A note on MEV models

Michel Bierlaire

April 22, 2022

Report TRANSP-OR xxxxxx
Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne
`transp-or.epfl.ch`

Series on Biogeme

This document contains lecture notes about Multivariate Extreme Value models, introduced by McFadden (1978).

1 Definition

A vector $\varepsilon_n = (\varepsilon_{1n}, \ldots, \varepsilon_{Jn})$ follows a *multivariate extreme value*[1] distribution if it is characterized by the following cumulative distribution function:

$$F_{\varepsilon_n}(\xi_1, \ldots, \xi_J) = e^{-G(e^{-\xi_1}, \ldots, e^{-\xi_J})}, \tag{1}$$

where $G : \mathbb{R}_+^{J_n} \to \mathbb{R}_+$ is a positive function accepting positive arguments, denoted by $y_1, \ldots, y_J$ in the following. Note from (1) that, in our context, $y_i = \exp(-\xi_i)$, guaranteeing the positivity of the arguments. To be a valid CDF, the function F_{ε_n} must verify some properties, and so does G:

1. F_{ε_n} goes to zero when any argument goes to $-\infty$, that is

   $$F_{\varepsilon_n}(\xi_1, \ldots, -\infty, \ldots, \xi_J) = 0.$$

   When ξ_i goes to $-\infty$, then $y_i = \exp(-\xi_i)$ goes to $+\infty$. Consequently, the corresponding condition on G is

   $$G(y_1, \ldots, +\infty, \ldots, y_J) = +\infty, \tag{2}$$

   that is the function must go to infinity whenever one of its arguments does. It is called the *limit* property.

2. F_{ε_n} goes to one when all of its arguments go to $+\infty$, that is

   $$F_{\varepsilon_n}(+\infty, \ldots, +\infty) = 1.$$

   When ξ_i goes to $+\infty$, then $y_i = \exp(-\xi_i)$ goes to 0. Consequently, the corresponding condition on G is

   $$G(0, \ldots, 0) = 0. \tag{3}$$

3. Any partial derivative of F_{ε_n} defines a density function of a marginal distribution. To be a valid density function, it has to be non negative. More precisely, for any set of $\widehat{J}_n \leq J_n$ distinct indices $i_1, \ldots, i_{\widehat{J}_n}$,

   $$\frac{\partial^{\widehat{J}_n} F_{\varepsilon_n}}{\partial \varepsilon_{i_1 n} \cdots \partial \varepsilon_{i_{\widehat{J}_n} n}}(\varepsilon_{1n}, \ldots, \varepsilon_{J_n n}) \geq 0.$$

[1] There are several families of multivariate extreme value distributions. We refer the interested reader to Pickands (1981), Joe (1997) or Kotz and Nadarajah (2001), among others.

In particular, if $\widehat{J}_n = J_n$, we obtain the density function of the entire distribution of ε_n, that is

$$f_{\varepsilon_n}(\varepsilon_{1n}, \ldots, \varepsilon_{J_n n}) = \frac{\partial^{J_n} F_{\varepsilon_n}}{\partial \varepsilon_{1n} \cdots \partial \varepsilon_{J_n n}}(\varepsilon_{1n}, \ldots, \varepsilon_{J_n n}) \geq 0.$$

Considering (1), the above condition says that any level of differentiation must correspond to a non-negative result. It appears that the right-hand side of (1) changes sign each time it is differentiated, except the first time. Indeed, $\partial y_{in}/\partial \varepsilon_{in} = -\exp(-\varepsilon_{in}) = -y_{in}$. To compensate that and always obtain a non negative sign, the function G must also change sign each time it is differentiated. This condition is called the *strong alternating sign property*, and states that the cross partial derivatives of G have alternative signs[2]. That is, at the first degree,

$$G_i = \partial G/\partial y_i \geq 0,$$

for $i = 1, \ldots, J_n$. At the second degree,

$$G_{ij} = \partial G_i/\partial y_j = \partial^2 G/\partial y_i \partial y_j \leq 0,$$

for $i \neq j$. For higher degrees, and for any set of $\widehat{J}_n$ distinct indices $i_1, \ldots, i_{\widehat{J}_n}$,

$$(-1)^{\widehat{J}_n - 1} G_{i_1, \ldots, i_{\widehat{J}_n}} \geq 0. \tag{4}$$

If these properties are verified, (1) is a valid CDF. We also need an additional condition on G: *homogeneity*. A function G is homogeneous of degree μ, or μ-homogeneous, if

$$G(\alpha y) = \alpha^\mu G(y), \ \forall \alpha > 0 \text{ and } y \in \mathbb{R}^J_+. \tag{5}$$

The homogeneity condition implies two important properties of the model. We show below that, if G is homogeneous,

- the marginals of (1) are univariate extreme value distributions, so that the valid CDF indeed corresponds to a multivariate extreme value distribution, and,
- the corresponding choice model has a closed form.

[2]The *weak alternating sign property* requires that $\ln G$ changes sign each time it is differentiated. See Fosgerau et al. (2013) for a detailed discussion.

The ith marginal distribution of (1) is given by

$$F_{\varepsilon_n}(+\infty,\ldots,+\infty,\varepsilon_{in},+\infty,\ldots,+\infty) = e^{-G(0,\ldots,0,e^{-\varepsilon_{in}},0,\ldots,0)}. \tag{6}$$

If G is μ-homogeneous, we have

$$G(0,\ldots,0,e^{-\varepsilon_{in}},0,\ldots,0) = e^{-\mu\varepsilon_{in}}G(0,\ldots,0,1,0,\ldots,0),$$

or equivalently,

$$G(0,\ldots,0,e^{-\varepsilon_{in}},0,\ldots,0) = e^{-\mu\varepsilon_{in}+\ln G(0,\ldots,0,1,0,\ldots,0)},$$

The quantity $\ln G(0,\ldots,0,1,0,\ldots,0)$ is a constant. Call it $\mu\eta$, so that the CDF of the i^{th}marginal distribution of ε_n is

$$F_{\varepsilon_n}(+\infty,\ldots,+\infty,\varepsilon_{in},+\infty,\ldots,+\infty) = \exp\left(-e^{-\mu(\varepsilon_{in}-\eta)}\right), \tag{7}$$

which is the CDF of a univariate extreme value distribution with location parameter η and scale parameter μ.

