Recent developments in discrete choice modeling

Properties of the Multivariate Extreme Value models and their impact on applications

Michel Bierlaire

transp-or.epfl.ch

Transport and Mobility Laboratory, EPFL

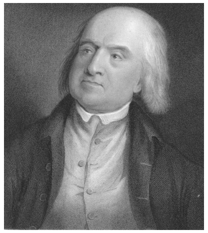
Outline

- Introduction
- Discrete choice models
- Relax the independence assumption
- MEV models
- Impact on practice
  - Software.
  - Choice-based samples.
  - Large choice sets.
  - Design of new MEV models.

Introduction

Homo Economicus (source: D. McFadden)

Jeremy Bentham (1789) — My notion of man is that ...he aims at happiness ...in every thing he does.



Frank Taussig (1912) — The fact that [the consumer] is willing to give up something in order to procure an article proves once for all that for him it has utility

Herb Simon (1956) — The rational man of economics is a maximizer, who will settle for nothing less than the best.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Discrete choice models

- Finite and discrete set of alternatives
  - Choice of transportation mode: car, bus, etc.
  - Choice of brand: Leonidas, Lindt, Suchard, Toblerone, etc.
  - Choice of university: ETHZ, EPFL, etc.
- Individual n associates a utility to alternative i
- Represented by a random function

$$U_{in} = V_{in} + \varepsilon_{in} = \sum_k \beta_k x_{ink} + \varepsilon_{in}$$

For instance

$$U_{\text{Leonidas,mb}} = \beta_1 \text{sugar} + \beta_2 \text{bitterness} + \beta_3 \text{Belgian} + \ldots + \varepsilon_{\text{Leonidas,mb}}$$

Example: Choice of Space-Water Heating System

- Data: Hydro-Québec 1989 (Bernard, Bolduc and Bélanger, 1996)
- 2897 single-family houses with a recent heating system (max. 3 years)

	Water heating			
Space heating	Natural gas	Oil	Electricity	Total
Natural gas	27		9	36
Dual energy		72	201	273
Oil		12	20	32
Electricity			2351	2351
Wood			124	124
Wood-electricity			81	81
Total	27	84	2786	2897

Example: Choice of Space-Water Heating System

Choice set: 9 alternatives

Name	Percent	Frequency	Availability
1. Gas/Gas	0.0093	27	543
2. Gas/Electricity	0.0031	9	543
3. Dual energy/oil	0.0249	72	2897
4. Dual energy/electricity	0.0694	201	2897
5. Oil/Oil	0.0041	12	2897
6. Oil/Electricity	0.0069	20	2897
7. Electricity/Electricity	0.8115	2351	2897
8. Wood/Electricity	0.0428	124	2897
9. Wood+Electricity/Electricity	0.0280	81	2897
Total	1.0000	2897	21365

Discrete choice models

- Individual n chooses alternative i if $U_{in} \geq U_{jn}$, for all j.
- Utility is random, so we have a probabilistic model

$$P_n(i|\mathcal{C}_n) = \Pr(U_{in} \geq U_{jn}) = \Pr(V_{in} + \varepsilon_{in} \geq V_{jn} + \varepsilon_{jn})$$

- Concrete models require
  - specification of V_{in}
  - assumptions about ε_{in}
  - estimation of the parameters from data

Example: Model specification

	G/G	G/E	D/O	D/E	O/O	O/E	E/E	W/E	WE/E
β_1	1	0	0	0	0	0	0	0	0
β_2	0	1	0	0	0	0	0	0	0
β_3	0	0	1	0	0	0	0	0	0
β_4	0	0	0	1	0	0	0	0	0
β_5	0	0	0	0	1	0	0	0	0
β_6	0	0	0	0	0	1	0	0	0
β_7	0	0	0	0	0	0	0	1	0
β_8	0	0	0	0	0	0	0	0	1
β_9	OC	OC	OC	OC	OC	OC	OC	OC	OC
β_{10}	FC	FC	FC	FC	FC	FC	FC	FC	FC
β_{11}	FC $\times$ I	FC $\times$ I	FC $\times$ I	FC $\times$ I	FC $\times$ I	FC $\times$ I	FC $\times$ I	FC $\times$ I	FC $\times$ I
β_{12}	I	0	0	0	0	0	0	0	0
β_{13}	0	I	0	0	0	0	0	0	0
β_{14}	0	0	I	0	0	0	0	0	0
β_{15}	0	0	0	I	0	0	0	0	0
β_{16}	0	0	0	0	I	0	0	0	0
β_{17}	0	0	0	0	0	I	0	0	0
β_{18}	0	0	0	0	0	0	0	I	0
β_{19}	0	0	0	0	0	0	0	0	I

