**An Innovative Forecasting Procedure for the MDCEV Model**

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

and

**Chandra Bhat**
The University of Texas at Austin
Dept of Civil, Architectural & Environmental Engineering
1 University Station C1761, Austin, TX 78712-0278
Tel: 512-471-4535, Fax: 512-475-8744
E-mail: bhat@mail.utexas.edu

*Submitted for presentation at the 3rd Innovations in Travel Modeling (ITM) Conference, Arizona State University, Tempe, Arizona.*

Unabridged version of the paper is available at the following link:
http://www.ce.utexas.edu/prof/bhat/ABSTRACTS/MDCEV_Forecasting_paper_25July09.pdf

# An Innovative Forecasting Procedure for the MDCEV Model

## 1. INTRODUCTION

Multiple discrete-continuous (MDC) choice situations are being increasingly recognized and modeled in the recent travel modeling literature. Example applications include, but are not limited to, individual activity participation and time-use studies (Bhat, 2005; Habib and Miller, 2009), household vehicle ownership and usage forecasting (Ahn *et al.*, 2007; Bhat *et al.*, 2008), and household travel expenditure analyses (Rajagopalan and Srinivasan, 2008). Further, a variety of modeling approaches have been used to analyze MDC choices. Among the available approaches, the recently developed multiple discrete-continuous extreme value (MDCEV) model structure has gained particular attention. Specifically, due to its closed form probability expressions and other elegant properties (Bhat, 2008), the MDCEV model has been used in several of the above-identified and other empirical applications.

Despite several empirical applications, a simple and practically feasible forecasting procedure has not yet been developed for the MDCEV model system. This has severely limited the applicability of the MDCEV model for practical travel forecasting and policy analysis purposes. Hence, a brief discussion is provided below to highlight the nature of the MDCEV forecasting problem and to outline the objective of this paper.

The MDCEV model is based on a resource allocation formulation. Specifically, it is assumed that consumers operate with a finite amount of available resources, such as time or money. The decision-making mechanism is assumed to be driven by an allocation of the limited amount of resources to consume various goods/alternatives in such a way as to maximize the utility of consumption. Further, a stochastic utility framework is used to recognize the analyst's lack of awareness of all factors affecting consumer decisions. In addition, a non-linear utility function is employed to incorporate important features of consumer choice making, including: (1) the diminishing nature of marginal utility with increasing consumption, and (2) the possibility of consuming multiple goods/alternatives as opposed to a single good/alternative. To summarize, the MDCEV model is based on a stochastic, constrained, non-linear utility optimization formulation. Specifically, the model estimation is based on the stochastic Kuhn-Tucker (KT) first order conditions for the optimization problem identified above. An assumption that stochasticity is type-I extreme value distributed leads to closed form consumption probability expressions and facilitates a straightforward maximum likelihood estimation of the model (Bhat, 2008).

Given the estimated model parameters and a budget amount for each individual, any forecasting exercise involves solving the stochastic, constrained, non-linear utility maximization problem for optimal consumption quantities. Unfortunately, there is no straight-forward analytical solution to this problem; a combination of simulation (to mimic stochasticity) and optimization (to solve the constrained non-linear optimization problem) methods needs to be employed. The analyst must carry out constrained non-linear optimization to obtain the consumption forecasts at each simulated value of stochasticity (or unobserved heterogeneity). Such conditional (on unobserved heterogeneity) consumption forecasts evaluated over the entire (simulated)distribution of unobserved heterogeneity are used to derive the distributions of unconditional consumption forecasts.

To solve the conditional constrained non-linear optimization problem, the forecasting procedures used in the literature so far use either enumerative or iterative optimization methods, which are saddled with large computation times and potential convergence issues (for iterative procedures). Hence, the objective of this paper is to develop an efficient, non-iterative forecasting algorithm for the MDCEV model. The algorithm builds on simple, yet insightful, analytical explorations that shed new light on the properties of the MDCEV model. For specific forms of utility functions, the algorithm becomes non-iterative and significantly reduces computational time. Preliminary application results, as discussed Section 3, indicate a significant computational efficiency of the proposed algorithm. For example, to forecast the expenditures of 4000 households in 7 transportation-related expenditure alternatives, for 500 sets of error term draws for each household, the proposed algorithm takes less than 2 minutes. On the other hand, the iterative forecasting routine would take around 2 days to do so.