We have now established that F is the CDF of a multivariate extreme value distribution if G verifies a handful of properties:

M1: the strong alternating sign property (4),

M2: the μ-homogeneity property (5), and,

M3: the limit property (2).

Note that the condition (3) is not included in the above list, as it is a direct consequence of the homogeneity property.

The corresponding choice model is:

$$P_n(i) = \frac{e^{V_{in}}G_i(e^V)}{\mu G(e^V)}, \tag{8}$$

or, equivalently,

$$P_n(i) = \frac{e^{V_{in}+\ln G_i\left(e^{V_{1n}},\ldots,e^{V_{J_n n}}\right)}}{\sum_j e^{V_{jn}+\ln G_j\left(e^{V_{1n}},\ldots,e^{V_{J_n n}}\right)}}. \tag{9}$$

This is the *multivariate extreme value* (MEV) model. The function G is called a *choice probability generating function* (CPGF).

Formulation (9) is interesting because it has a similar structure as the logit model. Indeed, it can be interpreted as a logit model, where each systematic utility V_i is shifted by $\ln G_i(\cdot)$. As a consequence, relaxing the independence assumption associated with the logit model can be accommodated by an appropriate correction of the utility functions, while keeping the functional form of the logit. However, it has to be remembered that the utility of an alternative i depends on the variables of all alternatives in the MEV context. Moreover, if the G function has a closed form, so has the choice model.

2 Properties of the MEV model

We present here various comments, properties and features of the MEV model.

- The multivariate extreme value model was first proposed by McFadden (1978), under the name "Generalized Extreme Value" model. In order to avoid any confusion with the Generalized Extreme Value distribution, and to emphasize that we are dealing with multivariate distributions, we refer to this model as *multivariate extreme value* (MEV).
- In the context of random utility, a random vector $U_n = (U_{1n}, \ldots, U_{J_n n}) = (V_{1n} + \varepsilon_{1n}, \ldots, V_{Jn} + \varepsilon_{Jn})$ with a MEV distribution is such that its CDF is

$$F_{U_n}(\xi_1, \ldots, \xi_{J_n}) = \Pr(U_n \leq \xi_n) = e^{-G(e^{V_{1n}-\xi_1}, \ldots, e^{V_{J_n n}-\xi_{J_n}})}. \tag{10}$$

- The marginal distributions of $(U_n)_j$ for $j = 1, \ldots, J_n$ are extreme value distributed, with
  - means:
  $$V_{jn} + \frac{\ln G(0, \ldots, 1, \ldots, 0) + \gamma}{\mu}, \tag{11}$$
  where γ is Euler's constant:
  $$\gamma = -\int_0^{+\infty} e^{-x} \ln x dx \approx 0.5772, \tag{12}$$
  - variances: $\pi^2/6\mu^2$, for each j, and
  - moment generating functions
  $$e^{tV_{jn}} G(0, \ldots, 1, \ldots, 0)^{\frac{t}{\mu}} \Gamma\left(1 - \frac{t}{\mu}\right), \tag{13}$$
  where $\Gamma(\cdot)$ is the Gamma function
  $$\Gamma(t) = \int_0^{+\infty} z^{t-1} e^{-z} dz.$$
- The variance covariance matrix of a MEV model is derived from its CDF (1). The covariance between the error terms of two alternatives i and j is given by

$$\begin{aligned} \mathrm{Cov}(\varepsilon_{in}, \varepsilon_{jn}) &= \mathrm{E}[\varepsilon_{in}\varepsilon_{jn}] - \mathrm{E}[\varepsilon_{in}]\,\mathrm{E}[\varepsilon_{jn}] \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \xi_i \xi_j \frac{\partial^2 F_{\varepsilon_n}(\xi_i, \xi_j)}{\partial \xi_i \partial \xi_j} d\xi_i d\xi_j - \gamma^2, \end{aligned} \tag{14}$$

where $E[\varepsilon_{in}] = \gamma$,

$$F_{\varepsilon_n}(\xi_i, \xi_j) = F_{\varepsilon_n}(\dots, +\infty, \xi_i, +\infty, \dots, +\infty, \xi_j, +\infty, \dots) \tag{15}$$

is the bivariate marginal cumulative distribution, and

$$\frac{\partial^2 F_{\varepsilon_{in},\varepsilon_{jn}}(\xi_i, \xi_j)}{\partial \xi_i \partial \xi_j} = F_{\varepsilon_{in},\varepsilon_{jn}}(\xi_i, \xi_j) e^{-\xi_i} e^{-\xi_j} (G_i^{ij} G_j^{ij} - G_{ij}^{ij}) \tag{16}$$

where

$$G_i^{ij} = \frac{\partial G(\dots, 0, e^{-\xi_i}, 0, \dots, 0, e^{-\xi_j}, 0, \dots)}{\partial y_i} \tag{17}$$

and

$$G_{ij}^{ij} = \frac{\partial^2 G(\dots, 0, e^{-\xi_i}, 0, \dots, 0, e^{-\xi_j}, 0, \dots)}{\partial y_i \partial y_j}. \tag{18}$$

In the general case, this double integral can only be computed numerically. It is recommended to apply first the change of variables $z_i = \exp(-\exp(-\xi_i))$ and $z_j = \exp(-\exp(-\xi_j))$.