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Discrete choice models

Assumptions about ε_{in}

- If V_{in} contains a constant, the mean can be assumed to be 0 wlog.
- Logit model: independent and identically distributed (i.i.d., across i and n), extreme value distribution

$$\varepsilon_{in} \sim \text{EV}(0, \mu)$$

EV(η, μ), with $\mu > 0$:

$$\begin{aligned} f(t) &= \mu e^{-\mu(t-\eta)} e^{-e^{-\mu(t-\eta)}} \\ P(t \geq \varepsilon) = F(t) &= e^{-e^{-\mu(t-\eta)}} \end{aligned}$$

Discrete choice models

Logit model:

$$P_n(i|\mathcal{C}_n) = \frac{e^{\mu V_{in}}}{\sum_{j \in \mathcal{C}_n} e^{\mu V_{jn}}}.$$

Parameters to be estimated:

- β_k parameters within V_{in}
- Scale parameter μ

Issues:

- Utility is not observed. It is latent. Cannot regress directly.
- Choice is observed: $y_{in} = 1$ if n actually chooses i, 0 otherwise.

Discrete choice models

Consequences:

- Maximum likelihood estimation:

$$\max_{\theta} \mathcal{L}(\theta) = \sum_{n=1}^{N} \left(\sum_{j=1}^{J} y_{jn} \ln P_n(j|\mathcal{C}_n;\theta) \right)$$

- One of the constants must be normalized to 0.

$$\Pr(U_{in} \geq U_{jn}) = \Pr(U_{in} + \alpha \geq U_{jn} + \alpha) \quad \forall \alpha.$$

- Scale μ not identified

$$\Pr(U_{in} \geq U_{jn}) = \Pr(\alpha U_{in} \geq \alpha U_{jn}) \quad \forall \alpha > 0.$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Example: Choice of Space-Water Heating System

Parameter number	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	-5.35	1.02	-5.26	0.00
2	-3.67	1.13	-3.24	0.00
3	0.410	0.672	0.61	0.54
4	-0.0740	0.488	-0.15	0.88
5	0.820	0.897	0.91	0.36
6	0.535	0.704	0.76	0.45
7	-6.06	0.443	-13.69	0.00
8	-2.60	0.402	-6.47	0.00
9	-10.8	0.489	-22.04	0.00
10	-2.90	0.642	-4.52	0.00
11	0.840	0.161	5.23	0.00
12	-0.173	0.182	-0.95	0.34
13	-0.576	0.249	-2.31	0.02
14	-0.769	0.175	-4.40	0.00
15	-0.669	0.126	-5.33	0.00
16	-0.652	0.222	-2.93	0.00
17	-0.593	0.179	-3.31	0.00
18	-0.804	0.0862	-9.33	0.00
19	-0.565	0.0961	-5.88	0.00

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Relax the independence assumption

- The red bus/blue bus paradox
- MEV models

Red bus/Blue bus paradox

- Mode choice example
- Two alternatives: car and bus
- There are red buses and blue buses
- Car and bus travel times are equal: T

Red bus/Blue bus paradox

Model 1

$$\begin{aligned} U_{\text{car}} &= \beta T + \varepsilon_{\text{car}} \\ U_{\text{bus}} &= \beta T + \varepsilon_{\text{bus}} \end{aligned}$$

Logit gives,

$$P(\text{car}|\{\text{car}, \text{bus}\}) = P(\text{bus}|\{\text{car}, \text{bus}\}) = \frac{e^{\beta T}}{e^{\beta T} + e^{\beta T}} = \frac{1}{2}$$

Red bus/Blue bus paradox

Model 2

$$\begin{aligned} U_{\text{car}} &= \beta T + \varepsilon_{\text{car}} \\ U_{\text{blue bus}} &= \beta T + \varepsilon_{\text{blue bus}} \\ U_{\text{red bus}} &= \beta T + \varepsilon_{\text{red bus}} \end{aligned}$$

$$P(\text{car}|\{\text{car}, \text{blue bus}, \text{red bus}\}) = \frac{e^{\beta T}}{e^{\beta T} + e^{\beta T} + e^{\beta T}} = \frac{1}{3}$$

$$\left.\begin{array}{r} P(\text{car}|\{\text{car}, \text{blue bus}, \text{red bus}\}) \\ P(\text{blue bus}|\{\text{car}, \text{blue bus}, \text{red bus}\}) \\ P(\text{red bus}|\{\text{car}, \text{blue bus}, \text{red bus}\}) \end{array}\right\} = \frac{1}{3}.$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Red bus/Blue bus paradox

