative nested model (with all four, personal vehicle transportation related expenditures – vehicle purchase, gasoline, vehicle insurance, and vehicle maintenance – in the nest) provided the best data fit. Subsequently, to guide the cross-nesting explorations, the two-alternative nests that did not yield statistically significant nesting parameters were used to eliminate certain cross-nesting specifications. Thus, all cross-nesting explorations were with the four, personal vehicle transportation-related alternatives. Specifically, in each cross-nesting exploration, one (subsequently, two) of the four alternatives was (were) allocated to more than one nest. From all these trials, a cross-nested model with two overlapping nests – a {vehicle purchase, gasoline} nest, and a {gasoline, vehicle insurance, vehicle maintenance} nest – provided the best data fit.[6]

Model estimations were carried out by coding the log-likelihood functions within the maximum likelihood module of GAUSS matrix programming language; constraints on the nesting parameters and allocation parameters were ensured by employing logit functional forms.

---

[6]Exploration of the pair-wise correlated specification (within the four alternatives) was deemed unnecessary because some of the two-alternative, mutually exclusive nested specifications did not yield significant nesting parameters. Further, in some specifications, the allocation parameter estimates indicated the absence of certain cross nests.

After determining the best fitting cross-nested model, the allocation parameters of this model were parameterized as a logit function of household socio-demographic attributes to explore the presence of systematic heterogeneity in inter-alternative covariance.[7]

4.3 Model Estimation Results

Table 1 shows the results of the best MDCEV specification, the MDCNEV specification with the best model fit, and two cross-nested specifications – the best-fit cross-nested model with no heterogeneity in allocation parameters, and the best-fit cross-nested model with heterogeneity in allocation parameters. In all these models, the parameters in the deterministic part of the utility function include baseline utility parameters (baseline constants, household socio-demographic, residential location and regional attribute effects), and translation (γ_k) parameters (satiation parameters were constrained to be zero). All these parameters have reasonable and expected substantive interpretations that are similar to those found in Ferdous *et al.,* (2010). However, the magnitudes (and t-statistics) of several parameters corresponding to the alternatives included in the nests (especially the translation parameter for the vehicle purchase alternative) are substantially different between the MDCEV and MDCNEV models, and considerably different between the MDCNEV and cross-nested models.

In the MDCNEV model as well as in both the cross-nested models, the nesting parameters are highly significantly different from 1 (see the block of rows labeled “Nesting Parameters” in the table). In the cross-nested specification with no heterogeneity, the allocation parameters, 0.873 and 0.127, are statistically different from 1 and 0, respectively, indicating the presence of cross-nested covariance structure. However, the allocation of the gasoline alternative is more skewed toward the {gasoline, vehicle insurance, vehicle maintenance} nest (see the first two rows in the block of rows labeled “Allocation Parameters”).

In terms of model fit, the MDCNEV model substantially outperforms the MDCEV model based on a nested likelihood ratio test, but it can be rejected in favor of both the cross-nested specifications at a significance level less than 0.001 based on a non-nested test. Thus the cross-nested models clearly outperform the MDCNEV model. Further, it can be observed from the log-likelihood values (last row of the table) that while the homogenous cross-nested model shows a

---

[7]See Newman (2009), who proposed this idea in the context of single discrete choice models, for a detailed discussion on the relative ease and advantages of accommodating heterogeneity in allocation parameters as opposed to doing so with the nesting parameters.

substantial improvement in the log-likelihood from the MDCNEV model, the heterogeneous cross-nested model shows a modest (albeit statistically significant) improvement from the homogenous cross-nested model. That is, parameterization of the allocation parameters as a logit function of socio-demographic attributes resulted in modest improvement in the model fit. However, the parameter estimates (of the logit function of the allocation parameters) indicate interesting differences in the nesting (hence covariance) patterns among different population segments based on income (see the last note at the bottom of the table). Specifically, the allocation parameters for low-income (income < 30K) households are highly skewed toward the {gasoline, vehicle insurance, vehicle maintenance} nest with a very small value of the allocation parameter (0.07) for the {vehicle purchase, gasoline} nest. This result suggests that the {vehicle purchase, gasoline} nest is almost absent for low-income households. On the other hand, for high-income households, the allocation of the gasoline alternative between the {vehicle purchase, gasoline} nest and the {gasoline, vehicle insurance, vehicle maintenance} nest is a bit more balanced for than that for low-income households. Such differences in allocation parameters between low-income and high-income households can be explained as follows. Income is likely to be the primary constraint driving the vehicle purchase decisions of low-income households (90% of the low-income households in the current estimation data do not show any vehicle purchase-related expenditures, while 40% of the high-income households show vehicle purchase-related expenditures). Once this effect of income (and other variables, such as the number of vehicles already owned by the household) is incorporated in the deterministic utility specification, the magnitude of unobserved factors influencing vehicle purchase decisions of low-income households may be relatively small compared to that for high-income households. As a result, the co-variation between the unobserved factors influencing vehicle purchase expenditures and those influencing gasoline expenditures may be relatively small for low-income households (hence the absence of the nest with vehicle-purchase gasoline alternatives for low-income households). Such co-variation, if any, is more likely to be in the context of higher-income households.