- Contrarily to the probit model, the variance-covariance matrix does not characterize the distribution. Indeed, higher moments of the MEV distribution exist, and different MEV models can share the same variance-covariance (or the same correlation) matrix.

- McFadden's original result was derived from the assumption that G is a 1-homogeneous function. It is always possible through the normalization

$$G(y_1, \dots, y_J) = G^*(y_1^{1/\mu}, \dots, y_J^{1/\mu}) \tag{19}$$

to convert a μ-homogeneous function G^* into a 1-homogeneous function G. Indeed,

$$\begin{aligned} G(\alpha y_1, \dots, \alpha y_J) &= G^*((\alpha y_1)^{1/\mu}, \dots, (\alpha y_J)^{1/\mu}) \\ &= G^*(\alpha^{1/\mu} y_1^{1/\mu}, \dots, \alpha^{1/\mu} y_J^{1/\mu}) \\ &= (\alpha^{1/\mu})^{\mu} G^*(y_1^{1/\mu}, \dots, y_J^{1/\mu}) \\ &= \alpha G(y_1, \dots, y_J), \end{aligned}$$

where the first and last equations use (19), and the third is a consequence of the μ-homogeneity of G^*. In the choice model (9), the arguments y of the G function (or its derivatives) are e^V. Therefore, the normalization $y^{1/\mu}$ is $e^{V/\mu}$, which amounts to rescale the V's. Like for the logit model, the μ parameter is not identified from data and can be normalized to one.

- The logarithm of the CPGF is the expected maximum utility of the choice set for the model, that is

$$E[\max_{j\in\mathcal{C}_n} U_{jn}] = \frac{1}{\mu}(\ln G(e^{V_{1n}}, \ldots, e^{V_{J_n n}}) + \gamma), \tag{20}$$

  where γ is Euler's constant. As utilities are defined up to a constant, it is common to ignore the γ. Also, the parameter μ is usually normalized to one. Under this interpretation, the choice model can also be obtained from:

$$P_n(i) = \frac{\partial\, E[\max_{j\in\mathcal{C}_n} U_{jn}]}{\partial V_{in}}, \ \forall i \in \mathcal{C}_n,$$

  which is

$$P_n(i) = \frac{\partial \ln G(e^{V_{1n}}, \ldots, e^{V_{J_n n}})}{\partial V_{in}} = \frac{e^{V_{in}} G_i(e^{V_{1n}}, \ldots, e^{V_{J_n n}})}{G(e^{V_{1n}}, \ldots, e^{V_{J_n n}})}, \tag{21}$$

  justifying the name "choice probability generating function". Comparing (21) with (9), it is seen that G must be such that

$$G(e^{V_{1n}}, \ldots, e^{V_{J_n n}}) = \sum_j e^{V_{jn} + \ln G_j\left(e^{V_{1n}}, \ldots, e^{V_{J_n n}}\right)} \tag{22}$$

  or, equivalently,

$$G(e^{V_{1n}}, \ldots, e^{V_{J_n n}}) = \sum_j e^{V_{jn}} G_j\left(e^{V_{1n}}, \ldots, e^{V_{J_n n}}\right). \tag{23}$$

  This latter condition actually characterizes homogeneous functions, and is known as *Euler's theorem*.

- It can be shown that some operations maintain the properties of CPGF functions. Therefore, MEV functions can be constructed from others, and they all correspond to valid choice models. The following results are adapted from the *inheritance theorem* proposed by Daly and Bierlaire (2006). A MEV function which is homogeneous of degree μ is called here a μ-MEV function.

  Consider a choice set $\mathcal{C}$ with J alternatives. Consider also M subsets of alternatives $\mathcal{C}_m$, $m = 1, \ldots, M$, and let J_m be the number of alternatives in subset m. Let $G^m : \mathbb{R}_+^{J_m} \longrightarrow \mathbb{R}$, $m = 1, \ldots, M$ be M μ_m-MEV functions on $\mathcal{C}_m$. Then, the function

$$G : \mathbb{R}_+^J \longrightarrow \mathbb{R} : y \rightsquigarrow G(y) = \sum_{m=1}^{M} \left(\alpha_m G^m([y]_m)\right)^{\frac{\mu}{\mu_m}} \tag{24}$$

is a μ-MEV function if $\alpha_m > 0$, $\mu > 0$ and $\mu_m \geq \mu$, $m = 1, \ldots, m$, where $[y]_m$ denotes a vector of dimension J_m with entries y_i, where the indices i correspond to the elements in $\mathcal{C}_m$.

This result has some interesting corollaries.

1. If $G(y)$ is a μ-MEV function, so is $\alpha G(y)$, with $\alpha > 0$. The inheritance theorem can be invoked with $M = 1$ and $\mu_m = \mu$.
2. If $G(y)$ is a μ-MEV function and $\widehat{\mu} \geq 1$, then $G(y^{\widehat{\mu}})^{1/\widehat{\mu}}$ is also a μ-MEV function. Indeed, $G^*(y) = G(y^{\widehat{\mu}})$ is a $(\mu\widehat{\mu})$-MEV function. By the theorem,
$$G^*(y)^{\frac{\mu}{\mu\widehat{\mu}}} = G(y^{\widehat{\mu}})^{\frac{1}{\widehat{\mu}}}$$
is a μ-MEV function, as $\mu\widehat{\mu} \geq \mu$.
3. Any linear combination of μ-MEV functions is also a μ-MEV function if the multipliers are non negative and at least one is strictly positive.
4. If $P_m(i)$ is the choice model derived from the μ_m-MEV function G^m, then the choice model $P(i)$ derived from the μ-MEV function G defined by (24) is
$$P(i) = \sum_{m=1}^{M} \frac{(\alpha_m G^m(e^V))^{\frac{\mu}{\mu_m}}}{\sum_{p=1}^{M}(\alpha_p G^p(e^V))^{\frac{\mu}{\mu_p}}} P_m(i). \tag{25}$$

