Estimation of Random Utility Models in R: The mlogit Package

Yves Croissant
Université de la Réunion

Abstract

mlogit is a package for R which enables the estimation of random utility models with individual and/or alternative specific variables. The main extensions of the basic multinomial model (heteroscedastic, nested and random parameter models) are implemented.

Keywords: discrete choice models, maximum likelihood estimation, R, econometrics.

Introduction

Random utility models is the reference approach in economics when one wants to analyze the choice by a decision maker of one among a set of mutually exclusive alternatives. Since the seminal works of Daniel Mc Fadden (McFadden 1974) who won the Nobel prize in economics "for his development of theory and methods for analyzing discrete choice", a large amount of theoretical and empirical literature have been developed in this field.[1]

These models rely on the hypothesis that the decision maker is able to rank the different alternatives by an order of preference represented by a utility function. For a choice of an alternative l among a set of J alternatives, each alternative is therefore characterized by an utility value U_l and the alternative l is chosen if and only if $U_l > U_j \; \forall \; j \neq l$.

These models are called random utility models because the researcher is unable to measure the whole level of utility, but only part of it. Therefore, the utility for alternative l is written as: $U_l = V_l + \epsilon_l$ where V_l is a function of some observable covariates and unknown parameters to be estimated, and ϵ_l a random deviate which contains all the unobserved determinants of the utility. Alternative l is therefore chosen if $\epsilon_j < (V_l - V_j) + \epsilon_l \; \forall \; j \neq l$ and this choice can be written in probabilistic terms:

$$\mathrm{P}(\epsilon_1 < V_l - V_1 + \epsilon_l, \epsilon_2 < V_l - V_2 + \epsilon_l, \ldots, \epsilon_J < V_l - V_J + \epsilon_l)$$

Different hypothesis on the distribution of ϵ lead to different flavors of random utility

[1] For an extensive presentation of this literature, see Train (2003) ; the theoretical parts of this paper draw heavily on Kenneth Train's book.

models. Early developments of these models were based on the hypothesis of identically and independent errors following a Gumbell distribution.[2] Much more general models have since been proposed, based on much less restrictive distribution hypothesis, and often estimated using simulations.

The first version of **mlogit** was posted in 2008, it was the first R package allowing the estimation of random utility models. Since then, other package have emerged (see Sarrias and Daziano 2017, page 4 for a survey of revelant R pakages). **mlogit** still provides the widests set of estimators for random utility models and, moreover, its syntax has been adopted by other R packages, especially by **gmnl** (Sarrias and Daziano 2017) and **mnlogit** (Hasan, Wang, and Mahani 2016) which, respectively, implements advanced mixed logit models and estimates efficiently multinomial logit models on large data sets.

The article is organized as follow. Section 1 explains how the usual formula-data and testing interface can be extended in order to describes in a very natural way the model to be estimated. Section 2 describe the landmark multinomial logit model. Section 3 and 4 present two important extensions of this basic model: section 3 presents models that relax the *iid* Gumbell hypothesis and section 4 introduces slope heterogeneity by considering some parameters as random. Section 5 concludes.

1. Data management, model description and testing

The formula-data interface is a critical advantage of the R software. It provides a practical way to describe the model to be estimated and to store data. However, the usual interface is not flexible enough to deal correctly with random utility models. Therefore, `mlogit` introduce tools to construct richer `data.frame`s and `formula`s.

1.1. Data management

mlogit is loaded using:

```
library("mlogit")
```

It comes with several data sets that we'll use to illustrate the features of the library. Data sets used for multinomial logit estimation concern some individuals, that make one or a sequential choice of one alternative among a set of mutually exclusive alternatives. The determinants of these choices are covariates that can depends on the alternative and the choice situation, only on the alternative or only on the choice situation.

Such data have therefore a specific structure that can be characterized by three indexes: the alternative, the choice situation and the individual. These three indexes will be denoted `alt`, `chid` and `id`. Note that the distinction between `chid` and `id` is only relevant if we have repeated observations for the same individual.

[2] This distribution has the distinctive advantage that it leads to a probability which can be written has an integral which has a closed form.

Data sets can have two different shapes: a *wide* shape (one row for each choice situation) or a *long* shape (one row for each alternative and, therefore, as many rows as there are alternatives for each choice situation).

mlogit deals with both format. It provides a mlogit.data function that take as first argument a data.frame and returns a data.frame in "long" format with some supplementary information about the structure of the data.