Next section discusses the structure and some new properties of the MDCEV model. Section 3 presents the new forecasting algorithm and application results. Section 4 concludes the paper.

## 2 THE MDCEV MODEL: STRUCTURE AND PROPERTIES

### 2.1 Model Structure (drawn from Bhat, 2008)

Consider the following additively separable utility function as in Bhat (2008):

$$U(\boldsymbol{t}) = \frac{1}{\alpha_1}\psi_1 t_1^{\alpha_1} + \sum_{k=2}^{K} \frac{\gamma_k}{\alpha_k}\psi_k \left\{ \left( \frac{t_k}{\gamma_k} + 1 \right)^{\alpha_k} - 1 \right\}; \;\; \psi_k > 0,\; 0 < \alpha_k \le 1,\; \gamma_k > 0 \tag{1}$$

In the above expression, $U(\boldsymbol{t})$ is the total utility accrued from consuming $\boldsymbol{t}$ (a $K$x1 vector with non-negative consumption quantities $t_k$; $k$ = 1,2,…,$K$) amount of the $K$ alternatives available to the decision maker. The $\psi_k$ terms ($k$ = 1,2,…,$K$), called as baseline utility parameters, represent the marginal utility for alternative $k$ at the point of zero consumption for that alternative. Through the $\psi_k$ terms, the impact of observed and unobserved alternative attributes, decision-maker attributes, and the choice environment attributes may be introduced as $\psi_k = \exp(\beta' z_k + \varepsilon_k)$, where $z_k$ contains the observed attributes and $\varepsilon_k$ captures the unobserved factors. The $\alpha_k$ terms ($k$ = 1,2,…,$K$), labeled as satiation parameters $(0 < \alpha_k \le 1)$, capture satiation effects by reducing the marginal utility accrued from each unit of additional consumption of alternative $k$. The $\gamma_k$ terms ($k$ = 2,3,…,$K$), labeled as translation parameters, play a similar role of satiation as that of $\alpha_k$ terms, and an additional role of translating the indifference curves associated with the utility function to allow corner solutions. Note that there is no $\gamma_k$ term for the first alternative for it is assumed to be an essential Hicksian composite good (or outside good) that is always consumed (hence no need for corner solution). Finally, the consumption-based utility function in (1) can be expressed in terms of expenditures ($e_k$) and prices ($p_k$) as:

$$U(\boldsymbol{e}) = \frac{1}{\alpha_1}\psi_1 \left( \frac{e_1}{p_1} \right)^{\alpha_1} + \sum_{k=2}^{K} \frac{\gamma_k}{\alpha_k}\psi_k \left\{ \left( \frac{e_k}{\gamma_k p_k} + 1 \right)^{\alpha_k} - 1 \right\}, \text{ where } \frac{e_k}{p_k} = t_k \tag{2}$$

From the analyst's perspective, consumers maximize the random utility given by Equation (2) subject to a linear budget constraint and non-negativity constraints on $t_k$:

$$\sum_{k=1}^{K} p_k t_k = E \text{ (where } E \text{ is the total budget) and } t_k \ge 0 \;\forall\; k \;(k = 1,2,...,K) \tag{3}$$

The optimal consumptions (or expenditure allocations) can be found by forming the Lagrangian and applying the Kuhn-Tucker (KT) conditions. The Lagrangian function for the problem is:

$$\mathscr{L} = \frac{1}{\alpha_1}\psi_1\left(\frac{e_1}{p_1}\right)^{\alpha_1} + \sum_{k=2}^{K}\frac{\gamma_k}{\alpha_k}\psi_k\left\{\left(\frac{e_k}{\gamma_k p_k}+1\right)^{\alpha_k} - 1\right\} - \lambda\left[\sum_{k=1}^{K} e_k - E\right],$$

where $\lambda$ is the Lagrangian multiplier associated with the budget constraint. The KT first-order conditions for the optimal expenditure allocations $(e_k^*; k=1,2,...,K)$ are:

$$\frac{\psi_1}{p_1}\left(\frac{e_1^*}{p_1}\right)^{\alpha_1-1} - \lambda = 0, \text{ since } e_1^* > 0,$$

$$\frac{\psi_k}{p_k}\left(\frac{e_k^*}{\gamma_k p_k}+1\right)^{\alpha_k-1} - \lambda = 0, \text{ if } e_k^* > 0, \ (k = 2,\dots, K) \tag{4}$$

$$\frac{\psi_k}{p_k}\left(\frac{e_k^*}{\gamma_k p_k}+1\right)^{\alpha_k-1} - \lambda < 0, \text{ if } e_k^* = 0, \ (k = 2,\dots, K)$$