4.4 Model Evaluation

The above discussed model estimation results indicate that the cross-nested models are preferred specifications (over the MDCEV and MDCNEV specifications) in the current empirical context, and that ignoring cross-nesting (and any heterogeneity in cross-nesting) can potentially lead to inferior model fit (and distorted inferences regarding the covariance patterns). Better data fit (than the nested and simple non-nested specifications) is certainly a desirable (and perhaps expected) feature of the cross-nested models, but it should not be the sole criterion for model selection; prediction superiority plays an important role. Thus, the performance of the estimated models was tested against a hold-out sample (i.e., a validation sample) of 101 households. Specifically, a predictive log-likelihood function (PLLF) was computed using the hold-out sample for all the models estimated in the paper. The PLLF was computed by plugging in the out-of-sample observations into the log-likelihood function, while retaining the estimated parameters from the estimation sample. Table 2 reports the PLLF values for the entire validation sample (of 101 households) as well as for different income segments within the sample. As can be observed from the first row, The PLLF value for the MDCNEV model is substantially higher than that for the MDCEV model, while the PLLF values for both the cross-nested models are considerably higher than that for the MDCNEV model. This result suggests that the cross-nested models perform better than the simpler MDCNEV model even in the validation sample. This trend is consistent for all income-classes. However, between the MDCNEV and the cross-nested models, the medium-income segment shows a notable improvement in the PLLF, while the improvement in the PLLF value for the other two segments (low- and high-income households) is moderate. Further, between the two cross-nested models (see the last two rows in the table), the improvement in the PLLF value due to the incorporation of heterogeneity in allocation parameters is rather small.

Overall, the predictive log-likelihood based measures discussed in this section suggest that the cross-nested models perform better than the MDCNEV model. However, there is no convincing evidence that the cross-nested model with income-based heterogeneity in allocation parameters performs better than the cross-nested model without such heterogeneity. To better understand the value of cross-nested models over simpler models and to further assess the value of incorporating heterogeneity in the allocation parameters, it is important to conduct extensive policy prediction analyses or simulation experiments. In this context, it is necessary to first develop methods to simulate multivariate extreme value distributions with cross-nested and other GEV-based correlation patterns.[8]

---

[8]While a handful of studies (McNeil, 2005; Bodea and Garrow, 2006; Wu *et al.,* 2006) discuss procedures to simulate nested extreme value distributions, the literature seems to be lacking in studies that simulate cross-nested

4.5 Model Computation Time

An important aspect in the context of complex models as described in this paper is the computation time (or run-time) for model estimation. Pinjari and Bhat (2010) report very high run times (about 10 hours) for an MDNEV model estimation in their empirical context with 32 choice alternatives and 6167 decision-makers. As they indicated, such a high run-time may be attributed largely to the fact that only the likelihood function was coded and the analytical gradients were not explicitly provided to the log-likelihood optimization routine. The cross-nested model estimation code used for the current paper builds on the MDCNEV code used in Pinjari and Bhat (2010), but also incorporates some changes that enhance computational efficiency, although the analytical gradients have not been coded yet. In the current empirical context with 4000 households and 7 choice alternatives, it took 11 minutes to estimate the MDCEV model, 15 minutes to estimate the MDCNEV model, 33 minutes to estimate the cross-nested model without heterogeneity, and 58 minutes to estimate the cross-nested model with heterogeneity in allocation parameters. For comparison purposes, all the run-times reported here, including the run-time for the MDCEV model are based on model estimation codes without analytical gradients.[9] In addition, all these run-times are based on default starting values for the parameters, such as zeros for the coefficients in the deterministic utility function and ones for nesting parameters.[10] When compared to the run times reported in Pinjari and Bhat (2010), the relatively lower run-times in the current empirical context can be attributed primarily to the smaller number of choice alternatives. However, in complex choice situations with large number of choice alternatives and very general cross-nesting structures, the issue of model run times is anything but trivial. In such situations, it is anticipated that coding the gradients of the log-likelihood function can help reduce the run-times substantially.