Wide format

Train[3] is an example of a *wide* data set:

```
data("Train", package = "mlogit")
head(Train, 3)

##   id choiceid choice price_A time_A change_A comfort_A price_B
## 1  1        1      A    2400    150        0         1    4000
## 2  1        2      A    2400    150        0         1    3200
## 3  1        3      A    2400    115        0         1    4000
##   time_B change_B comfort_B
## 1    150        0         1
## 2    130        0         1
## 3    115        0         0
```

This data set contains data about a stated preference survey in Netherlands. Each individual has responded to several (up to 16) scenarios. For every scenario, two train trips are proposed to the user, with different combinations of four attributes: price (the price in cents of guilders), time (travel time in minutes), change (the number of changes) and comfort (the class of comfort, 0, 1 or 2, 0 being the most comfortable class).

This "wide" format is suitable to store individual specific variables. Otherwise, it is cumbersome for alternative specific variables because there are as many columns for such variables that there are alternatives.

For such a wide data set, the shape argument of mlogit.data is mandatory, as it default value is "long". The alternative specific variables are indicated with the varying argument which is a numeric vector that indicates their position in the data frame. This argument is then passed to reshape that coerced the original data.frame in "long" format. Further arguments may be passed to reshape. For example, as the names of the variables are of the form price_A, one must add sep = '_' (the default value being "."). The choice argument is also mandatory because the response has to be transformed in a logical value in the long format. To take the panel dimension into account, one has to add an argument id.var which is the name of the individual index.

[3] used by Ben-Akiva, Bolduc, and Bradley (1993) and Meijer and Rouwendal (2006).

```
Tr <- mlogit.data(Train, shape = "wide", choice = "choice",
                  varying = 4:11, sep = "_",
                  alt.levels = c("A", "B"), id.var = "id",
                  opposite = c("price", "comfort", "time", "change"))
```

Note the use of the `opposite` argument for the 4 covariates: we expect negative coefficients for all of these, taking the opposite of the covariates will lead to expected positive coefficients. We next convert price and time in more meaningful unities, hours and euros (1 guilder was 2.20371 euros):

```
Tr$price <- Tr$price / 100 * 2.20371
Tr$time <- Tr$time / 60
```

```
head(Tr, 3)

##     id choiceid choice alt     price time change comfort chid
## 1.A  1        1   TRUE   A -52.88904 -2.5      0      -1    1
## 1.B  1        1  FALSE   B -88.14840 -2.5      0      -1    1
## 2.A  1        2   TRUE   A -52.88904 -2.5      0      -1    2
```

An index attribute is added to the data, which contains the two relevant indexes: `chid` is the choice situation index and `alt` the alternative index. As a `id.var` is provided, the index contains a third column, the individual index. This attribute is a `data.frame` that can be extracted using the index function, which returns it:this `data.frame`.

```
head(index(Tr), 3)

##     chid alt id
## 1.A    1   A  1
## 1.B    1   B  1
## 2.A    2   A  1
```

Long format

ModeCanada,[4] is an example of a data set in long format. It presents the choice of individuals for a transport mode for the Montreal-Toronto corridor:

[4] Used in particular by Forinash and Koppleman (1993), Bhat (1995), Koppelman and Wen (1998) and Koppelman and Wen (2000).

```
data("ModeCanada", package = "mlogit")
head(ModeCanada)

##   case   alt choice dist  cost ivt ovt freq income urban noalt
## 1    1 train      0   83 28.25  50  66    4     45     0     2
## 2    1   car      1   83 15.77  61   0    0     45     0     2
## 3    2 train      0   83 28.25  50  66    4     25     0     2
## 4    2   car      1   83 15.77  61   0    0     25     0     2
## 5    3 train      0   83 28.25  50  66    4     70     0     2
## 6    3   car      1   83 15.77  61   0    0     70     0     2
```

There are four transport modes (air, train, bus and car) and most of the variable are alternative specific (cost for monetary cost, ivt for in vehicle time, ovt for out of vehicle time, freq for frequency). The only individual specific variables are dist (the distance of the trip), income (household income), urban (a dummy for trips which have a large city at the origin or the destination) and noalt the number of available alternatives. The advantage of this shape is that there are much fewer columns than in the wide format, the caveat being that values of dist, income and urban are repeated four times.

