BIOGEME: an open source package for the estimation of advanced discrete choice models

transp-or.epfl.ch

Prof. Michel Bierlaire

Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne

Introduction

- Transportation demand
- Disaggregate representation
- Focus on the traveler
- Modeling individual decisions such as
  - Choice of housing location
  - Choice of activity participation and location
  - Choice of transportation mode
  - Choice of company, carrier
  - Choice of itinerary
  - Choice of parking

Discrete Choice

- Discrete choice models:

$$P(i|\mathcal{C}_n) \text{ where } \mathcal{C}_n = \{1, \ldots, J\}$$

- Random utility models:

$$U_{in} = V_{in} + \varepsilon_{in}$$

  and

$$P(i|\mathcal{C}_n) = P(U_{in} \geq U_{jn}, j = 1, \ldots, J)$$

- Utility is a latent concept

Discrete Choice

$$V_{in} = V(z_{in}, S_n)$$

where

- z_{in} is a vector of attributes of alternative i for individual n
- S_n is a vector of socio-economic characteristics of n

Modeling assumptions:

- Functional form
- Definition of z_{in} and S_n

Notation:

$$x_{in} = (z_{in}, S_n)$$

In general, linear-in-parameters utility functions are used

$$V_{in} = V(z_{in}, S_n) = V(x_{in}) = \sum_p \beta_p (x_{in})_p$$

Multinomial Logit Model

- Assumption: ε_{in} are i.i.d. Extreme Value distributed.
- Independence is both across i and n
- Choice model: MNL

$$P(i|\mathcal{C}_n) = \frac{e^{V_{in}}}{\sum_{j \in \mathcal{C}_n} e^{V_{jn}}} = \frac{e^{\sum_p \beta_p (x_{in})_p}}{\sum_{j \in \mathcal{C}_n} e^{\sum_p \beta_p (x_{jn})_p}}$$

- Maximum likelihood estimation of β_p is easy
- Available in a great deal of software packages
- Intensively used in many transportation applications
- But... may be inappropriate

MNL inappropriate

MNL with linear-in-parameter V's is not always adequate

- Correlation among alternatives
  - Red bus / blue bus “paradox”
  - Route choice
- Heterogeneity in the population
  - β varies stochastically
  - β varies deterministically
  - Scale (variance of ε_{in}) varies across n

Relaxing the independence assumption

...across alternatives

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

that is

$$U_n = V_n + \varepsilon_n$$

and ε_n is a vector of random variables.

Relaxing the independence assumption

- $\varepsilon_n \sim N(0, \Sigma)$: multinomial probit model
  - No closed form for the multifold integral
  - Numerical integration is computationally infeasible
- Extensions of multinomial logit model
  - Nested logit model
  - Multivariate Extreme Value (MEV) models

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

MEV models

Family of models proposed by McFadden (1978)
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) of ε_n
- The probability model
- The expected maximum utility

Called Generalized EV models in DCM community

MEV models

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

$$G(\alpha x) = \alpha^{\mu} G(x)$$

2. $\lim_{x_i \to +\infty} G(x_1, \ldots, x_i, \ldots, x_J) = +\infty$, $\forall i$,

3. the kth partial derivative with respect to k distincts x_i is non negative if k is odd and non positive if k is even, i.e., for all (disctincts) indices $i_1, \ldots, i_k \in \{1, \ldots, J\}$, we have

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

MEV models

- Density function: $F(\varepsilon_1, \ldots, \varepsilon_J) = e^{-G(e^{-\varepsilon_1}, \ldots, e^{-\varepsilon_J})}$
- 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 $G_i = \frac{\partial G}{\partial x_i}$. This is a closed form
- Expected maximum utility: $V_C = \frac{\ln G(\ldots) + \gamma}{\mu}$ where γ is Euler's constant.
- Note: $P(i|C) = \frac{\partial V_C}{\partial V_i}$.