5. SUMMARY AND CONCLUSIONS

This paper derives the class of multiple discrete-continuous generalized extreme value (MDCGEV) models, a general class of multiple discrete-continuous choice models based on

---

and other GEV distributions (see McFadden, 1999 for a Markov Chain Monte Carlo procedure to generate random draws that approximate GEV distributions).

[9] For the MDCEV model, the estimation code with analytical gradients takes less than 2 minutes for convergence. All the run times reported here are for a dual core computer of 2.66GHz processing speed and 3.5GB RAM.

[10] The estimation times of the cross-nested models reduced by half when the MDCEV model estimates were provided as the starting values for the parameters in the deterministic utility function.

generalized extreme value (GEV) error specifications. Specifically, the paper proves the existence of, and derives the general form of, closed-form consumption probability expressions for multiple discrete-continuous choice models with generalized extreme value (GEV) error structures. In addition to deriving the general form, the paper derives a more compact and readily usable form of consumption probability expressions that can be used to estimate multiple discrete-continuous choice models with cross-nested error structures. While the former task builds on McFadden (1978)'s derivation of the GEV models for single discrete choice occasions, the latter task builds on the recent MDCNEV model development by Pinjari and Bhat (2010).

The cross-nested model is applied to analyze household annual expenditure patterns in various transportation-related expenses using data from a consumer expenditure survey in the United States. The model estimation results highlight the superiority of the cross-nested model over the MDCNEV and the simpler MDCEV model specifications in terms of model fit. Further, the cross-nested model was amenable to accommodation of demographic heterogeneity in inter-alternative covariance across decision-makers through a parameterization of the allocation parameters. Such a heterogeneous cross-nested model suggested heterogeneity in the nesting patterns in the population based on household income levels. When evaluated against a validation sample, both the cross-nested models performed better than the simpler nested (MDCNEV) and non-nested (MDCEV) models. However, there is no significant difference in performance between the cross-nested model with demographic heterogeneity (in the allocation parameters) and the cross-nested model without heterogeneity.

In the context of these developments, it is important to understand the substitution properties of closed-form multiple discrete-continuous choice models with cross-nested error structures vis-à-vis models with simple nested and non-nested errors. To this end, there is an immediate need for methods to easily and accurately simulate GEV distributions. Another interesting avenue for subsequent work would be to incorporate network-GEV type of error structures (as in Daly and Bierlaire, 2006) in multiple discrete-continuous choice models.

ACKNOWLEDGEMENTS

The author is grateful to Chandra Bhat, discussions with whom inspired and benefited this research. Two anonymous reviewers provided helpful comments on an earlier manuscript. Of course, any errors are the author's own responsibility. Thanks to Nazneen Ferdous for providing the empirical data used in this paper. This research was funded by a faculty start-up grant at the University of South Florida.

REFERENCES

BLS (2003) 2002 Consumer Expenditure Interview Survey Public Use Microdata Documentation. U.S. Department of Labor Bureau of Labor Statistics, Washington, D.C. http://www.bls.gov/cex/csxmicrodoc.htm#2002.

Ben-Akiva, M., and B. Francois (1983) μ-homogeneous Generalized Extreme Value Model. Working paper, Massachusetts Institute of Technology.

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.

Bierlaire, M. (2006) A Theoretical Analysis of the Cross-nested Logit Model. *Annals of Operations Research*, 144(1), 287-300.

Bodea, T.D., and L.A. Garrow (2006) The Importance of Synthetic Datasets in Empirical Testing: Comparison of NL with MNL with Error Components Models. Proceedings of the European Transport Conference, Strasbourg, September 2006.

Chu, C.A (1989) Paired Combinatorial Logit Model for Travel Demand Analysis. Proceedings of 5th World Conference on Transportation Research, vol 4. Ventura, CA, 295-309.

Daly, A., and M. Bierlaire (2006) A General and Operational Representation of Generalized Extreme Value Models. *Transportation Research Part B*, 40(4), 285-305.

Ferdous, N., Pinjari A.R., Bhat C.R., and R.M. Pendyala (2010). A Comprehensive Analysis of Household Transportation Expenditures Relative to Other Goods and Services: An Application to United States Consumer Expenditure Data. *Transportation*, 37(3), 363-390.

Hanemann, W.M., (1979). A Methodological and Empirical Study of Recreation Benefits from Water Quality Improvement. Ph.D. Dissertation, Department of Economics, Harvard University.

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

Kapur, A., and C.R. Bhat (2007) On Modeling Adults' Daily Time Use by Activity Purpose and Accompaniment Arrangement. *Transportation Research Record*, 2021, 18-27.