3 The logit model as MEV

The logit model is a MEV model derived from the following choice probability generating function:
$$G(y) = \sum_{i=1}^{J} y_i^{\mu}. \tag{26}$$

Properties [M1]–[M3] are trivially verified:

M1 The strong alternating sign property is a consequence of the fact that $\mu > 0$, $y_i > 0$ and
$$G_i(y) = \frac{\partial G}{\partial y_i} = \mu y_i^{\mu-1}. \tag{27}$$

Derivatives of higher orders are all zero:
$$G_{ij}(y) = \frac{\partial^2 G}{\partial y_i \partial y_j} = 0, \text{ if } i \neq j.$$

M2 The function is μ-homogeneous as

$$G(\alpha y) = \sum_{i=1}^{J} (\alpha y_i)^\mu = \alpha^\mu \sum_{i=1}^{J} y_i^\mu = \alpha^\mu G(y).$$

M3 The limit property is also verified, as

$$\begin{aligned} G(y_1, \ldots, y_{j-1}, +\infty, y_{j+1}, \ldots, y_J) &= \lim_{y_j \to +\infty} \sum_{i=1}^{J} y_i^\mu \\ &= \sum_{i \neq j} y_i^\mu + \lim_{y_j \to +\infty} y_j^\mu = +\infty. \end{aligned}$$

Consequently, (26) defines a μ-MEV function. From (1), the CDF is

$$\begin{aligned} F_\varepsilon(\xi_1, \ldots, \xi_J) &= e^{-G(e^{-\xi_1}, \ldots, e^{-\xi_J})} \\ &= e^{-\sum_{i=1}^{J} e^{-\mu \xi_i}} \\ &= \prod_{i=1}^{J} e^{-e^{-\mu \xi_i}}. \end{aligned}$$

Substituting

$$e^{V_i + \ln G_i(e^{V_1}, \ldots, e^{V_J})} = e^{V_i + \ln \mu + (\mu - 1) \ln e^{V_i}} = e^{\ln \mu + \mu V_i}$$

into (9), we obtain the choice probability

$$P(i) = \frac{e^{\ln \mu + \mu V_{in}}}{\sum_{j \in \mathcal{C}} e^{\ln \mu + \mu V_{jn}}} = \frac{e^{\mu V_{in}}}{\sum_{j \in \mathcal{C}} e^{\mu V_{jn}}},$$

which is indeed the choice probability of a logit model.

From (20), the expected maximum utility is

$$\frac{1}{\mu} \ln G\left(e^{V_{1n}}, \ldots, e^{V_{Jn}}\right) = \frac{1}{\mu} \ln \sum_{i=1}^{J} e^{\mu V_{in}},$$

sometimes called the "logsum" formula.

We also illustrate the computation of the covariance from (14). We consider two distinct alternatives i and j. The double integral in (14) writes

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \xi_i \xi_j \frac{\partial^2 F_{\varepsilon_i, \varepsilon_j}(\xi_i, \xi_j)}{\partial \xi_i \partial \xi_j} d\xi_i d\xi_j =$$

$$\int_{-\infty}^{+\infty} \xi_i e^{-e^{-\xi_i}} e^{-\xi_i} d\xi_i \int_{-\infty}^{+\infty} \xi_j e^{-e^{-\xi_j}} e^{-\xi_j} d\xi_j,$$

where index n has been dropped for notational convenience. Applying the change of variable $t_i = e^{-\xi_i}$, we have

$$\int_{-\infty}^{+\infty} \xi_i e^{-e^{-\xi_i}} e^{-\xi_i} d\xi_i = -\int_0^{+\infty} \ln t_i e^{-t_i} dt_i = \gamma,$$

so that

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \xi_i \xi_j \frac{\partial^2 F_{\varepsilon_i,\varepsilon_j}(\xi_i, \xi_j)}{\partial \xi_i \partial \xi_j} d\xi_i d\xi_j = \gamma^2. \tag{28}$$

Using (28) in (14), we obtain that the covariance between any pair of alternatives is 0, which is expected for the logit model.

4 The nested logit model as MEV

We consider a nested logit model, where the choice set $\mathcal{C}$ is partitioned into mutually exclusive nests $\mathcal{C}_m$, $m = \ldots, M$. The nested logit model with M nests is a MEV model, with CPGF

$$G(y) = \sum_{m=1}^{M} \left(\sum_{\ell \in \mathcal{C}_m} y_\ell^{\mu_m} \right)^{\mu/\mu_m}. \tag{29}$$

If $0 < \mu \leq \mu_m$, for all m, (29) defines a μ-MEV function. Properties [M2] and [M3] are trivially verified. In order to check the strong alternating sign property [M1], we consider alternative i in nest m. We have

$$G_i(y) = \frac{\partial G}{\partial y_i}(y) = \mu y_i^{\mu_m - 1} \left(\sum_{\ell \in \mathcal{C}_m} y_\ell^{\mu_m} \right)^{\frac{\mu}{\mu_m} - 1}. \tag{30}$$

It is non-negative when $y \in \mathbb{R}_+^J$. Consider now another alternative $j \neq i$. If j does not belong to the same nest as i, then

$$\frac{\partial^2 G}{\partial y_i y_j}(y) = 0. \tag{31}$$

If j belongs to m, we have

$$\frac{\partial^2 G}{\partial y_i y_j}(y) = \mu \mu_m (\frac{\mu}{\mu_m} - 1) y_i^{\mu_m - 1} y_j^{\mu_m - 1} \left(\sum_{\ell \in \mathcal{C}_m} y_\ell^{\mu_m} \right)^{\frac{\mu}{\mu_m} - 2}. \tag{32}$$

We need to verify that (32) is non-positive. It is the case as $\mu > 0$ and $\mu \leq \mu_m$. Indeed, all terms in the product are non-negative, except for the term

$(\mu/\mu_m - 1)$, which is non-positive. Each additional differentiation involves an additional factor of the form $(\mu/\mu_m - k)$, $k > 1$, which is always negative. Therefore, each differentiation leads to a change of sign, and [M1] is verified.