- Assumption of logit: ε i.i.d
- $\varepsilon_{\text{blue bus}}$ and $\varepsilon_{\text{red bus}}$ contain common unobserved attributes:
  - fare
  - headway
  - comfort
  - convenience
  - etc.
- Invalid assumption may lead to wrong results.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Example: Choice of Space-Water Heating System

	Water heating		
Space heating	Natural gas	Oil	Electricity
Natural gas	×		×
Dual energy		×	×
Oil		×	×
Electricity			×
Wood			×
Wood-electricity			×

Alternatives are likely to share common unobserved attributes

Relax the independence assumption

$$\begin{pmatrix} U_{1n} \\ \vdots \\ U_{Jn} \end{pmatrix} = \begin{pmatrix} V_{1n} \\ \vdots \\ V_{Jn} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1n} \\ \vdots \\ \varepsilon_{Jn} \end{pmatrix}$$

that is

$$U_n = V_n + \varepsilon_n$$

and ε_n is a vector of random variables.

Assumption about the random term:
multivariate distribution

Relax the independence assumption

- A multivariate random variable ε is represented by a density function

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

and

$$P(\varepsilon \leq x) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_J} f(\varepsilon) d\varepsilon_J \ldots d\varepsilon_1$$

where $x \in \mathbb{R}^J$ is a $J \times 1$ vector of constants.
- Main operational issue: the multifold integral.

Relax the independence assumption

- If the CDF $F(\varepsilon_1,\ldots,\varepsilon_J)$ of the distribution is known

$$f(\varepsilon_1,\ldots,\varepsilon_J)=\frac{\partial^J F}{\partial\varepsilon_1\cdots\partial\varepsilon_J}(\varepsilon_1,\ldots,\varepsilon_J)$$

- The choice probability is

$$\begin{aligned}P(i) &= \Pr(V_1+\varepsilon_1\leq V_i+\varepsilon_i,\ldots,V_J+\varepsilon_J\leq V_i+\varepsilon_i)\\ &= \Pr(\varepsilon_1\leq V_i+\varepsilon_i-V_1,\ldots,\varepsilon_J\leq V_i+\varepsilon_i-V_J)\\ &= \int_{\varepsilon_i=-\infty}^{\infty} F_i(V_i+\varepsilon_i-V_1,\ldots,\varepsilon_i,\ldots,V_i+\varepsilon_i-V_J)d\varepsilon_i\end{aligned}$$

where $F_i=\partial F/\partial\varepsilon_i$.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

MEV models

Family of models proposed by McFadden (1978) (called GEV)
Idea: a model is generated by a function

$$G : \mathbb{R}_+^J \rightarrow \mathbb{R}_+$$

From G, we can build

- The cumulative distribution function (CDF)
- The probability model
- The expected maximum utility

MEV models

Technical sufficient conditions:

1. G is homogeneous of degree $\mu > 0$, that is

$$G(\alpha y) = \alpha^\mu G(y)$$

2. $\lim_{y_i \to +\infty} G(y_1, \ldots, y_i, \ldots, y_J) = +\infty$, for each $i = 1, \ldots, J$,
3. the kth partial derivative with respect to k distinct y_i is non negative if k is odd and non positive if k is even, i.e., for all (distinct) indices $i_1, \ldots, i_k \in \{1, \ldots, J\}$, we have

$$(-1)^k \frac{\partial^k G}{\partial y_{i_1} \ldots \partial y_{i_k}}(y) \leq 0, \ \forall y \in \mathbb{R}^J_+.$$