Koppelman, F.S., and C.-H. Wen (2000) The Paired Combinatorial Logit Model: Properties, Estimation and Application. *Transportation Research Part B*, 34(2), 75-89.

McFadden, D. (1978) Modelling the choice of residential location, in A. Karlquist *et al.* (eds.), *Spatial Interaction Theory and Residential Location*, North-Holland, Amsterdam, 75-96.

McFadden, D. (1999) Computing Willingness-to-pay in Random Utility Models, in J.C. Moore *et al.* (eds.), *Trade, Theory and Econometrics: Essays in Honor of John S. Chipman*, Routledge, London, 253-274.

McNeil, A.J., R. Frey, and P. Embrechts (2005) Quantitative Risk Management: Concepts, Techniques and Tools. Princeton University Press.

NBER (1980-2003) The National Bureau of Economic Research (NBER) Archive of Consumer Expenditure Survey Microdata Extracts. http://www.nber.org/data/ces_cbo.html.

Newman, J. (2009) Systematically Heterogeneous Covariance in Network GEV Models. Proceedings of the 1st International Choice Modeling Conference, Harrogate, April 2009.

Pinjari, A.R., and C.R. Bhat (2010) A Multiple Discrete-Continuous Nested Extreme Value Model: Formulation and Application to Non-Worker Activity Time-use and Timing Behavior on Weekdays. *Transportation Research Part B*, 44(4), 562-583.

Sener, I.N., R.B. Copperman, R.M. Pendyala, and C.R. Bhat (2008) An Analysis of Children's Leisure Activity Engagement: Examining the Day of Week, Location, Physical Activity Level, and Fixity Dimensions. *Transportation*, 35(5), 673-696.

Small, K. (1987) A Discrete Choice Model for Ordered Alternatives. *Econometrica*, 55(2), 409-424.

Spissu, E., A.R. Pinjari, C.R. Bhat, R.M. Pendyala, and K.W. Axhausen (2009) An Analysis of Weekly Out-of-Home Discretionary Activity Participation and Time-Use Behavior. *Transportation*, 36(5), 483-510.

von Haefen, R.H., and D.J. Phaneuf (2005) Kuhn-Tucker Demand System Approaches to Nonmarket Valuation, in R. Scarpa and A.A. Alberini (Eds.), *Applications of Simulation Methods in Environmental and Resource Economics*, Springer.

Vovsha, P. (1997). Cross-nested logit model: An Application to Mode Choice in Tel-Aviv Metropolitan Area. *Transportation Research Record*, 1645, 133-142.

Wen, C.-H., and F.S. Koppelman (2001) The Generalized Nested Logit Model. *Transportation Research Part B*, 35(7), 627-641.

Wu, F., E.A. Valdez, and M. Sherris (2006) Simulating exchangeable multivariate Archimedean copulas and its applications, *Communications in Statistics - Simulation and Computation* 36 (5), 1019–1034.

APPENDIX A

From Equation (8), the multiple discrete-continuous choice probability is as below[11]:

$$P(t_1^*,...,t_M^*,0,...,0)=|J|\int_{\varepsilon_1=-\infty}^{+\infty}\left[\left\{\prod_{i=1}^{M}e^{-\varepsilon_i}Q_M F_{GEV}\right\}\Bigg|_{\substack{\varepsilon_i=V_1-V_i+\varepsilon_1,\\ \forall i=1,2,...,K}}\right]d\varepsilon_1$$

$$=|J|\int_{\varepsilon_1=-\infty}^{+\infty}\left[\left\{\prod_{i=1}^{M}e^{-\varepsilon_i}\begin{pmatrix}\left\{\pm\left(G_1G_2..G_M\right)\right\}+\\ \left\{\pm\left(G_{12}^2G_3..G_M\right)\pm...\pm\left(G_1G_2..G_{(M-1)M}^2\right)\right\}+\\ \left\{\pm\left(G_{123}^3G_4..G_M\right)\pm\left(G_{12}^2G_{34}^2..G_M\right)\pm...\right\}+\\ ...+\\ \left\{\pm\left(G_{123..M-1}^{M-1}G_M\right)\pm...\pm\left(G_1G_{23...M}^{M-1}\right)\right\}+\\ \left\{\pm(G_{123...M}^{M})\right\}\end{pmatrix}F_{GEV}\right\}\Bigg|_{\substack{\varepsilon_i=V_1-V_i+\varepsilon_1,\\ \forall i=1,2,...,K}}\right]d\varepsilon_1 \tag{A1}$$