MEV models

Example: Multinomial logit:

$$G(e^{V_1}, \ldots, e^{V_J}) = \sum_{i=1}^{J} e^{\mu V_i}$$

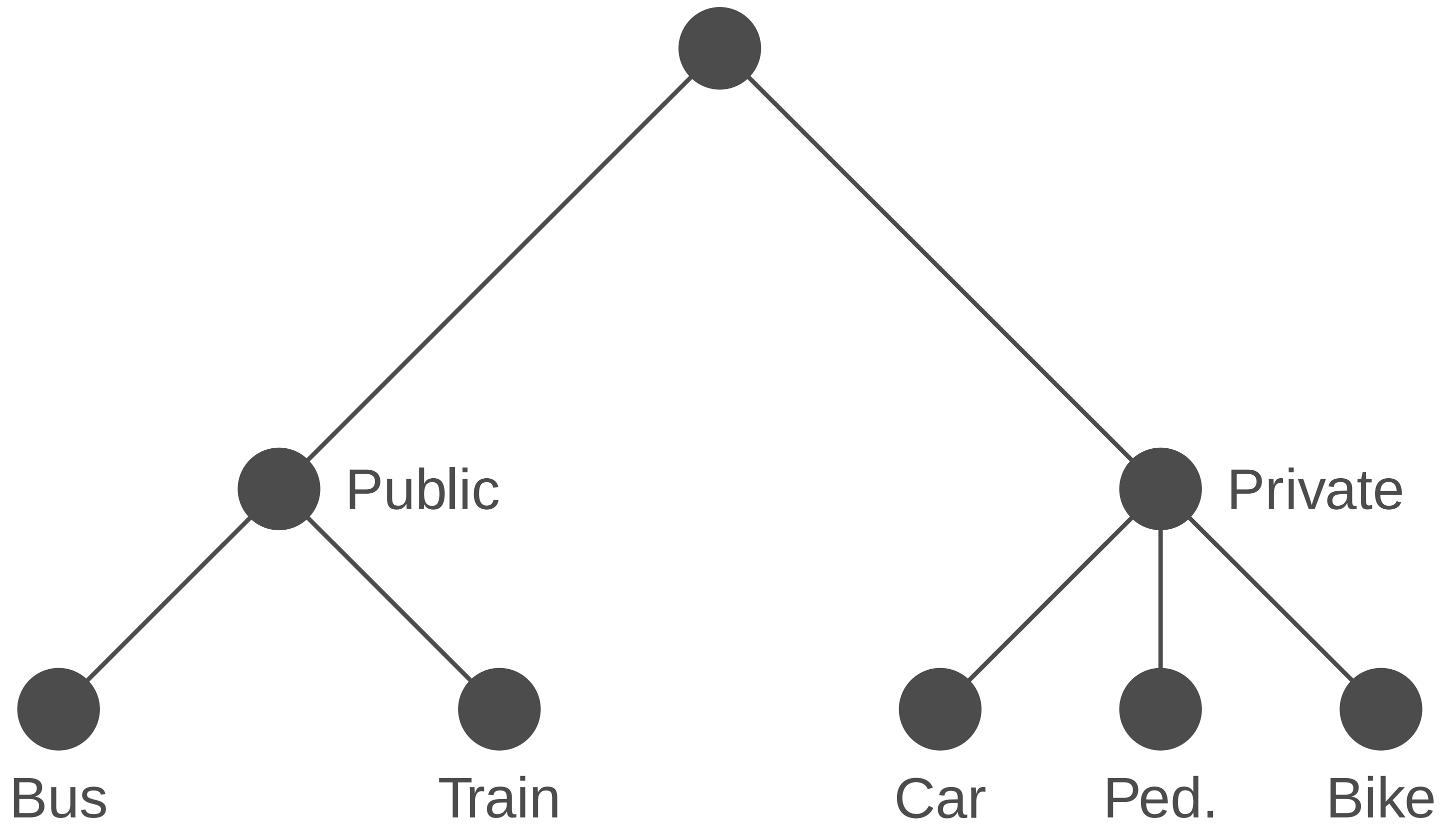
Example: Nested logit

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