MEV models

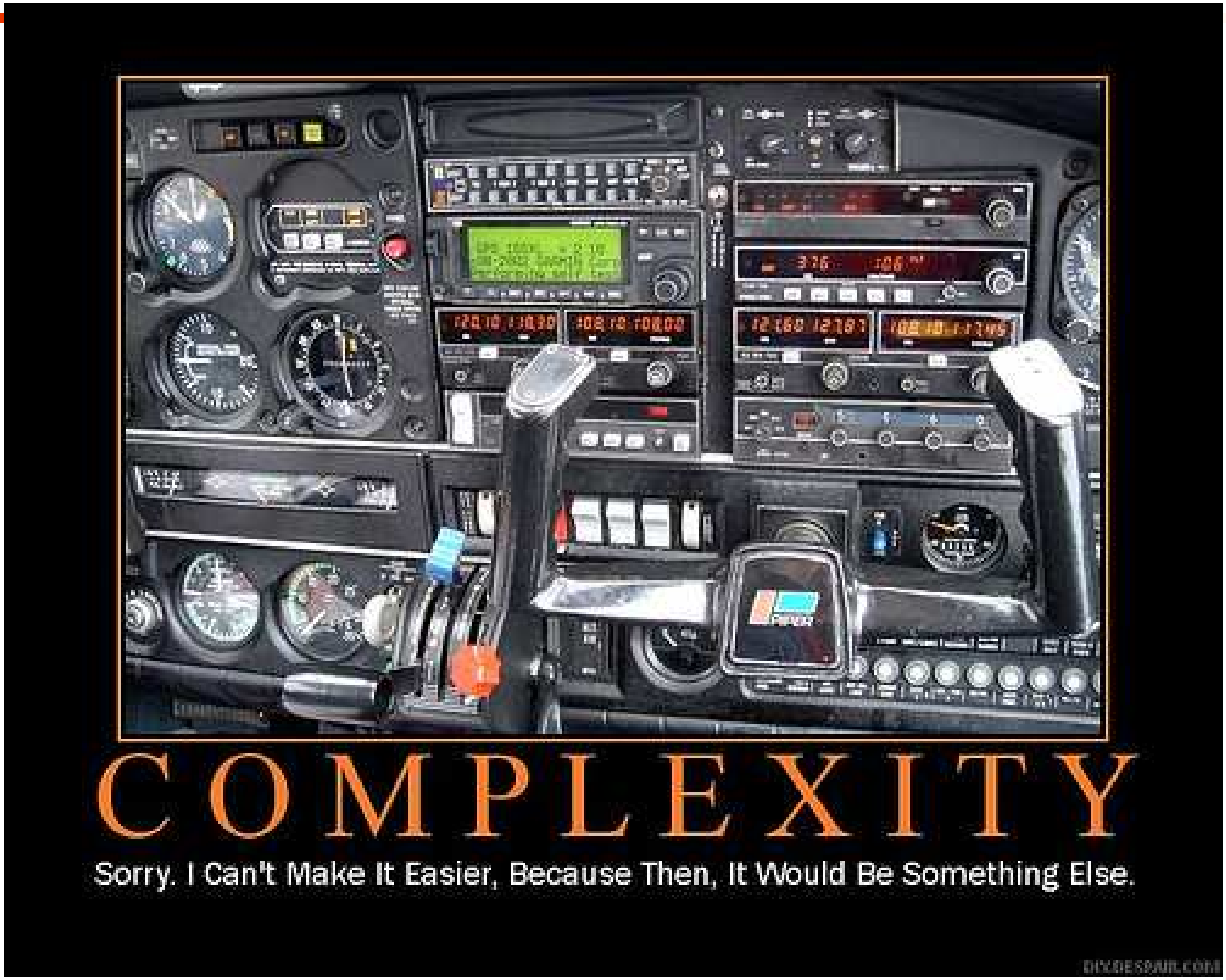
- CDF: $F(\varepsilon_1,\ldots,\varepsilon_J) = e^{-G(e^{-\varepsilon_1},\ldots,e^{-\varepsilon_J})}$
- Probability: $P(i|C) = \frac{e^{V_i+\ln G_i(e^{V_1},\ldots,e^{V_J})}}{\sum_{j\in C} e^{V_j+\ln G_j(e^{V_1},\ldots,e^{V_J})}}$ with $G_i = \frac{\partial G}{\partial x_y}$. This is a closed form
- Expected maximum utility: $V_C = \frac{\ln G(\ldots)+\gamma}{\mu}$ where γ is Euler's constant.
- Note: $P(i|C) = \frac{\partial V_C}{\partial V_i}$.
- Euler's constant

$$\gamma = -\int_0^{+\infty} e^{-x}\ln x dx = \lim_{n\to\infty}\left(\sum_{k=1}^{n}\frac{1}{k} - \ln n\right) \approx 0.577215665$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Complicated...



TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

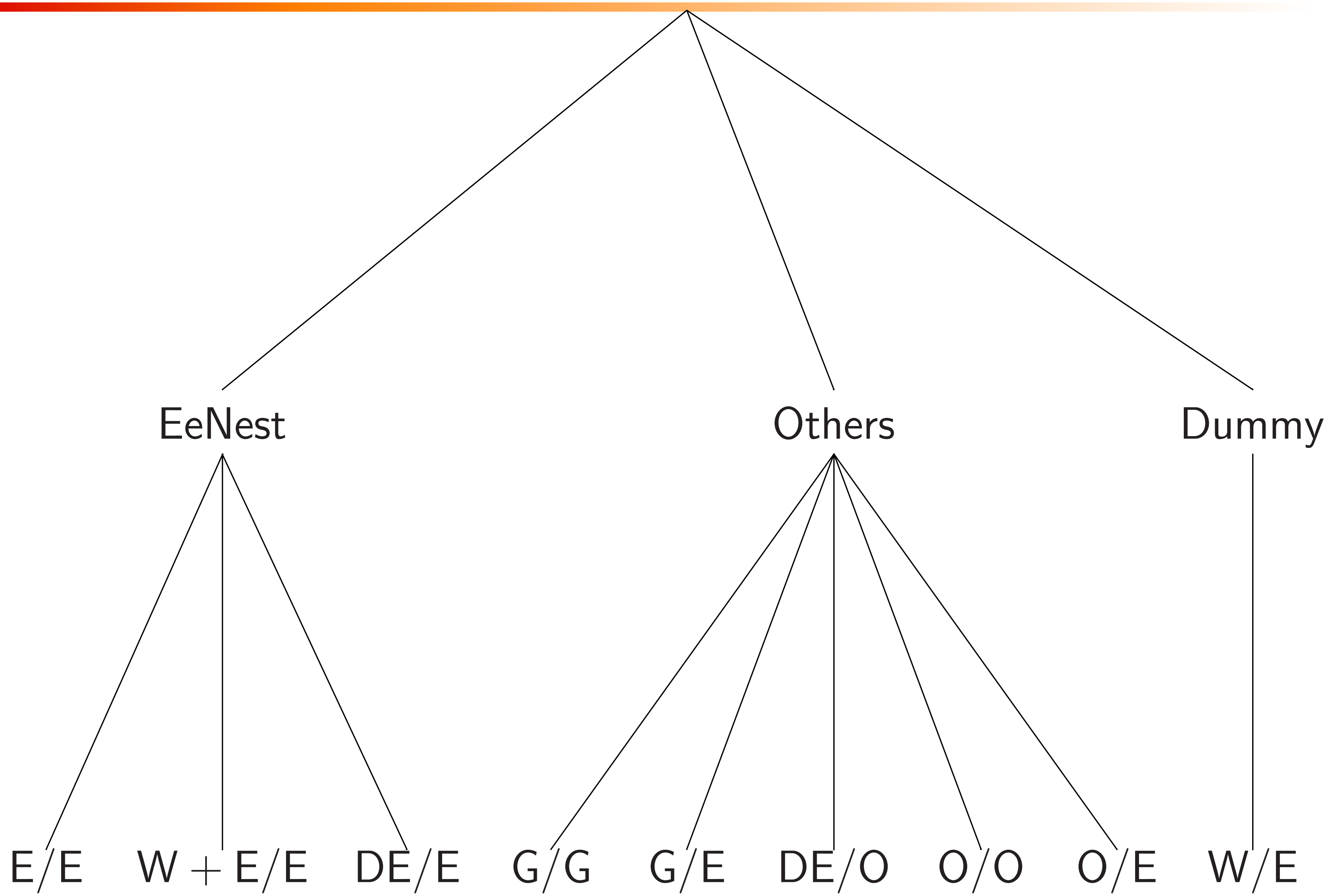
Example: the nested logit model

- Group the alternatives into M nests
- Nest m contains J_m alternatives potentially correlated
- Define the G function as

$$G(y) = \sum_{m=1}^{M} \left(\sum_{i=1}^{J_m} y_i^{\mu_m} \right)^{\frac{\mu}{\mu_m}}$$