The above expression can be split into a sum of several integrals as below:

$$P(t_1^*,...,t_M^*,0,...,0)=$$

$$|J|\left[\begin{array}{l}\displaystyle\int_{\varepsilon_1=-\infty}^{+\infty}\prod_{i=1}^{M}e^{-\varepsilon_i}\left(G_1G_2..G_M\right)F_{GEV}\Big|_{\substack{\varepsilon_i=V_1-V_i+\varepsilon_1,\\ \forall i=1,2,...,K}}d\varepsilon_1\pm\\ \displaystyle\int_{\varepsilon_1=-\infty}^{+\infty}\prod_{i=1}^{M}e^{-\varepsilon_i}\left(G_{12}^2G_3..G_M\right)F_{GEV}\Big|_{\substack{\varepsilon_i=V_1-V_i+\varepsilon_1,\\ \forall i=1,2,...,K}}d\varepsilon_1\pm...\pm\int_{\varepsilon_1=-\infty}^{+\infty}\prod_{i=1}^{M}e^{-\varepsilon_i}\left(G_1G_2..G_{(M-1)M}^2\right)F_{GEV}\Big|_{\substack{\varepsilon_i=V_1-V_i+\varepsilon_1,\\ \forall i=1,2,...,K}}d\varepsilon_1\pm\\ \displaystyle\int_{\varepsilon_1=-\infty}^{+\infty}\prod_{i=1}^{M}e^{-\varepsilon_i}\left(G_{123}^3G_3..G_M\right)F_{GEV}\Big|_{\substack{\varepsilon_i=V_1-V_i+\varepsilon_1,\\ \forall i=1,2,...,K}}d\varepsilon_1\pm...\pm\int_{\varepsilon_1=-\infty}^{+\infty}\prod_{i=1}^{M}e^{-\varepsilon_i}\left(G_{12}^2G_{34}^2..G_M\right)F_{GEV}\Big|_{\substack{\varepsilon_i=V_1-V_i+\varepsilon_1,\\ \forall i=1,2,...,K}}d\varepsilon_1\pm...\pm\\ ...+\\ \displaystyle\int_{\varepsilon_1=-\infty}^{+\infty}\prod_{i=1}^{M}e^{-\varepsilon_i}\left(G_{123...(M-1)}^{M-1}G_M\right)F_{GEV}\Big|_{\substack{\varepsilon_i=V_1-V_i+\varepsilon_1,\\ \forall i=1,2,...,K}}d\varepsilon_1\pm...\pm\int_{\varepsilon_1=-\infty}^{+\infty}\prod_{i=1}^{M}e^{-\varepsilon_i}\left(G_1G_{23...M}^{M-1}\right)F_{GEV}\Big|_{\substack{\varepsilon_i=V_1-V_i+\varepsilon_1,\\ \forall i=1,2,...,K}}d\varepsilon_1\pm\\ \displaystyle\int_{\varepsilon_1=-\infty}^{+\infty}\prod_{i=1}^{M}e^{-\varepsilon_i}\left(G_{123...M}^{M}\right)F_{GEV}\Big|_{\substack{\varepsilon_i=V_1-V_i+\varepsilon_1,\\ \forall i=1,2,...,K}}d\varepsilon_1\end{array}\right] \tag{A2}$$

[11]In this equation, Q_M is expanded and arranged in such a way that each row of the terms in the expansion of Q_M is a sum of several terms, with each term in the row having the same number of partial derivatives. That is, the term in the first row, $\left(G_1G_2..G_M\right)$, is a product of *M* number of partial derivatives. Similarly, each term in the second row is a product of *M-1* number of partial derivatives, each term in the third row is a product of *M-2* number of partial derivatives, and so on. This arrangement will be useful toward the end of the proof.

To solve each integral in (A2), consider: $G^n_{123...n}\Big|_{\substack{\varepsilon_i=V_1-V_i+\varepsilon_1 \\ \forall i=1,2,...,K}} = \left.\dfrac{\partial^n G(e^{-\varepsilon_1},..,e^{-\varepsilon_n})}{\partial e^{-\varepsilon_1}..\partial e^{-\varepsilon_n}}\right|_{\substack{\varepsilon_i=V_1-V_i+\varepsilon_1 \\ \forall i=1,2,..,K}}$, which is an n^{th} order partial derivative of G evaluated at $\varepsilon_i = V_1 - V_i + \varepsilon_1$ $(\forall i = 1,2,...,K)$. Using mathematical induction (details are available with the author), one can show that:

$$G^n_{123...n}\Big|_{\substack{\varepsilon_i=V_1-V_i+\varepsilon_1 \\ \forall i=1,2,...,K}} = \left(e^{-\varepsilon_1}e^{-V_1}\right)^{-(n-1)} \frac{\partial^n G(e^{V_1},...,e^{V_K})}{\partial e^{V_1}...\partial e^{V_n}} \tag{A3}$$

