<u>**Online supplement to**</u>

A New Closed Form Multiple Discrete-Continuous Extreme Value (MDCEV) Choice Model with Multiple Linear Constraints

Aupal Mondal and Chandra R. Bhat (corresponding author)

Other forecasting methods

Another approach, in addition to the one described in the main manuscript, is to repeat steps 1 through 3 for the first method for many sets of realizations. Count the number of times each of the possible $(2^{K-R}-1)$ combinations of discrete consumption of the inside goods appear as the chosen combination. Also, estimate the probability P_n of each discrete consumption combination *n* as the number of times it appears as the chosen combination relative to the total number of sets of realizations. Next, for each combination *n* (*n*=1,2,…,*N*, *N*= $2^{K-R}-1$), compute the mean value $\overline{x}_{kn}^*$ of the continuous consumption values across the many realizations. Finally, forecast the continuous amount of consumption for each inside alternative *k* as $x_k^* = \sum_n P_n \overline{x}_{kn}^*$. This approach will provide more accurate aggregate-level predictions (that is, predictions of consumption quantities across multiple individuals) than the first approach with small forecasting samples. But, for a given individual, given enough number of sets of realizations, it will always forecast a positive value of consumption for each and every alternative.

A third approach is to first compute the discrete probability P_n for each combination *n*, then use the usual discrete probability-to-deterministic choice procedure (used in traditional simulation approaches) to determine the most likely market basket of consumption, and forecast the consumption quantities for this single market basket. Specifically, the procedure is as follows.

- Step 1: Compute the discrete consumption probability for each possible consumption bundle *n* as follows:

$$P(d_{R+1}=1, d_{R+2}=1, \cdots d_{R+M}=1, d_{R+M+1}=0, \cdots d_{K-1}=0, d_K=0)$$
$$= \int_{\eta_{R+1}=\hat{\theta}_{R+1}}^{\eta_{R+1}=\infty} \int_{\eta_{R+2}=\hat{\theta}_{R+2}}^{\eta_{R+2}=\infty} \cdots \int_{\eta_{R+M}=\hat{\theta}_{R+M}}^{\eta_{R+M}=\infty} \int_{\eta_{R+M+1}=-\infty}^{\eta_{R+M+1}=\hat{\theta}_{R+M+1}} \cdots \int_{\eta_{K-1}=-\infty}^{\eta_{K-1}=\hat{\theta}_{K-1}} \int_{\eta_K=-\infty}^{\eta_K=\hat{\theta}_K} \boldsymbol{f}(\eta_{R+1}, \eta_{R+2}, \ldots, \eta_K)\, d\eta_K\, d\eta_{K-1,} \ldots d\eta_{R+1} \,,$$

where $\boldsymbol{f}(\eta_{R+1}, \eta_{R+2}, ..., \eta_K)$ represents the multivariate density function (pdf) of the random variates $\eta_{R+1}, \eta_{R+2}, ..., \eta_K$. The above expression may be written as:

$$P(d_{R+1} = 1, d_{R+2} = 1, \cdots d_{R+M} = 1, d_{R+M+1} = 0, \cdots d_{K-1} = 0, d_K = 0)$$
$$= S_M(\hat{\theta}_{R+1}, \hat{\theta}_{R+2}, ...\hat{\theta}_{R+M}) + \sum_{D \subset \{R+M+1, ..., K-1, K\}, |D| \geq 1} (-1)^{|D|} S_{M+|D|}(\hat{\theta}_{R+1}, \hat{\theta}_{R+2}, ...\hat{\theta}_{R+M}, \hat{\boldsymbol{\theta}}_D),$$

where $\boldsymbol{S}_D(.)$ for any dimension D is the multivariate survival distribution function given by Equation (11), D represents a specific combination of the non-consumed goods (there are a total of $2^{K-M-R} - 1$ possible combinations of the non-consumed goods), $|D|$ is the cardinality of the specific combination D, and $\hat{\boldsymbol{\theta}}_D$ is a vector of utility elements drawn from $\{\hat{\theta}_{R+M+1}, ...\hat{\theta}_{K-1}, \hat{\theta}_K\}$ that belong to the specific combination D. The discrete consumption probability for the case of none of the inside goods being consumed is already provided in Equation (21), while the discrete consumption probability for the case of all the inside goods being consumed is given by:

$$P(d_{R+1} = 1, d_{R+2} = 1, \cdots d_{R+M} = 1, d_{R+M+1} = 1, \cdots d_{K-1} = 1, d_K = 1)$$
$$= S_{K-R}(\hat{\theta}_{R+1}, \hat{\theta}_{R+2}, ...\hat{\theta}_K)$$

- Step 2: Order the combinations from 1 to N in an arbitrary order (but retain this from hereon), and, for each combination n up to the penultimate combination (n=1,2,…,N–1), obtain the cumulative probability from combination 1 to combination n as $CP_n = \sum_{d=1}^{n} P_d$.
- Step 3: Partition the 0-1 line into N segments (each corresponding to a specific combination n) using the (N–1) CP_n values. Draw a random uniformly distributed realization from {0,1} and superimpose this value over the 0-1 line with the N segments. Identify the segment where the realization falls, and declare the combination corresponding to that line segment as the deterministic discrete event of consumption for the individual.
- Step 4: For the specific combination declared as the discrete bundle of consumption from Step 3, forecast the continuous consumption as follows. Draw R independent realizations of ε_k from $\text{EV}(0,1)$ for the outside goods $(k = 1, 2, ..., R)$ and compute an estimate of $\hat{\xi}_k$ for each inside good. For each of the consumed goods in the bundle, draw a realization of ε_k (say μ_k)

from EV$(0, \hat{\sigma})$ truncated from below at $\hat{\xi}_k + \hat{\theta}_k$ (that is, such that $\mu_k > \hat{\xi}_k + \hat{\theta}_k$). Predict the continuous consumption value for the consumed goods as: $\hat{x}_k^* = \left[\exp\left\{\mu_k - \left(\hat{\theta}_k + \hat{\xi}_k\right)\right\} - 1\right]\hat{\gamma}_k$ and set $x_k^* = 0$ for the non-consumed goods. A variant of this step (4) would be to repeat step (4) multiple times with different sets of realizations, and take the mean across the resulting $\hat{x}_k^*$ predictions.