We now derive the covariance between the error terms of two alternatives i and j in nest m. We first normalize all utilities by μ is order to transform G into a 1-homogenous function, without loss of generality, as suggested by (19). Then we consider the bivariate marginal cumulative distribution

$$\begin{aligned}
F_{\varepsilon_n}(\xi_i, \xi_j) &= F_{\varepsilon_n}(\ldots, +\infty, \xi_i/\mu, +\infty, \ldots, +\infty, \xi_j/\mu, +\infty, \ldots) \\
&= \exp\left(-G(\ldots, 0, \exp(-\xi_i/\mu), 0, \ldots, 0, \exp(-\xi_j/\mu), 0, \ldots)\right) \\
&= \exp\left(-\left(\exp(-\frac{\mu_m}{\mu}\xi_i) + \exp(-\frac{\mu_m}{\mu}\xi_i)\right)^{\frac{\mu}{\mu_m}}\right).
\end{aligned}$$

This is the CDF of a bivariave logistic model (see Kotz et al., 2000, p. 628) with parameter $m = \mu_m/\mu$ and correlation

$$\rho = 1 - m^{-2} = 1 - \frac{\mu^2}{\mu_m^2}. \tag{33}$$

Finally, we derive the choice model. For each nest m, we denote

$$S_m = \sum_{j \in \mathcal{C}_m} e^{\mu_m V_{jn}},$$

so that

$$G = \sum_m S_m^{\frac{\mu}{\mu_m}},$$

and

$$G_i = \mu e^{\mu_m V_{in}} e^{-V_{in}} S_m^{\frac{\mu}{\mu_m} - 1}.$$

We use (8) to obtain

$$\begin{aligned}
P_n(i) &= \frac{e^{V_{in}} \mu e^{\mu_m V_{in}} e^{-V_{in}} S_m^{\frac{\mu}{\mu_m} - 1}}{\mu \sum_\ell S_\ell^{\frac{\mu}{\mu_\ell}}} \\
&= \frac{e^{\mu_m V_{in}} S_m^{\frac{\mu}{\mu_m} - 1}}{\sum_\ell S_\ell^{\frac{\mu}{\mu_\ell}}} \\
&= \frac{e^{\mu_m V_{in}}}{S_m} \frac{S_m^{\frac{\mu}{\mu_m}}}{\sum_\ell S_\ell^{\frac{\mu}{\mu_\ell}}} \\
&= \frac{e^{\mu_m V_{in}}}{\sum_{j \in \mathcal{C}_m} e^{\mu_m V_{jn}}} \frac{\left(\sum_{j \in \mathcal{C}_m} e^{\mu_m V_{jn}}\right)^{\frac{\mu}{\mu_m}}}{\sum_\ell \left(\sum_{j \in \mathcal{C}_\ell} e^{\mu_\ell V_{jn}}\right)^{\frac{\mu}{\mu_\ell}}},
\end{aligned}$$

which is the nested logit model.

The multiple level nested logit model is also a MEV model. Consider for instance a 3-level case, where the choice set is partitioned into p groups, each of them partitioned into M_p nests. The model is derived from the following CPGF:

$$G(y) = \sum_{p=1}^{P} \left(\sum_{m=1}^{M_p} \left(\sum_{i=1}^{J_{mp}} y_i^{\mu_{mp}} \right)^{\mu_p/\mu_{mp}} \right)^{\mu/\mu_p}, \tag{34}$$

where J_{mp} is the number of alternatives in the m^{th} nest within group p. It can be verified that the condition:

$$0 \leq \mu \leq \mu_m \leq \mu_{pm}, \text{ for all } m, p,$$

is sufficient for (34) to define a CPGF function.

5 The cross nested logit model

If we consider M nests, the cross nested logit model is a MEV model based on the following CPGF:

$$G(y) = \sum_{m=1}^{M} \left(\sum_{j=1}^{J} \alpha_{jm}^{\frac{\mu_m}{\mu}} y_j^{\mu_m} \right)^{\mu/\mu_m}, \tag{35}$$

where μ_m is a parameter associated with nest m (it plays a similar role as the μ_m parameter in the nested logit model), and α_{jm} are parameters capturing the level of membership of alternative j in nest m. We immediately note that the nested logit model is a special instance of the cross nested logit model where $\alpha_{im} = 1$ if alternative i belongs to nest m, and $\alpha_{im} = 0$ otherwise. For this reason, Wen and Koppelman (2001) prefer to call the model the *Generalized Nested Logit Model*, although this terminology does not prevail in the literature.

Bierlaire (2006) has shown that the conditions

1. $\alpha_{im} \geq 0$, $\forall i, m$,
2. $\sum_m \alpha_{im} > 0$, $\forall i$, and
3. $0 < \mu \leq \mu_m$, $\forall m$,

are sufficient for (35) to be a CPGF. Note that condition 3 is the same as for the nested logit model (see Section 4).