For notational ease, denote $G(e^{V_1},...,e^{V_n})$ as H, and $\dfrac{\partial^n G(e^{V_1},...,e^{V_n})}{\partial e^{V_1}...\partial e^{V_n}}$ as $H^n_{123...n}$, to rewrite the above Equation as:

$$G^n_{123...n}\Big|_{\substack{\varepsilon_i=V_1-V_i+\varepsilon_1 \\ \forall i=1,2,...,K}} = \left(e^{-\varepsilon_1}e^{-V_1}\right)^{-(n-1)} H^n_{123...n} \tag{A4}$$

Now, from equation (A2), consider an arbitrarily chosen integral, and express it in a general form as:

$$I = \int_{\varepsilon_1=-\infty}^{+\infty} \left.\prod_{i=1}^{M} e^{-\varepsilon_i}\left(\text{Prod}(\tilde{G})\right)F_{GEV}\right|_{\varepsilon_i=V_1-V_i+\varepsilon_1} d\varepsilon_1 = \int_{\varepsilon_1=-\infty}^{+\infty} \left(\prod_{i=1}^{M} e^{-\varepsilon_i}\Big|_{\varepsilon_i=V_1-V_i+\varepsilon_1} \left(\text{Prod}(\tilde{G})\right)\Big|_{\varepsilon_i=V_1-V_i+\varepsilon_1} F_{GEV}\Big|_{\varepsilon_i=V_1-V_i+\varepsilon_1}\right) d\varepsilon_1 \tag{A5}$$

In this integral, $\left(\text{Prod}(\tilde{G})\right)$ represents the product of partial derivatives of G of various orders. Without loss of generality, let there be N partial derivatives in $\left(\text{Prod}(\tilde{G})\right)$ and let the order of these partial derivatives be $M_1, M_2,..., M_N$ (the reader will note that $M_1 + M_2 + ... + M_N = M$). Based on this information, the integral I can be rewritten as:

$$I = \int_{\varepsilon_1=-\infty}^{+\infty} \left(\left(e^{-\varepsilon_1}e^{-V_1}\right)^M \left(\prod_{i=1}^{M} e^{V_i}\right)\left(e^{-\varepsilon_1}e^{-V_1}\right)^{-\sum_{j=1}^{N}(M_j-1)} \left(\text{Prod}(\tilde{H})\right)F_{GEV}\Big|_{\varepsilon_i=V_1-V_i+\varepsilon_1}\right) d\varepsilon_1 \tag{A6}$$

where $\left(\text{Prod}(\tilde{H})\right)$ represents the product of partial derivatives of H of various orders.

Further, based on Equation (5) and the properties of the homogeneous function G, by expanding F_{GEV}, the integral can be rewritten as:

$$I = \int_{\varepsilon_1=-\infty}^{+\infty} \left\{\left(e^{-\varepsilon_1}e^{-V_1}\right)^M \left(\prod_{i=1}^{M} e^{V_i}\right)\left(e^{-\varepsilon_1}e^{-V_1}\right)^{-\sum_{j=1}^{N}(M_j-1)} \left(\text{Prod}(\tilde{H})\right)\exp\left[-e^{-\varepsilon_1}e^{-V_1}H\left(e^{-V_1},e^{-V_2},...,e^{-V_K}\right)\right]\right\} d\varepsilon_1$$

$$= e^{-V_1\left(M-\sum_{j=1}^{N}(M_j-1)\right)}\left(\prod_{i=1}^{M}e^{V_i}\right)\left(\text{Prod}(\tilde{H})\right)\int_{\varepsilon_1=-\infty}^{+\infty}\left\{e^{-\varepsilon_1\left(M-\sum_{j=1}^{N}(M_j-1)\right)}\exp\left[-e^{-\varepsilon_1}e^{-V_1}H\left(e^{-V_1},e^{-V_2},...,e^{-V_K}\right)\right]\right\}d\varepsilon_1$$

(A7)

The above integral, after applying integration by parts in a repeated fashion, results in the following closed-form expression (details are available with the author):

$$I = e^{-V_1\left(M-\sum_{j=1}^{N}(M_j-1)\right)}\left(\prod_{i=1}^{M}e^{V_i}\right)\left(\text{Prod}(\tilde{H})\right)\frac{\left(M-\sum_{j=1}^{N}(M_j-1)-1\right)!}{\left[e^{-V_1}H\left(e^{-V_1},e^{-V_2},...,e^{-V_K}\right)\right]^N} \tag{A8}$$