Next, using these same stochastic KT conditions, we derive a few properties of the MDCEV model that can be exploited to develop a highly efficient forecasting algorithm.

## 2.2 Model Properties

***Property 1:*** *The price-normalized baseline utility of a chosen good is always greater than that of a good that is not chosen.*

$$\left(\frac{\psi_i}{p_i}\right) > \left(\frac{\psi_j}{p_j}\right) \text{ if 'i' is a chosen good and 'j' is not a chosen good.} \tag{5}$$

***Proof:*** The KT conditions in (4) can be rewritten as:

$$\frac{\psi_1}{p_1}\left(\frac{e_1^*}{p_1}\right)^{\alpha_1-1} = \lambda,$$

$$\frac{\psi_k}{p_k}\left(\frac{e_k^*}{\gamma_k p_k}+1\right)^{\alpha_k-1} = \lambda, \text{ if } e_k^* > 0, \ (k = 2,\dots, K) \ (\textit{i.e.}, \text{ for all chosen goods}) \tag{6}$$

$$\frac{\psi_k}{p_k} < \lambda, \text{ if } e_k^* = 0, \ (k = 2,\dots, K) \ (\textit{i.e.}, \text{ for all goods that are not chosen})$$

The above KT conditions can further be rewritten as:

$$\left(\frac{\psi_k}{p_k}\right) = \left(\frac{e_k^*}{\gamma_k p_k}+1\right)^{1-\alpha_k}\left(\frac{\psi_1}{p_1}\right)\left(\frac{e_1^*}{p_1}\right)^{\alpha_1-1}, \quad \text{if } e_k^* > 0, \ (k = 2,3,...,K) \tag{7}$$

$$\left(\frac{\psi_k}{p_k}\right) < \left(\frac{\psi_1}{p_1}\right)\left(\frac{e_1^*}{p_1}\right)^{\alpha_1-1}, \qquad \text{if } e_k^* = 0, \ (k = 2,3,...,K)$$

Now, consider two alternatives *'i'* and *'j'*, of which *'i'* is chosen and *'j'* is not chosen by a consumer. For that consumer, the above KT conditions for alternatives *'i'* and *'j'* can be written as:

$$\left(\frac{\psi_i}{p_i}\right) = \left(\frac{e_i^*}{\gamma_i p_i} + 1\right)^{1-\alpha_i} \left(\frac{\psi_1}{p_1}\right)\left(\frac{e_1^*}{p_1}\right)^{\alpha_1-1}, \ \text{and} \tag{8}$$

$$\left(\frac{\psi_j}{p_j}\right) < \left(\frac{\psi_1}{p_1}\right)\left(\frac{e_1^*}{p_1}\right)^{\alpha_1-1}$$

Further, since $\left(\frac{e_j^*}{\gamma_j p_j} + 1\right)^{1-\alpha_j}$ is always greater than 1, one can write the following inequality:

$$\left(\frac{\psi_j}{p_j}\right) < \left(\frac{\psi_1}{p_1}\right)\left(\frac{e_1^*}{p_1}\right)^{\alpha_1-1} < \left(\frac{e_i^*}{\gamma_i p_i} + 1\right)^{1-\alpha_i} \left(\frac{\psi_1}{p_1}\right)\left(\frac{e_1^*}{p_1}\right)^{\alpha_1-1} \tag{9}$$

As one can observe, the third term in the above inequality is nothing but $\left(\frac{\psi_i}{p_i}\right)$. Thus, by the transitive property of inequality of real numbers, the above inequality implies a fundamental property of the MDCEV model that $\left(\frac{\psi_i}{p_i}\right) > \left(\frac{\psi_j}{p_j}\right)$. *In words, the price-normalized baseline utility of a chosen good is always greater than that of a good that is not chosen.*