- Ratio μ/μ_m captures the correlation among alternatives in nest m

Example: the nested logit model



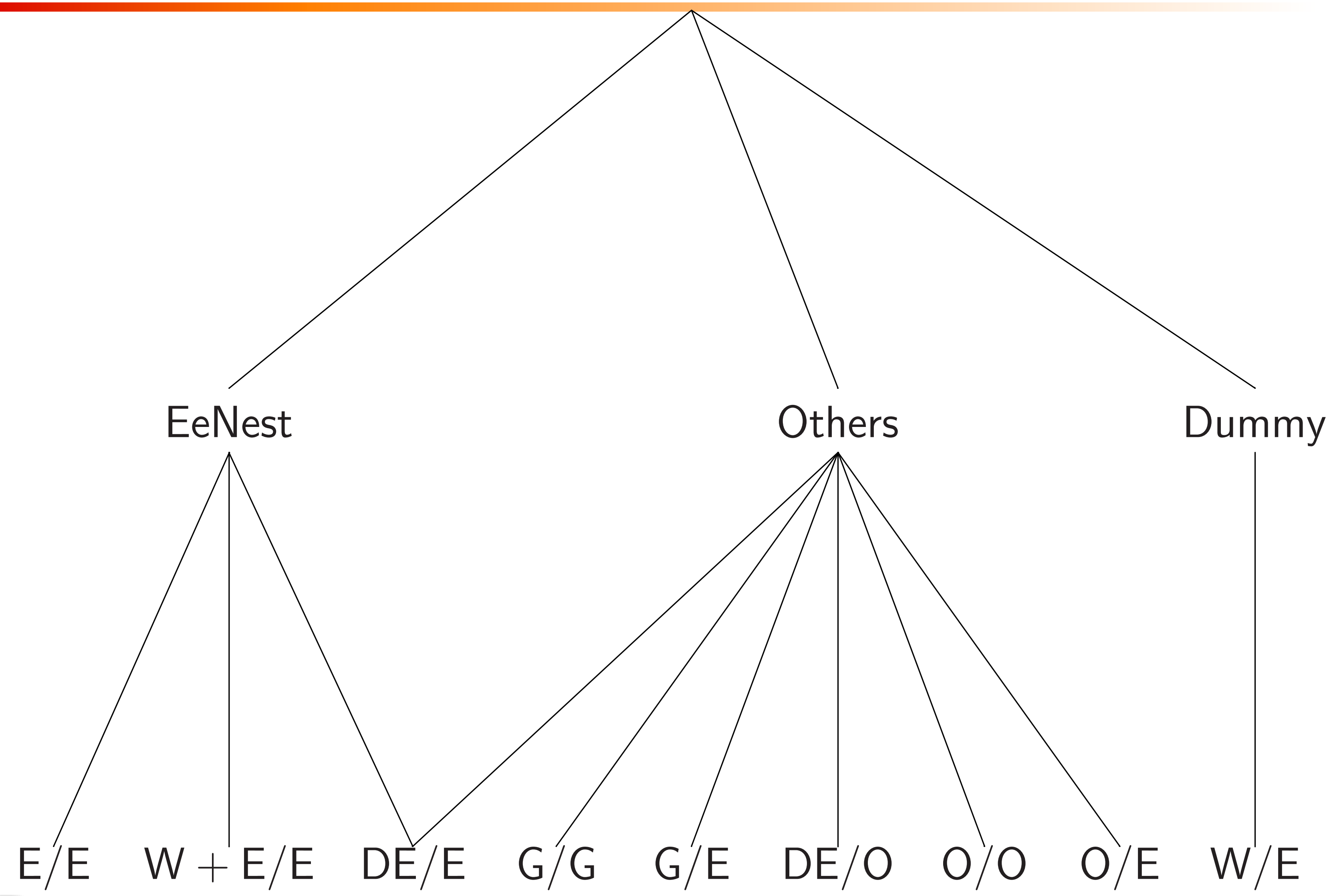
Example: the cross-nested logit model

- Group the alternatives into M nests
- Define the G function as

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

- α_{jm} defines the "degree of membership"
- Ratio μ/μ_m captures the correlation among alternatives in nest m

Example: the cross-nested logit model



Properties

- Closed form model.
- Logit-like formulation.
- Inheritance theorems.

Impact on practice:

- Software
- Choice-based samples
- Large choice sets
- Design of new MEV models

Biogeme

- Closed form model.
- Log likelihood has also a closed form.
- Easy to implement and fast to compute.

biogeme.epfl.ch

- Free software, open source.
- Designed to estimate the parameters of MEV models.

Choice-based samples

- Simple random samples are not available in practice
- Choice-based samples are convenient
- Example:
  - Hydroquébec have easier access to its clients than to its non-clients.
  - The proportion of electricity users in the sample likely to be larger than in the population.
  - It may introduce significant bias in the estimates.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Choice-based samples

Manski & Lerman (1977)

- If logit is used, and the usual maximum likelihood estimator is used, all parameters are correctly estimated, except the constants.
- Contribution to the log likelihood function

$$\frac{e^{V_{in}+\delta_{in}}}{\sum_j e^{V_{jn}+\delta_{jn}}}$$

- The biases δ_{in} are confounded with the constants during estimation

Choice-based samples

Bierlaire, Bolduc & McFadden (2008)

- Idea: MEV models have a logit-like form:

$$P(i|C) = \frac{e^{V_i + \ln G_i(e^{V_1},\ldots,e^{V_J})}}{\sum_{j\in C} e^{V_j + \ln G_j(e^{V_1},\ldots,e^{V_J})}}$$

  with $G_i = \partial G/\partial y$.

- The same result may apply...

$$P(i|C) = \frac{e^{V_i + \ln G_i(e^{V_1},\ldots,e^{V_J}) + \delta_i}}{\sum_{j\in C} e^{V_j + \ln G_j(e^{V_1},\ldots,e^{V_J}) + \delta_j}}$$

Choice-based samples

But...

$$P(i|C) = \frac{e^{V_i+\delta_i+\ln G_i(e^{V_1},\ldots,e^{V_J})}}{\sum_{j\in C} e^{V_j+\delta_j+\ln G_j(e^{V_1},\ldots,e^{V_J})}}$$

shifted not shifted

Usual estimator cannot be used

Choice-based samples

But...

$$P(i|C) = \frac{e^{V_i + \ln G_i(e^{V_1},\ldots,e^{V_J}) + \delta_i}}{\sum_{j \in C} e^{V_j + \ln G_j(e^{V_1},\ldots,e^{V_J}) + \delta_j}}$$

- Bias δ_i can be estimated from data.
- Not the original estimator, but the modification is minor and easy to implement.
- Available in Biogeme

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Large choice set

Examples:

- Route choice
- Destination choice
- House location choice
- etc.