Since $\sum_{j=1}^{N}M_j = M$, the integral can be further simplified as:

$$I = \frac{\left(\prod_{i=1}^{M}e^{V_i}\right)\left(\text{Prod}(\tilde{H})\right)\left(N-1\right)!}{\left[H\left(e^{-V_1},e^{-V_2},...,e^{-V_K}\right)\right]^N} \tag{A9}$$

Using the above form of analytical expression for the integrals, the probability expression of Equation (A2) can be expressed in a closed-form as below:

$$P(t_1^*,...,t_M^*,0,...,0) = |J|\prod_{i=1}^{M}e^{V_i}\times\begin{bmatrix}(M-1)!\left\{\pm\frac{H_1H_2..H_M}{H^M}\right\}+\\ (M-2)!\left\{\pm\frac{H_{12}^2H_3..H_M}{H^{M-1}}\pm\frac{H_1H_{34}^2..H_M}{H^{M-1}}\pm...\pm\frac{H_1H_2..H_{(M-1)M}^2}{H^{M-1}}\right\}+\\ (M-3)!\left\{\pm\frac{H_{123}^3H_3..H_M}{H^{M-2}}\pm...\pm\frac{H_{12}^2H_{34}^2..H_M}{H^{M-2}}\pm...\right\}+\\ ...\\ 1!\left\{\pm\frac{H_{123...M-1}^{M-1}H_M}{H^2}\pm...\pm\frac{H_1H_{234...M}^{M-1}}{H^2}\right\}+\\ 0!\left\{\pm\frac{H_{123...M}^M}{H^1}\right\}\end{bmatrix}$$

(A10)

where $H_{123...n}^n = \frac{\partial^n H(e^{V_1},..,e^{V_K})}{\partial e^{V_1}..\partial e^{V_n}}$, $H_i = \frac{\partial H(e^{V_1},..,e^{V_K})}{\partial e^{V_i}}$, and all other terms are defined in a similar fashion.

Table 1. Model Estimation Results

	MDCEV		MDCNEV		Cross-Nested Model with No Heterogeneity		Heterogeneous Cross-nested Model	
Baseline Utility Parameters	Parameter	t-stat	Parameter	t-stat	Parameter	t-stat	Parameter	t-stat
Baseline Constants								
Vehicle purchase	-6.955	-67.12	-6.487	-74.17	-6.757	-70.12	-6.861	-65.19
Gasoline/oil	-1.255	-10.21	-1.913	-18.71	-1.675	-14.59	-1.691	-14.60
Vehicle insurance	-3.333	-57.10	-3.800	-53.06	-3.671	-51.60	-3.668	-51.57
Vehicle maintenance	-2.751	-42.00	-3.487	-44.88	-3.246	-41.45	-3.243	-41.42
Air travel	-6.058	-72.53	-6.035	-73.85	-6.050	-73.73	-6.052	-73.80
Public transit	-6.092	-49.49	-6.082	-48.40	-6.063	-48.36	-6.063	-48.35
No of workers in household								
Vehicle purchase	0.160	4.19	0.117	3.46	0.108	2.90	0.097	2.63
Gasoline/Oil	0.190	5.46	0.183	5.76	0.197	6.22	0.199	6.41
Vehicle Insurance	0.066	2.50	0.066	2.28	0.072	2.43	0.071	2.42
Vehicle Maintenance	0.157	6.32	0.140	4.94	0.152	5.34	0.150	5.31
Annual HH income 30-70K								
Vehicle purchase	0.869	8.49	0.607	9.33	0.795	8.96	0.929	8.85
Gasoline/oil	-0.379	-4.89	-0.334	-8.12	-0.343	-7.18	-0.334	-7.01
Air travel	0.871	9.68	0.875	10.03	0.905	10.20	0.907	10.23
Annual HH income >70K								
Vehicle purchase	0.873	7.26	0.684	8.92	0.765	6.88	0.967	7.88
Gasoline/oil	-0.902	-8.27	-0.733	-12.34	-0.745	-11.01	-0.735	-10.67
Vehicle insurance	-0.343	-4.79	-0.206	-4.57	-0.234	-4.53	-0.241	-4.63
Air travel	1.434	12.91	1.441	13.83	1.456	13.71	1.460	13.74
Number of vehicles in household								
Vehicle purchase	0.208	9.94	0.211	11.52	0.195	9.46	0.188	9.14
Gasoline/oil	0.123	4.92	0.156	7.85	0.139	6.89	0.140	7.23
Vehicle insurance	0.095	5.45	0.113	6.39	0.093	4.92	0.093	4.98
Vehicle maintenance	0.099	5.19	0.115	6.24	0.096	4.94	0.096	5.00
Zero-car household (dummy variable)								
Gasoline/oil	-3.109	-25.23	-2.355	-25.65	-2.895	-24.81	-2.900	-24.48
Vehicle insurance	-3.246	-16.65	-2.667	-21.86	-3.571	-19.82	-3.580	-19.68
Vehicle maintenance	-2.474	-19.17	-1.932	-21.54	-2.601	-19.08	-2.606	-19.04
Public Transit	1.099	10.99	1.315	13.47	1.126	11.38	1.120	11.31
Non-Caucasian HH – Public transit	0.288	4.01	0.256	3.32	0.265	3.48	0.264	3.47
Urban location – Public transit	0.543	4.60	0.550	4.52	0.552	4.54	0.552	4.54
North East Region – Public transit	0.711	10.07	0.740	9.74	0.734	9.76	0.734	9.77
Western Region – Public transit	0.608	9.41	0.604	8.70	0.606	8.83	0.605	8.83