The CNL model must be normalized before being estimated. As any MEV model, the normalization $\mu = 1$ is applicable. Moreover, if the α parameters are estimated, not all of them are identified and the following normalization is appropriate:

$$\sum_{m=1}^{M} \alpha_{im} = 1, \quad \forall i = 1, \ldots, J. \tag{36}$$

This normalization is consistent with the interpretation of α_{im} as the level of membership of alternative i in nest m. The derivative of (35) is

$$G_i(y) = \frac{\partial G(y)}{\partial y_i} = \mu \sum_{m=1}^{M} \alpha_{im}^{\frac{\mu_m}{\mu}} y_i^{\mu_m - 1} \left(\sum_{j=1}^{J} \alpha_{jm}^{\frac{\mu_m}{\mu}} y_j^{\mu_m} \right)^{\frac{\mu}{\mu_m} - 1}. \tag{37}$$

Substituting (37) in (9), we obtain the cross nested logit model:

$$P_n(i) = \sum_{m=1}^{M} \frac{\left(\sum_{j \in \mathcal{C}_n} \alpha_{jm}^{\mu_m/\mu} e^{\mu_m V_{jn}}\right)^{\frac{\mu}{\mu_m}}}{\sum_{p=1}^{M} \left(\sum_{j \in \mathcal{C}_n} \alpha_{jp}^{\mu_p/\mu} e^{\mu_p V_{jn}}\right)^{\frac{\mu}{\mu_p}}} \frac{\alpha_{im}^{\mu_m/\mu} e^{\mu_m V_{in}}}{\sum_{j \in \mathcal{C}_n} \alpha_{jm}^{\mu_m/\mu} e^{\mu_m V_{jn}}}, \tag{38}$$

which can nicely be interpreted as

$$P_n(i) = \sum_{m=1}^{M} P_n(m|\mathcal{C}_n) P_n(i|m), \tag{39}$$

where

$$P_n(i|m) = \frac{\alpha_{im}^{\mu_m/\mu} e^{\mu_m V_{in}}}{\sum_{j \in \mathcal{C}_n} \alpha_{jm}^{\mu_m/\mu} e^{\mu_m V_{jn}}}, \tag{40}$$

is the choice probability conditional to nest m, and

$$P_n(m|\mathcal{C}_n) = \frac{\left(\sum_{j \in \mathcal{C}_n} \alpha_{jm}^{\mu_m/\mu} e^{\mu_m V_{jn}}\right)^{\frac{\mu}{\mu_m}}}{\sum_{p=1}^{M} \left(\sum_{j \in \mathcal{C}_n} \alpha_{jp}^{\mu_p/\mu} e^{\mu_p V_{jn}}\right)^{\frac{\mu}{\mu_p}}}, \tag{41}$$

is the probability associated with nest m.

The choice model can also be written in a form where the utilities are shifted:

$$P_n(i) = \sum_{m=1}^{M} \frac{\left(\sum_{j \in \mathcal{C}_n} \exp(\mu_m(V_{jn} + \frac{1}{\mu} \ln \alpha_{jm}))\right)^{\frac{\mu}{\mu_m}}}{\sum_{p=1}^{M} \left(\sum_{j \in \mathcal{C}_n} \exp(\mu_p(V_{jn} + \frac{1}{\mu} \ln \alpha_{jp}))\right)^{\frac{\mu}{\mu_p}}} \frac{e^{\mu_m(V_{in} + \frac{1}{\mu} \ln \alpha_{im})}}{\sum_{j \in \mathcal{C}_n} e^{\mu_m(V_{jn} + \frac{1}{\mu} \ln \alpha_{jm})}}. \tag{42}$$

The correlation structure must be computed using (14), where

$$F_{\varepsilon_i,\varepsilon_j}(\xi_i, \xi_j) = \exp\left(-\sum_{m=1}^{M}\left(\left(\alpha_{im}^{\frac{1}{\mu}} e^{-\xi_i}\right)^{\mu_m} + \left(\alpha_{jm}^{\frac{1}{\mu}} e^{-\xi_j}\right)^{\mu_m}\right)^{\frac{1}{\mu_m}}\right). \quad (43)$$

The cross nested logit model provides an intuitive way to capture complex correlation structures. Indeed, any source of correlation assumed by the analyst can be represented by a nest, and the alternatives involved are associated with the nest. As each alternative can potentially belong to more than one nest, a great deal of flexibility is provided. Actually, Fosgerau et al. (2013) have shown than any additive random utility model can be approximated by a cross nested logit model.

6 Derivation of the MEV model

The choice model is obtained by incorporating (1)

$$F_{\varepsilon_n}(\varepsilon_{1n}, \ldots, \varepsilon_{Jn}) = e^{-G(e^{-\varepsilon_{1n}},\ldots,e^{-\varepsilon_{Jn}})},$$

into the general definition of a choice model from random utility theory:

$$P_n(i) = \int_{\varepsilon=-\infty}^{+\infty} \frac{\partial F_{\varepsilon_{1n},\varepsilon_{2n},\ldots,\varepsilon_{Jn}}}{\partial \varepsilon_i}(\ldots, V_{in} - V_{(i-1)n} + \varepsilon, \varepsilon, V_{in} - V_{(i+1)n} + \varepsilon, \ldots) d\varepsilon.$$