***Corollary 1.1:*** *It naturally follows from the above property that when all the K alternatives available to a consumer are arranged in the descending order of their price-normalized baseline utility values (with the outside good being first in the order), and if it is known that the number of chosen alternatives is M, then the first M alternatives in this arrangement are the chosen alternatives.*

***Property 2:*** *When all the satiation parameters* $(\alpha_k)$ *are equal, and if the corner solutions are known (i.e., if the chosen and not-chosen alternatives are known), the optimal consumptions of the chosen goods can be expressed in an analytical form.*

***Proof:*** Using the first and second KT conditions in (6), and assuming without loss of generality that the first *M* goods are chosen, one can express the optimal consumptions as:

$$\frac{e_1^*}{p_1}=\left(\lambda\frac{p_1}{\psi_1}\right)^{\frac{1}{\alpha_1-1}}, \text{ and}$$

$$\frac{e_k^*}{p_k}=\left\{\left(\lambda\frac{p_k}{\psi_k}\right)^{\frac{1}{\alpha_k-1}}-1\right\}\gamma_k;\ \forall k=(2,3,...,M) \tag{11}$$

Using these expressions, the budget constraint in (3) can be written as:

$$p_1\left(\lambda\frac{p_1}{\psi_1}\right)^{\frac{1}{\alpha_1-1}}+\sum_{k=2}^{M}p_k\left\{\left(\lambda\frac{p_k}{\psi_k}\right)^{\frac{1}{\alpha_k-1}}-1\right\}\gamma_k=E$$

From the above equation, and assuming that all satiation $(\alpha_k)$ parameters as equal to $\alpha$, the Lagrange multiplier $\lambda$ can be expressed analytically as:

$$\lambda=\left(\frac{E+\sum_{k=2}^{M}p_k\gamma_k}{p_1\left(\frac{\psi_1}{p_1}\right)^{\frac{1}{1-\alpha}}+\sum_{k=2}^{M}p_k\gamma_k\left(\frac{\psi_k}{p_k}\right)^{\frac{1}{1-\alpha}}}\right)^{\alpha-1} \tag{12}$$

The above expression for $\lambda$ can be substituted back into the expressions in (11) to obtain the following analytical expressions for optimal consumptions:

$$\frac{e_1^*}{p_1}=\frac{\left(\frac{\psi_1}{p_1}\right)^{\frac{1}{1-\alpha}}\left(E+\sum_{k=2}^{M}p_k\gamma_k\right)}{p_1\left(\frac{\psi_1}{p_1}\right)^{\frac{1}{1-\alpha}}+\sum_{k=2}^{M}p_k\gamma_k\left(\frac{\psi_k}{p_k}\right)^{\frac{1}{1-\alpha}}} \tag{13}$$

$$\frac{e_k^*}{p_k}=\left(\frac{\left(\frac{\psi_k}{p_k}\right)^{\frac{1}{1-\alpha}}\left(E+\sum_{k=2}^{M}p_k\gamma_k\right)}{p_1\left(\frac{\psi_1}{p_1}\right)^{\frac{1}{1-\alpha}}+\sum_{k=2}^{M}p_k\gamma_k\left(\frac{\psi_k}{p_k}\right)^{\frac{1}{1-\alpha}}}-1\right)\gamma_k;\ \forall k=(2,3,...,M)\,. \tag{14}$$

## 3 AN EFFICIENT FORECASTING ALGORITHM

In this section, using the properties identified in the preceding section, we develop an efficient, non-iterative forecasting algorithm for the MDCEV model with the following utility functional form:

$$U(\boldsymbol{e}) = \frac{1}{\alpha}\exp(\beta' z_1 + \varepsilon_1)\left(\frac{e_1}{p_1}\right)^{\alpha} + \sum_{k=2}^{K} \frac{\gamma_k}{\alpha}\exp(\beta' z_k + \varepsilon_k)\left\{\left(\frac{e_k}{\gamma_k p_k} + 1\right)^{\alpha} - 1\right\}, \tag{15}$$

### 3.1 Outline of the Forecasting Algorithm

***Step 0:*** Assume that only the outside good is chosen and let the number of chosen goods $M = 1$.

***Step 1:*** Given the input data ($z_k$, $p_k$), model parameters ($\beta, \gamma_k, \alpha$) and the simulated error term ($\varepsilon_k$) draws, compute the price-normalized baseline utility values $\left(\exp(\beta' z_k + \varepsilon_k)/p_k\right)$ for all alternatives. Arrange all the $K$ alternatives available to the consumer in the descending order of their price-normalized baseline utility values (with the outside good in the first place). Go to step 2.