For data in "long" format, the `shape` and the `choice` arguments are no more mandatory.

To replicate published results (latter in the text), we'll use only a subset of the choice situations, namely those for which the 4 alternatives are available. This can be done using the subset function with the `subset` argument set to `noalt == 4`. This can also be done within mlogit.data, using the `subset` argument.

The information about the structure of the data can be explicitly indicated using choice situations and alternative indexes (respectively `case` and `alt` in this data set) or, in part, guessed by the mlogit.data function. Here, after subsetting, we have 2779 choice situations with 4 alternatives, and the rows are ordered first by choice situation and then by alternative (`train`, `air`, `bus` and `car` in this order).

The first way to read correctly this data frame is to ignore completely the two index variables. In this case, the only supplementary argument to provide is the `alt.levels` argument which is a character vector that contains the name of the alternatives in their order of appearance:

```
MC <- mlogit.data(ModeCanada, subset = noalt == 4,
                  alt.levels = c("train", "air", "bus", "car"))
```

Note that this can only be used if the data set is "balanced", which means than the same set of alternatives is available for all choice situations. It is also possible to provide an argument `alt.var` which indicates the name of the variable that contains the alternatives

```
MC <- mlogit.data(ModeCanada , subset = noalt == 4,
                  alt.var = "alt")
```

The name of the variable that contains the information about the choice situations can be indicated using the `chid.var` argument:

```
MC <- mlogit.data(ModeCanada, subset = noalt == 4,
                  chid.var = "case",
                  alt.levels = c("train", "air", "bus", "car"))
```

Both alternative and choice situation variable can also be provided:

```
MC <- mlogit.data(ModeCanada, subset = noalt == 4,
                 chid.var = "case", alt.var = "alt")
```

and dropped from the data frame using the `drop.index` argument:

```
MC <- mlogit.data(ModeCanada, subset = noalt == 4,
                 chid.var = "case", alt.var = "alt",
                 drop.index = TRUE)
head(MC)

##           choice dist   cost ivt ovt freq income urban noalt
## 109.train      0  377  58.25 215  74    4     45     0     4
## 109.air        1  377 142.80  56  85    9     45     0     4
## 109.bus        0  377  27.52 301  63    8     45     0     4
## 109.car        0  377  71.63 262   0    0     45     0     4
## 110.train      0  377  58.25 215  74    4     70     0     4
## 110.air        1  377 142.80  56  85    9     70     0     4
```

1.2. Model description

Standard `formulas` are not very practical to describe random utility models, as these models may use different sets of covariates. Working with random utility models, one has to consider at most five sets of covariates:

- alternative and individual specific covariates x_{ij} with generic coefficients β,
- individual specific covariates z_i with alternative specific coefficients γ_j,
- alternative and individual specific covariates w_{ij} with alternative specific coefficients δ_j,

- alternative specific covariates t_j with a generic coefficient ν,
- individual specific covariates v_i that influence the variance of the errors.

Ignoring for the moment the 5th set of covariates, the observable part of the utility index for alternative j is:

$$V_{ij} = \alpha_j + \beta x_{ij} + \gamma_j z_i + \delta_j w_{ij} + \nu t_j$$

As the absolute value of utility is irrelevant, only utility differences are useful to modelize the choice for one alternative. For two alternatives j and k, we obtain:

$$V_{ij} - V_{ik} = (\alpha_j - \alpha_k) + \beta(x_{ij} - x_{ik}) + (\gamma_j - \gamma_k) z_i + (\delta_j w_{ij} - \delta_k w_{ik}) + \nu(t_j - t_k)$$

It is clear from the previous expression that coefficients of individual specific variables (the intercept being one of those) should be alternative specific, otherwise they would disappear in the differentiation. Moreover, only differences of these coefficients are relevant and may be identified. For example, with three alternatives 1, 2 and 3, the three coefficients $\gamma_1, \gamma_2, \gamma_3$ associated to an individual specific variable cannot be identified, but only two linear combinations of them. Therefore, one has to make a choice of normalization and the simplest one is just to set $\gamma_1 = 0$.

Coefficients for alternative and individual specific variables may (or may not) be alternative specific. For example, transport time is alternative specific, but 10 mn in public transport may not have the same impact on utility than 10 mn in a car. In this case, alternative specific coefficients are relevant. Monetary cost is also alternative specific, but in this case, one can consider than 1$ is 1$ whatever it is spent for the use of a car or in public transports. In this case, a generic coefficient is relevant.