Estimation:

- Alternatives can be sampled.
- Estimation can be performed with the sampled subset.
- It may introduce bias.
- Technically, it is handled exactly like for choice-based sampling.

See Bierlaire, Bolduc & McFadden (2008) for details

Inheritance theorems

- Sufficient conditions are technical and non-intuitive
- How do we derive a concrete generating function G from modeling assumptions?
- Solution: the network MEV model
- Theoretical foundation: inheritance theorems

Daly & Bierlaire (2006)

Network MEV

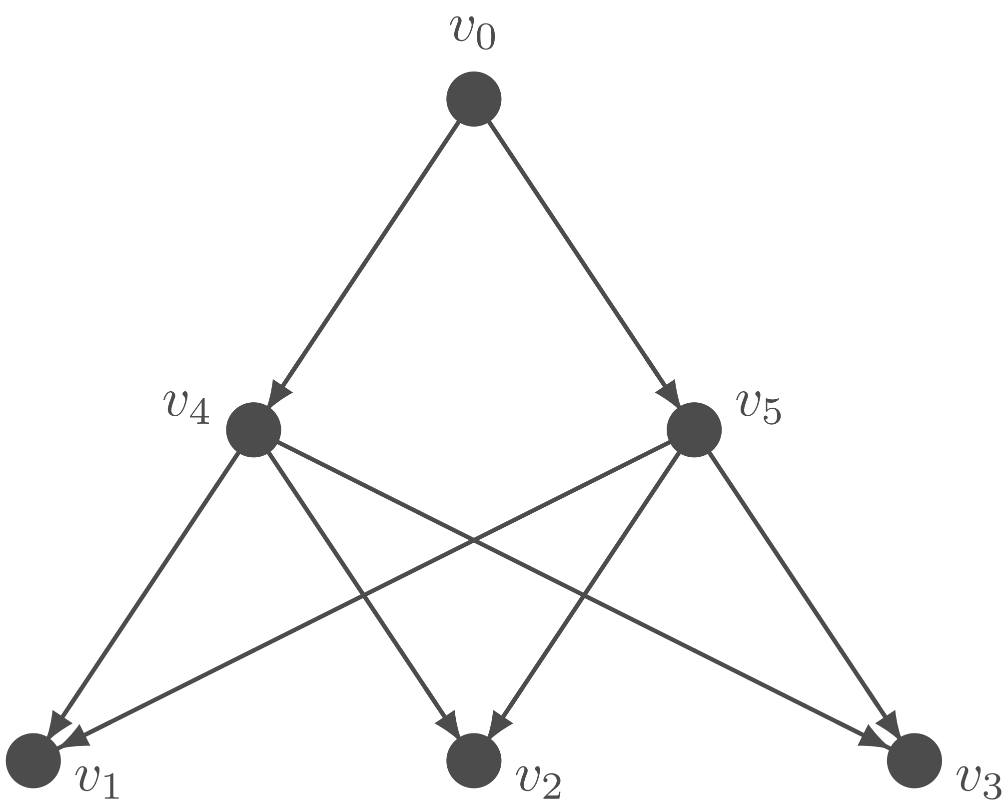
Let (V, E) be a network with link parameters $\alpha_{(i,j)} \geq 0$
Assumptions:

1. No circuit.
2. One node without predecessor: *root*.
3. J nodes without successor: *alternatives*.
4. For each node v_i, there exists at least one path from the root to v_i such that $\prod_{k=1}^{P} \alpha_{(i_{k-1}, i_k)} > 0$.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Network MEV



Network MEV

For each node v_i, we define

- a set of indices $I_i \subseteq \{1, \ldots, J\}$ of J_i relevant alternatives,
- a homogeneous function $G^i : \mathbb{R}^{J_i} \longrightarrow \mathbb{R}$, and
- a parameter μ_i.

Recursive definition of I_i:

- $I_i = \{i\}$ for alternatives,
- $I_i = \bigcup_{j \in \text{succ}(i)} I_j$ for all other nodes.