***Step 2:*** Compute the value of $\lambda$ using equation (12). Go to step 3.

***Step 3:*** If ($\lambda$ > the price-normalized baseline utility of the alternative in position $M+1$)

Compute the optimal consumptions of the first $M$ alternatives in the above descending order using equations in (13) and (14). Set the consumptions of other alternatives as zero and stop.

Else, go to step 4.

***Step 4:*** $M = M+1$.

If ($M = K$)

Compute the optimal consumptions using equations in (13) and (14) and stop.

Else, go to step 2.

The above-outlined algorithm can be applied a large number of times with different simulated values of the $\varepsilon_k$ terms to sufficiently cover the simulated distribution of unobserved heterogeneity and obtain the distributions of the consumption forecasts.

### 3.2 Intuitive Interpretation of the Algorithm

The proposed algorithm builds on the insight from corollary 1.1 that if the number of chosen alternatives is known, one can easily identify the chosen alternatives by arranging the price-normalized baseline utility values in a descending order. Subsequently, one can compute the optimal consumptions of the chosen alternatives using Equations (13) and (14). The only issue, however, is that the number of chosen alternatives is unknown *apriori*. To find this out, the algorithm begins with an assumption that only one alternative (*i.e.,* the outside good) is chosen and verifies this assumption by verifying the KT conditions (*i.e.,* the condition in Step 3) for other alternatives. If the KT conditions are met, the algorithm stops. Else, at least the next alternative (in the order of the price-normalized baseline utilities) has to be among the chosen alternatives. Thus, the KT conditions (*i.e.*, the condition in step 3) are verified again by assuming that the next alternative is among the chosen alternatives. These basic steps are repeated until either the KT conditions (*i.e.*, the condition in step 3) are met or the assumed number of chosen alternatives reaches the maximum number ($K$).

The algorithm involves enumeration of the choice baskets in the most efficient fashion. In fact, the algorithm begins with identifying a single alternative (outside good) that may be chosen. If the KT conditions are not met for this choice basket, the algorithm identifies a two-alternative choice basket and so on, till the number of chosen alternatives is determined. Thus, the number of times the algorithm enumerates choice baskets is equal to the number of chosen alternatives in the optimal consumption portfolio, which is at most equal to (but many times less than) the total number of available alternatives ($K$).

Another feature of the algorithm is that it is non-iterative in nature, which makes it highly efficient compared to other, iterative approaches. Further, coding the algorithm using vector and matrix notation in matrix programming languages significantly reduces the computational burden even with large number of choice alternatives and observations.

In summary, the proposed algorithm is simple and efficient. The only disadvantage of this algorithm (in its current form) is it is designed to be used with the $\gamma$-profile utility specification (i.e., all $\alpha_k$ parameters are constrained to be equal). However, as indicated in Bhat (2008), both $\gamma_k$ and $\alpha_k$ parameters serve the role of allowing differential satiation effects across the choice alternatives. Due to such overlapping roles, "*for a given $\psi_k$ value, it is possible to closely approximate a sub-utility function based on a combination of $\gamma_k$ and $\alpha_k$ values with a sub-utility function solely based on $\gamma_k$ or $\alpha_k$ values*" (Bhat, 2008). Hence, and given the ease of forecasting with the proposed algorithm, we suggest an estimation of the $\gamma$-profile utility function. Nevertheless, the insights obtained from the properties discussed in the paper can be used to design an efficient (albeit iterative) algorithm for cases when $\alpha_k$ parameters vary across alternatives.