We have

$$\begin{aligned}
&\frac{\partial F_{\varepsilon_{1n},\varepsilon_{2n},\ldots,\varepsilon_{Jn}}}{\partial \varepsilon_i}(\ldots, V_{in} - V_{(i-1)n} + \varepsilon, \varepsilon, V_{in} - V_{(i+1)n} + \varepsilon, \ldots) \\
&= e^{-\varepsilon} G_i(\ldots, e^{-V_{in}+V_{(i-1)n}-\varepsilon}, e^{-\varepsilon}, e^{-V_{in}+V_{(i+1)n}-\varepsilon}, \ldots) \\
&\qquad \exp\left(-G(\ldots, e^{-V_{in}+V_{(i-1)n}-\varepsilon}, e^{-\varepsilon}, e^{-V_{in}+V_{(i+1)n}-\varepsilon}, \ldots)\right) \\
&= e^{-\varepsilon} e^{-(\mu-1)\varepsilon} e^{-(\mu-1)V_{in}} G_i(\ldots, e^{V_{(i-1)n}}, e^{V_{in}}, e^{V_{(i+1)n}}, \ldots) \\
&\qquad \exp\left(-e^{-\mu\varepsilon} e^{-\mu V_{in}} G(\ldots, e^{V_{(i-1)n}}, e^{V_{in}}, e^{V_{(i+1)n}}, \ldots)\right),
\end{aligned}$$

because G is μ-homogeneous, which implies that G_i is $(\mu-1)$-homogeneous. We now denote

$$e^V = \left(\ldots, e^{V_{(i-1)n}}, e^{V_{in}}, e^{V_{(i+1)n}}, \ldots\right),$$

and simplify the terms to obtain

$$\begin{aligned}
&\frac{\partial F_{\varepsilon_{1n},\varepsilon_{2n},\ldots,\varepsilon_{Jn}}}{\partial \varepsilon_i}(\ldots, V_{in} - V_{(i-1)n} + \varepsilon, \varepsilon, V_{in} - V_{(i+1)n} + \varepsilon, \ldots) \\
&= e^{-\mu\varepsilon} e^{-\mu V_{in}} e^{V_{in}} G_i(e^V) \exp\left(-e^{-\mu\varepsilon} e^{-\mu V_{in}} G(e^V)\right).
\end{aligned}$$

Therefore,

$$P_n(i) = e^{-\mu V_{in}} e^{V_{in}} G_i(e^V) \int_{\varepsilon=-\infty}^{+\infty} e^{-\mu\varepsilon} \exp\left(-e^{-\mu\varepsilon} e^{-\mu V_{in}} G(e^V)\right) d\varepsilon.$$

Defining $t = -\exp(-\mu\varepsilon)$, so that $dt = \mu \exp(-\mu\varepsilon) d\varepsilon$, we write

$$P_n(i) = e^{-\mu V_{in}} e^{V_{in}} G_i(e^V) \frac{1}{\mu} \int_{t=-\infty}^{0} \exp\left(t e^{-\mu V_{in}} G(e^V)\right) dt,$$

which simplifies to

$$P_n(i) = \frac{e^{V_{in}} G_i(e^V)}{\mu G(e^V)}.$$

We finally invoke Euler's theorem that characterizes homogeneous functions to obtain (9):

$$P_n(i) = \frac{e^{V_{in} + \ln G_i\left(e^V\right)}}{\sum_j e^{V_{jn} + \ln G_j(e^V)}}.$$

7 Copulas

The concept of MEV functions is closely related to the concept of copulas in statistics. A copula is the CDF of a multivariate distribution such that every marginal distribution is uniform in the interval $[0, 1]$. We refer the reader to Nelsen (2006) for an introduction to copulas. A result by Sklar (1959) states that any multivariate distribution is entirely characterized by its marginals, and a copula. Intuitively, the copula captures the dependence among the various dimensions. More precisely, consider the CDF of a multivariate random vector ε

$$F(\varepsilon_1, \ldots, \varepsilon_J)$$

and denote $F_j(\varepsilon_j)$ the CDF of its univariate marginal distribution associated with dimension j. The copula of F is defined as

$$C : [0, 1]^J \to \mathbb{R} : (u_1, \ldots, u_J) \to C(u_1, \ldots, u_J) = F\left(F_1^{-1}(u_1), \ldots, F_J^{-1}(u_J)\right),$$

where $F_j^{-1} : [0, 1] \to \mathbb{R}$ denotes the inverse function of the marginals. Conversely, given a copula C, multivariate distributions can be constructed from the marginal distributions $F_j(\varepsilon_j)$:

$$F(\varepsilon_1, \ldots, \varepsilon_J) = C\left(F_1(\varepsilon_1), \ldots, F_J(\varepsilon_J)\right).$$

The link between copulas and MEV function is given by the following result (Joe, 1997): a multivariate random variable with CDF $F(\varepsilon_1, \ldots, \varepsilon_J)$ has a MEV distribution if and only if its copula satisfies the following condition:

$$C(u_1, \ldots, u_J)^\alpha = C(u_1^\alpha, \ldots, u_J^\alpha),$$

for $u \in [0, 1]^J$ and $\alpha > 0$. Actually, $\ln C$ plays the role of the choice probability generating function defined above.

These results, although quite technical, can be exploited to generate new MEV models from the theory of copulas. We refer the interested reader to Nikoloulopoulos and Karlis (2008), Bhat and Sener (2009) or Fosgerau et al. (2013) for more details.

8 Derivatives of the CDF of the cross-nested logit model

We write (35) as

$$G(y) = \sum_{m=1}^{M} \left(\sum_{k=1}^{J} t_{km} \right)^{\mu/\mu_m},$$

where

$$t_{km} = \alpha_{km}^{\frac{\mu_m}{\mu}} y_k^{\mu_m}.$$

Therefore,

$$\begin{aligned} G_i = \frac{dG}{dy_i} &= \sum_m \frac{\partial G}{\partial t_{im}} \frac{\partial t_{im}}{\partial y_i} \\ &= \sum_m \frac{\mu}{\mu_m} \left(\sum_{k=1}^{J} t_{km} \right)^{\frac{\mu}{\mu_m}-1} \mu_m \alpha_{im}^{\frac{\mu_m}{\mu}} y_i^{\mu_m - 1} \\ &= \mu \sum_m \left(\sum_{k=1}^{J} t_{km} \right)^{\frac{\mu}{\mu_m}-1} \alpha_{im}^{\frac{\mu_m}{\mu}} y_i^{\mu_m - 1}. \end{aligned}$$