Network MEV

Recursive definition of G^i:

For alternatives:

$$G^i \quad : \quad \mathbb{R} \longrightarrow \mathbb{R} \quad : \quad G^i(y_i) = y_i^{\mu_i} \quad i = 1, \ldots, J$$

For all others:

$$G^i \quad : \quad \mathbb{R}^{J_i} \longrightarrow \mathbb{R} \quad : \quad G^i(y) = \sum_{j \in \text{succ}(i)} \alpha_{(i,j)} G^j(y)^{\frac{\mu_i}{\mu_j}}$$

Theorem

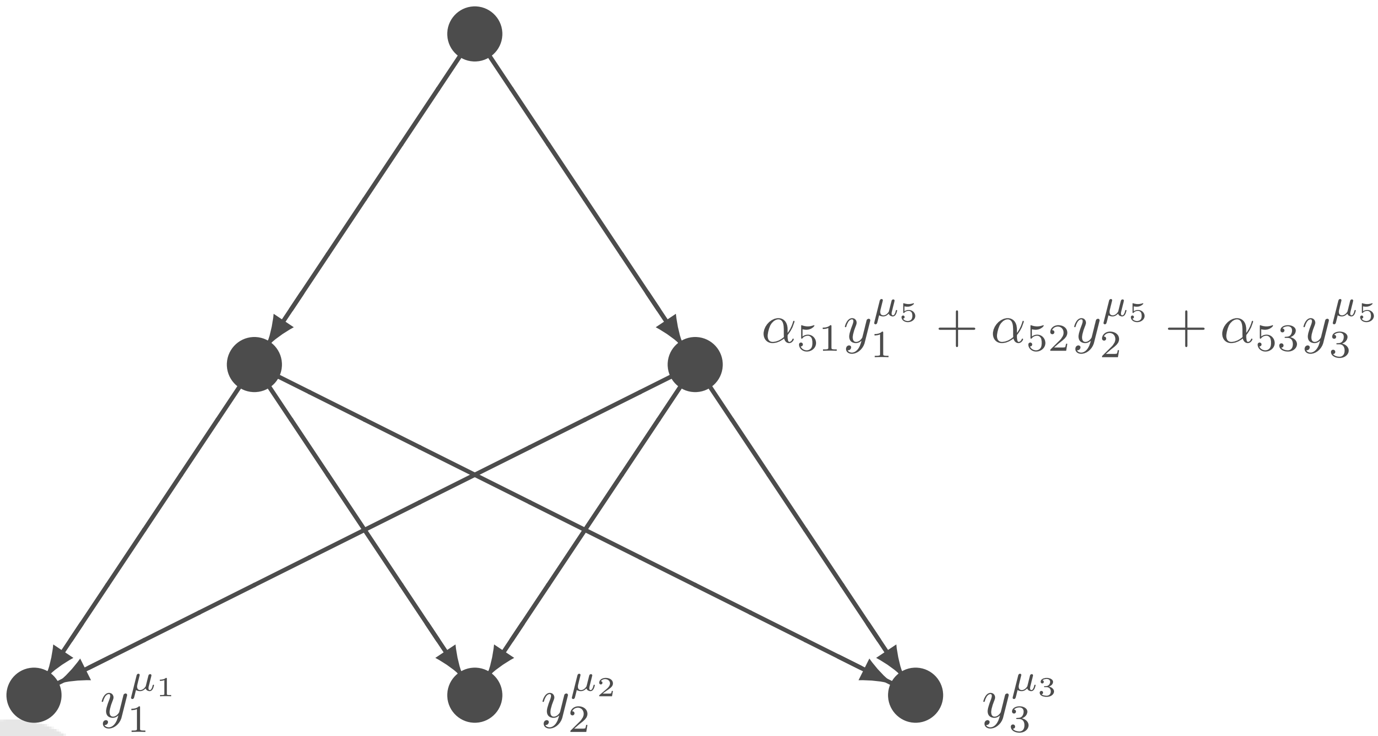
If all $G^j(y)$ are MEV generating functions, so is G^i

Network MEV

Example: Cross-Nested Logit

$$G = \sum_m \left(\sum_{j \in C} \alpha_{jm} y_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}}$$

$\sum_{i=4,5} \alpha_{0i} \left(\alpha_{i1} y_1^{\mu_i} + \alpha_{i2} y_2^{\mu_i} + \alpha_{i3} y_3^{\mu_i}\right)^{\frac{\mu_0}{\mu_i}}$



Network MEV

Similar idea: Daly (2001) *Recursive Nested EV Model*
Advantages :

- Easy to design
- No more proof necessary

Conclusion

- Discrete choice models are more and more used for disaggregate behavioral analysis
- The family of MEV models relaxes the independence assumption of the logit model
- Convenient due to close form (biogeme).
- Logit-like formulation enables to derive simple estimators for choice-based samples and sampling of alternatives.
- Inheritance theorems and the network MEV model enable practitioners to derive appropriate models without dealing with the technical burden.