The treatment of alternative specific variables don't differ much from the alternative and individual specific variables with a generic coefficient. However, if some of these variables are introduced, the ν parameter can only be estimated in a model without intercepts to avoid perfect multicolinearity.

Individual-related heteroscedasticity (see Swait and Louviere 1993) can be addressed by writing the utility of choosing j for individual i: $U_{ij} = V_{ij} + \sigma_i \epsilon_{ij}$, where ϵ has a variance that doesn't depend on i and j and σ_i is a parametric function of some individual-specific covariates. As the overall scale of utility is irrelevant, the utility can also be writen as: $U^*_{ij} = U_{ij}/\sigma_i = V_{ij}/\sigma_i + \epsilon_{ij}$, *i.e.* with homoscedastic errors. if V_{ij} is a linear combination of covariates, the associated coefficients are then divided by σ_i.

A logit model with only individual specific variables is sometimes called a *multinomial logit model*, one with only alternative specific variables a *conditional logit model* and one with both kind of variables a *mixed logit model*. This is seriously misleading: *conditional logit model* is also a logit model for longitudinal data in the statistical literature and *mixed logit* is one of the names of a logit model with random parameters. Therefore, in what follows, we'll use the name *multinomial logit model* for the model we've just described whatever the nature of the explanatory variables used.

mlogit package provides objects of class `mFormula` which are built upon `Formula` objects provided by the **Formula** package.[5] The **Formula** package provides richer `formulas`, which accept multiple responses (a feature not used here) and multiple set of covariates. It has in particular specific model.frame and model.matrix methods which can be used with one or several sets of covariates.

To illustrate the use of `mFormula` objects, let's use again the ModeCanada data set and consider three sets of covariates that will be indicated in a three-part formula:

- cost (monetary cost) is an alternative specific covariate with a generic coefficient (part 1),
- income and urban are individual specific covariates (part 2),
- ivt (in vehicle travel time) is alternative specific and alternative specific coefficients are expected (part 3).

```
f <- mFormula(choice ~ cost | income + urban | ivt)
```

Some parts of the formula may be omitted when there is no ambiguity. For example, the following sets of `formulas` are identical:

```
f2 <- mFormula(choice ~ cost + ivt | income + urban)
f2 <- mFormula(choice ~ cost + ivt | income + urban | 0)
```

```
f3 <- mFormula(choice ~ 0 | income | 0)
f3 <- mFormula(choice ~ 0 | income)
```

```
f4 <- mFormula(choice ~ cost + ivt)
f4 <- mFormula(choice ~ cost + ivt | 1)
f4 <- mFormula(choice ~ cost + ivt | 1 | 0)
```

By default, an intercept is added to the model, it can be removed by using `+ 0` or `- 1` in the second part.

```
f5 <- mFormula(choice ~ cost | income + 0 | ivt)
f5 <- mFormula(choice ~ cost | income - 1 | ivt)
```

`model.frame` and `model.matrix` methods are provided for `mFormula` objects. The latter is of particular interest, as illustrated in the following example:

[5] See Zeileis and Croissant (2010) for a description of the **Formula** package.

```
f <- mFormula(choice ~ cost | income  | ivt)
head(model.matrix(f, MC), 4)

##           air:(intercept) bus:(intercept) car:(intercept)   cost
## 109.train               0               0               0  58.25
## 109.air                 1               0               0 142.80
## 109.bus                 0               1               0  27.52
## 109.car                 0               0               1  71.63
##           air:income bus:income car:income train:ivt air:ivt
## 109.train          0          0          0       215       0
## 109.air           45          0          0         0      56
## 109.bus            0         45          0         0       0
## 109.car            0          0         45         0       0
##           bus:ivt car:ivt
## 109.train       0       0
## 109.air         0       0
## 109.bus       301       0
## 109.car         0     262
```

The model matrix contains $J-1$ columns for every individual specific variables (income and the intercept), which means that the coefficient associated to the first alternative (air) is set to 0. It contains only one column for cost because we want a generic coefficient for this variable. It contains J columns for ivt, because it is an alternative specific variable for which we want alternative specific coefficients.