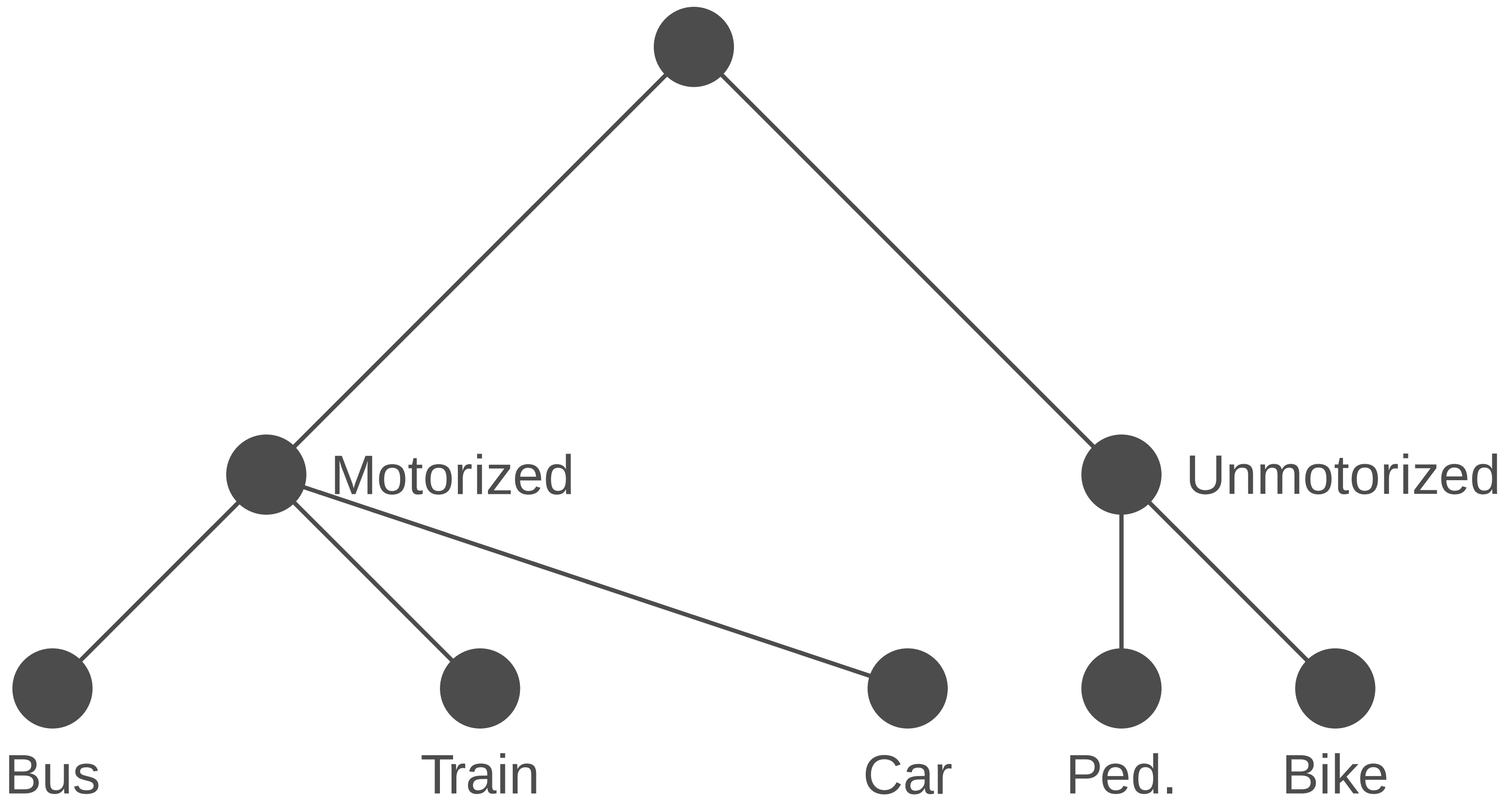
Example: Cross-Nested Logit

$$G(y_1, \ldots, y_J) = \sum_{m=1}^{M} \left(\sum_{j \in \mathcal{C}} (\alpha_{jm}{}^{1/\mu} y_j)^{\mu_m} \right)^{\frac{\mu}{\mu_m}}$$