Table 1 (Continued.) Model Estimation Results

	MDCEV		MDCNEV		Cross-Nested Model with No Heterogeneity		Heterogeneous Cross-nested Model	
Translation (γ_k) Parameters	Parameter	t-stat	Parameter	t-stat	Parameter	t-stat	Parameter	t-stat
Vehicle purchase	19.674	11.98	37.528	12.18	22.712	11.60	22.842	11.70
Gasoline/oil	0.098	8.58	0.149	10.64	0.120	9.61	0.120	9.53
Vehicle insurance	0.515	18.90	0.854	21.23	0.682	19.42	0.681	19.38
Vehicle maintenance	0.229	17.54	0.528	18.48	0.371	16.67	0.370	16.64
Air travel	0.631	15.13	0.598	15.04	0.605	15.04	0.605	15.04
Public transit	0.204	19.19	0.193	18.98	0.195	19.01	0.195	19.01
Nesting Parameters*								
θ {vehicle purchase, gasoline, vehicle insurance, maintenance}			0.628	34.99				
θ {gasoline, vehicle insurance, vehicle maintenance}					0.701	22.00	0.704	21.71
θ {vehicle purchase, gasoline}					0.409	20.32	0.364	24.96
Allocation Parameters								
Allocation parameter for the {gasoline, vehicle insurance, maintenance} nest**					0.873	8.08		
Allocation parameter for the {vehicle purchase, gasoline} nest**					0.127	8.08		
Logit function of the allocation parameter for the {gasoline, vehicle insurance, maintenance} nest ***								
Constant							2.568	13.44
Annual HH income 30-70K							-0.635	-3.05
Annual HH income >70K							-1.071	-4.53
Log-likelihood at constants	-42265		-42265		-42265		-42265	
Log-likelihood at convergence	-40703		-39111		-38786		-38778	

* The reported t-statistics for the nesting parameters are against a value of 1.

** The reported t-statistic for the allocation parameter corresponding to the {gasoline, vehicle insurance, maintenance} nest is against a value of 1, while that for the {vehicle purchase, gasoline} nest is against a value of 0.

***Based on the parameter estimates of this logit function, the allocation parameters for low-income households can be computed as: $\exp(2.568)/\left(1+\exp(2.568)\right) = 0.93$ for the {gasoline, vehicle insurance, maintenance} nest and $1/\left(1+\exp(2.568)\right) = 0.07$ for the {vehicle purchase, gasoline} nest. The allocation parameters for high-income households can be computed as: $\exp(2.568-1.071)/\left(1+\exp(2.568-1.071)\right) = 0.82$ for the {gasoline, vehicle insurance, maintenance} nest and $1/\left(1+\exp(2.568-1.071)\right) = 0.18$ for the {vehicle purchase, gasoline} nest.

Table 2. Predictive Log-Likelihood-based Measures of Fit in the Validation Sample

		Predictive Log-Likelihood Function (PLLF)			
	No. of Households	MDCEV	MDCNEV	Cross-Nested Model with No Heterogeneity	Heterogeneous Cross-nested Model
All Households	101	-1037.96	-1001.59	-994.51	-993.86
Low-income HHs	26	-245.30	-235.59	-234.75	-234.39
Medium-income HHs	51	-548.38	-529.37	-524.39	-524.12
High-income HHs	24	-244.28	-236.62	-235.37	-235.34