### 3.3 Application Results

Limited experiments were conducted to assess the performance of the algorithm. Specifically, the performance of the proposed algorithm and that of a traditionally used iterative optimization routine were compared. Empirical data on household transportation expenditures, obtained from the 2002 Consumer Expenditure Survey conducted by the Bureau of Labor Statistics, was used for the experiments. This data was used to estimate an MDCEV model for household expenditures in six transportation categories (or alternatives), including: (1) Vehicle purchases, (2) Gasoline and motor oil, (3) Vehicle insurance, (4) Vehicle maintenance, (5) Air travel, and (6) Public transportation, and a seventh, outside category that includes all other expenditures of a household. An MDCEV model with these seven expenditure alternatives was estimated using data from 4000 households. These model parameters and the data of the 4000 households were used for subsequent experiments.

The proposed algorithm was coded and executed in Gauss matrix programming language. In addition, for comparison purposes, the iterative constrained optimization routines of the Constrained Maximum Likelihood (CML) module of Gauss were also used for forecasting the household expenditures in the empirical data. The CML module uses a sequential programming method for non-linear optimization, in which the optimal consumption values are approximated iteratively using the first and second gradients of the Lagrangian function. To recognize stochasticity, both the forecasting procedures were run repeatedly using several sets of Halton draws of the $\varepsilon_k$ terms.

To forecast the expenditure patterns of 4000 households in the seven expenditure alternatives identified above, the proposed algorithm takes less than 2 minutes with 500 sets of error terms draws for each household. On the other hand, the optimization routine in the (CML) module of Gauss takes at least 6 minutes to compute the expenditure patterns of the same 4000 households for just one set of error term draws for each household. A linear extrapolation to 500 sets of error term draws implies a rather large computation time of more than 2 days. These run time differences (2 minutes versus 2 days) clearly highlight the efficiency of the proposed algorithm. Even in empirical contexts with a large number of alternatives, observations, or error term draws, the computation time of the proposed algorithm will not increase in a linear fashion. This is because, since most of the algorithm can be executed using matrix operations, the computations are performed simultaneously for all observations and over all error term draws. This contributes to the significant computational efficiency of the proposed algorithm. On the other hand, the iterative forecasting procedure does not exhibit such computational efficiencies. In addition to being computationally efficient, in certain (although a small number of) instances, the iterative procedure ran into convergence problems, and either yielded suboptimal solutions or did not even converge. On the other hand, the proposed algorithm did not run into any convergence issues thanks to its non-iterative nature.

## 4 SUMMARY AND CONCLUSIONS

This paper proposes a simple and efficient forecasting algorithm for the MDCEV model. The algorithm builds on simple, yet insightful, analytical explorations with the Kuhn-Tucker conditions of optimality that shed new light on the properties of the model. For specific, but reasonably general, forms of utility functions, the algorithm circumvents the need to carry out any iterative optimization-based forecasting procedures that have hitherto been used. The non-iterative nature of the algorithm contributes significantly to its computational efficiency.

As the proposed algorithm makes it easier to perform forecasting and policy analysis with MDCEV, it is hoped that these types of models will soon be utilized for practical travel forecasting and policy analysis purposes. In subsequent work, the proposed algorithm will be applied to use an MDCEV model of activity time-use for activity time-use forecasting (using data from the San Francisco Bay Area) and the results will be presented at this conference.

## REFERENCES


Ahn, J., Jeong, G., and K. Yeonbae (2007) A forecast of household ownership and use of alternative fuel vehicles: A multiple discrete-continuous choice approach. *Energy Economics*, 30(5), 2091-2104.

Bhat, C.R. (2008) The Multiple Discrete-Continuous Extreme Value (MDCEV) Model: Role of Utility Function Parameters, Identification Considerations, and Model Extensions. *Transportation Research-B*, 42(3), 274-303.

Bhat, C.R., S. Sen, and N. Eluru (2008). The Impact of Demographics, Built Environment Attributes, Vehicle Characteristics, and Gasoline Prices on Household Vehicle Holdings and Use. *Transportation Research-B,* 43(1) 1-18.

Habib, K. M. N., Miller, E. J. (2009) Modeling Activity Generation: A Utility-based Model for Activity Agenda Formation. *Transportmetrica* 5(1), 3-23.

Rajagopalan, B.S., and K.S., Srinivasan (2008) Integrating Household-Level Mode Choice and Modal Expenditure Decisions in a Developing Country: Multiple Discrete-Continuous Extreme Value Model. *Transportation Research Record*, No. 2076, 41-51.