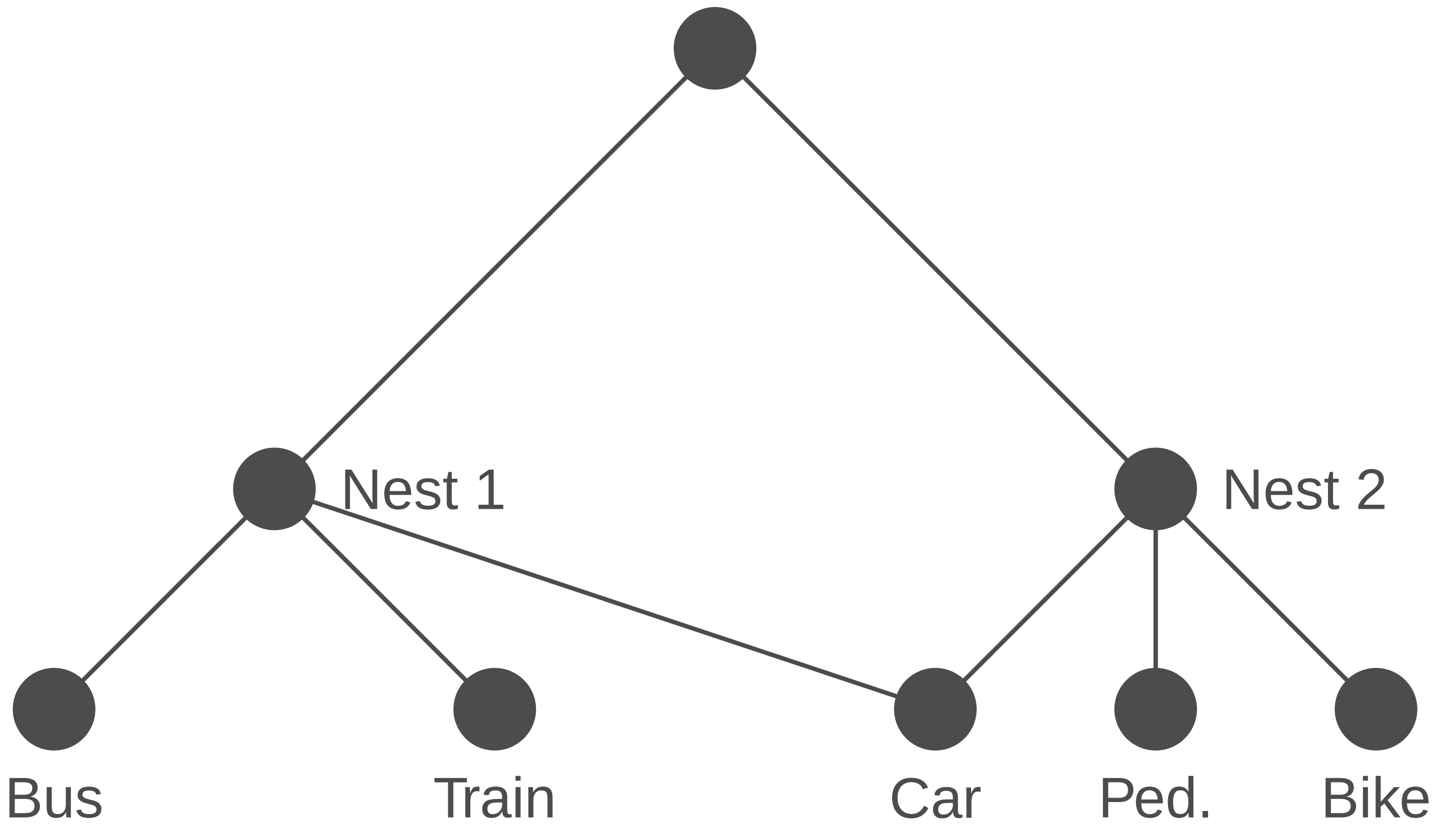
Nested Logit Model



Nested Logit Model



Cross-Nested Logit Model



MEV models

Advantages:

- Closed form probability model
- Provides a great deal of flexibility

Many transportation applications, including:

- Mode choice
- Departure time choice
- Destination choice

MNL inappropriate

MNL with linear-in-parameter V's is not always adequate

- *Correlation among alternatives*
  - *Red bus / blue bus "paradox"*
  - *Route choice*
- Heterogeneity in the population
  - β varies stochastically
  - β varies deterministically
  - Scale (variance of ε_{in}) varies across n

Mixture of MEV

In statistics, a mixture density is a pdf which is a convex linear combinations of other pdf's.
If $f(\varepsilon, \theta)$ is a pdf, and if $w(\theta)$ is a nonnegative function such that $\int_a w(a)da = 1$ then

$$g(\varepsilon) = \int_a w(a) f(\varepsilon, \theta) da$$

is also a pdf. We say that g is a mixture of f.

If f is the pdf of a MEV model, it is a mixture of MEV

Mixture of MEV

Discrete mixtures are also possible. If $f(\varepsilon, \theta)$ is a pdf, and if w_i, $i = 1, \ldots, n$ are nonnegative weights such that $\sum_{i=1}^{n} w_i = 1$ then

$$g(\varepsilon) = \sum_{i=1}^{n} w_i f(\varepsilon, \theta_i)$$

is also a pdf. We say that g is a discrete mixture of f.

Mixture of MEV

$$U_{in} = V_{in} + \varepsilon_{in}$$

- ε_{in} compliant with MEV theory
- V_{in} contains a random parameter.

$$V_{in} = \beta x_{i1} + \cdots \text{ where } \beta \sim N(\widehat{\beta}, \sigma)$$

- Note that

$$\beta = \widehat{\beta} + \sigma\zeta$$

and ζ is standard normal $N(0, 1)$.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Mixture of MEV

Derivation for MNL:
If β was known:

$$P(i|\beta) = \frac{e^{\beta x_{i1}+\cdots}}{\sum_j e^{\beta x_{j1}+\cdots}}$$

But β is distributed

$$P(i) = \int_\beta P(i|\beta) f(\beta) d\beta$$

It is a mixture of MNL.
Integral requires numerical simulation.

Exact same derivation for MEV

Mixture of MEV

- Very popular during the last 5-10 years
- Many applications in transportation
- Almost always mixture of MNL

Parameter varies deterministically

Example:

$$V = \beta \text{ income}^{\lambda} \text{ cost } + \cdots$$

- Cost parameter varies with income
- One additionnal parameter to estimate
- Used namely in the Swiss Value of Time study
- Nonlinear utility function

Scale

Scale unidentified with i.i.d. assumptions
Two methods to relax the identical scale across n:

- Estimate the scale for subgroups of the population: nonlinear

$$\mu V_{in} = \mu \sum_p \beta_p (x_{in})_p$$

- Multiplicative error terms (Fosgerau & Bierlaire, 2007)

$$\begin{aligned} U_{in} &= V_{in}\varepsilon_{in} \\ \ln U_{in} &= \ln V_{in} + \ln \varepsilon_{in} \end{aligned}$$

Summary

- MEV models
- Mixtures
- Nonlinear utility

Relaxing the assumptions comes with increased complexity

Short course

Lausanne, March 25-29, 2008

Ben-Akiva, McFadden, Bierlaire, Bolduc
+ Walker

http://transp-or.epfl.ch/dca

BIOGEME

Motivations

- No appropriate software package
- Most researchers use commercial packages: LIMDEP, ALOGIT, HieLoW or Gauss, Matlab, SAS
- Freeware: Kenneth Train (but based on Gauss)

BIOGEME

Objectives

- Maximum likelihood estimation of a wide variety of MEV models
- Complex formulations
- Use various nonlinear optimization algorithms
- Open source
- Designed for researchers, but also used by practitioners
- Flexible and easily extensible

`biogeme.epfl.ch`

Discussion

- Modern transportation issues require advanced models
- Examples presented today:
  - Scheduler for airlines
  - Traffic simulation
  - Estimation of choice models
- Practitioners and researchers need complex software

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Discussion

- Development costs can be very high
- The more complex the models, the higher the cost, the smaller the potential market
- Commercial software or free software?
  - Quality control
  - User friendliness
  - Support and documentation
- Who should cover the development cost?
- Can free software generate revenues?