Moreover, if $i \neq j$,

$$
\begin{aligned}
G_{ij} = \frac{\partial G_i}{\partial y_j} &= \sum_m \frac{\partial G_i}{\partial t_{jm}} \frac{\partial t_{jm}}{\partial y_j} \\
&= \sum_m \mu(\frac{\mu}{\mu_m} - 1) \left(\sum_{k=1}^{J} t_{km} \right)^{\frac{\mu}{\mu_m} - 2} \alpha_{im}^{\frac{\mu_m}{\mu}} y_i^{\mu_m - 1} \mu_m \alpha_{jm}^{\frac{\mu_m}{\mu}} y_j^{\mu_m - 1} \\
&= \mu \sum_m (\mu - \mu_m) \left(\sum_{k=1}^{J} t_{km} \right)^{\frac{\mu}{\mu_m} - 2} \alpha_{im}^{\frac{\mu_m}{\mu}} y_i^{\mu_m - 1} \alpha_{jm}^{\frac{\mu_m}{\mu}} y_j^{\mu_m - 1}.
\end{aligned}
$$

Finally,

$$
\begin{aligned}
G_{ii} = \frac{dG_i}{dy_i} =& \sum_m \frac{\partial G_i}{\partial t_{im}} \frac{\partial t_{im}}{\partial y_i} + \frac{\partial G_i}{y_i} \\
=&\mu \sum_m (\mu - \mu_m) \left(\sum_{k=1}^{J} t_{km} \right)^{\frac{\mu}{\mu_m} - 2} \alpha_{im}^{\frac{\mu_m}{\mu}} y_i^{\mu_m - 1} \alpha_{im}^{\frac{\mu_m}{\mu}} y_i^{\mu_m - 1} \\
&+ (\mu_m - 1) \left(\sum_{k=1}^{J} t_{km} \right)^{\frac{\mu}{\mu_m} - 1} \alpha_{im}^{\frac{\mu_m}{\mu}} y_i^{\mu_m - 2}.
\end{aligned}
$$

The derivatives of the CDF are calculated as

$$
\begin{aligned}
F_i = \frac{\partial F}{\partial \xi_i} &= \frac{\partial F}{\partial G} \frac{\partial G}{\partial y_i} \frac{\partial y_i}{\partial \xi_i} \\
&= -e^{-G} G_i (-e^{-\xi_i}) \\
&= F G_i e^{-\xi_i},
\end{aligned}
$$

and, if $i \neq j$,

$$
\begin{aligned}
F_{ij} = \frac{\partial F_i}{\partial \xi_j} &= \frac{\partial F}{\partial \xi_j} G_i e^{-\xi_i} + F \frac{\partial G_i}{\partial y_j} \frac{\partial y_j}{\partial \xi_j} e^{-\xi_i} \\
&= F_j G_i e^{-\xi_i} - F G_{ij} e^{-\xi_i} e^{-\xi_j} \\
&= F G_j e^{-\xi_j} G_i e^{-\xi_i} - F G_{ij} e^{-\xi_i} e^{-\xi_j} \\
&= F e^{-\xi_i} e^{-\xi_j} (G_i G_j - G_{ij}).
\end{aligned}
$$

Finally,

$$
F_{ii} = F e^{-\xi_i} e^{-\xi_i} (G_i G_i - G_{ii}) - F G_i e^{-\xi_i}.
$$

References

Bhat, C. and Sener, I. (2009). A copula-based closed-form binary logit choice model for accommodating spatial correlation across observational units, *Journal of Geographical Systems* **11**(3): 243–272–272.
URL: *http://dx.doi.org/10.1007/s10109-009-0077-9*

Bierlaire, M. (2006). A theoretical analysis of the cross-nested logit model, *Annals of operations research* **144**(1): 287–300.

Daly, A. and Bierlaire, M. (2006). A general and operational representation of generalised extreme value models, *Transportation Research Part B* **40**(4): 285–305.

Fosgerau, M., McFadden, D. and Bierlaire, M. (2013). Choice probability generating functions, *Journal of Choice Modelling* pp. 1–18.

Joe, H. (1997). *Multivariate models and dependence concepts*, Chapman & Hall.
URL: *http://books.google.com/books?id=iJbRZL2QzMAC*

Kotz, S., Balakrishnan, N. and Johnson, N. L. (2000). *Continuous multivariate distributions*, Vol. 1: Models and Applications of *Wiley Series in Probability and Statistics*, 2nd edn, John Wiley and Sons.

Kotz, S. and Nadarajah, S. (2001). *Extreme Value Distributions: Theory and Applications*, Imperial College Press.

McFadden, D. (1978). Modelling the choice of residential location, *in* A. Karlquist *et al.* (ed.), *Spatial interaction theory and residential location*, North-Holland, Amsterdam, pp. 75–96.

Nelsen, R. B. (2006). *An introduction to copulas*, Springer.
URL: *http://books.google.com/books?id=B3ONT5rBv0wC*

Nikoloulopoulos, A. K. and Karlis, D. (2008). Multivariate logit copula model with an application to dental data, *Statistics in Medicine* **27**(30): 6393–6406.
URL: *http://dx.doi.org/10.1002/sim.3449*

Pickands, J. (1981). Multivariate extreme value distributions, *Proceedings 43rd Session International Statistical Institute*, pp. 859–878.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges, *Publ. Inst. Statist. Univ. Paris* **8**: 229–231.

Wen, C.-H. and Koppelman, F. S. (2001). The generalized nested logit model, *Transportation Research Part B* **35**(7): 627–641.