1.3. Testing

As for all models estimated by maximum likelihood, three testing procedures may be applied to test hypothesis about models fitted using mlogit. The set of hypothesis tested defines two models: the unconstrained model that doesn't take these hypothesis into account and the constrained model that impose these hypothesis.

This in turns define three principles of tests: the *Wald test*, based only on the unconstrained model, the *Lagrange multiplier test* (or *score test*), based only on the constrained model and the *likelihood ratio test*, based on the comparison of both models.

Two of these three tests are implemented in the **lmtest** package (Zeileis and Hothorn 2002): waldtest and lrtest. The Wald test is also implemented as linearHypothesis in package **car** (Fox and Weisberg 2010), with a fairly different syntax. We provide special methods of `waldtest` and `lrtest` for `mlogit` objects and we also provide a function for the Lagrange multiplier (or score) test called scoretest.

We'll see later that the score test is especially useful for `mlogit` objects when one is interested in extending the basic multinomial logit model because, in this case, the unconstrained model may be difficult to estimate. For the presentation of further tests, we provide a convenient statpval function which extract the statistic and the p-value from the objects returned by the testing function, which can be either of class `anova` or `htest`.

```
statpval <- function(x){
    if (inherits(x, "anova"))
        result <- as.matrix(x)[2, c("Chisq", "Pr(>Chisq)")]
    if (inherits(x, "htest")) result <- c(x$statistic, x$p.value)
    names(result) <- c("stat", "p-value")
    round(result, 3)
}
```

2. Random utility model and the multinomial logit model

2.1. Random utility model

Remind from the introduction that, in a random utility model, the probability of choosing an alternative l among a set of mutually exclusive ones is:

$$\mathrm{P}(\epsilon_1 < V_l - V_1 + \epsilon_l, \epsilon_2 < V_l - V_2 + \epsilon_l, \ldots, \epsilon_K < V_l - V_J + \epsilon_l)$$

Denoting F_{-l} the cumulative density function of all the ϵs except ϵ_l, this probability is:

$$(\mathrm{P}_l \mid \epsilon_l) = F_{-l}(V_l - V_1 + \epsilon_l, \ldots, V_l - V_J + \epsilon_l) \tag{1}$$

Note that this probability is conditional on the value of ϵ_l. The unconditional probability (which depends only on β and on the value of the observed explanatory variables) is obtained by integrating out the conditional probability using the marginal density of ϵ_l, denoted f_l:

$$\mathrm{P}_l = \int F_{-l}(V_l - V_1 + \epsilon_l, \ldots, V_l - V_J) + \epsilon_l) f_l(\epsilon_l) d\epsilon_l \tag{2}$$

The conditional probability is an integral of dimension $J - 1$ and the computation of the unconditional probability adds on more dimension of integration.

2.2. The distribution of the error terms

The multinomial logit model (McFadden 1974) is a special case of the model developed in the previous section. It is based on three hypothesis

The first hypothesis is the independence of the errors. In this case, the univariate distribution of the errors can be used, which leads to the following conditional and unconditional probabilities:

$$(\mathrm{P}_l \mid \epsilon_l) = \prod_{j\neq l} F_j(V_l - V_j + \epsilon_l) \text{ and } \mathrm{P}_l = \int \prod_{j\neq l} F_j(V_l - V_j + \epsilon_l)\; f_l(\epsilon_l)\; d\epsilon_l \tag{3}$$

which means that the conditional probability is the product of $J-1$ univariate cumulative density functions and the evaluation of only a one-dimensional integral is required to compute the unconditional probability.

The second hypothesis is that each ϵ follows a Gumbel distribution, whose density and probability functions are respectively:

$$f(z) = \frac{1}{\theta} e^{-\frac{z-\mu}{\theta}} e^{-e^{-\frac{z-\mu}{\theta}}} \text{ and } F(z) = \int_{-\infty}^{+\infty} f(t)dt = e^{-e^{-\frac{z-\mu}{\theta}}} \tag{4}$$

where μ is the location parameter and θ the scale parameter. The first two moments of the Gumbel distribution are $\mathrm{E}(z) = \mu + \theta\gamma$, where γ is the Euler-Mascheroni constant (≈ 0.577) and $\mathrm{V}(z) = \frac{\pi^2}{6}\theta^2$. The mean of ϵ_js is not identified if V_j contains an intercept. We can then, without loss of generality suppose that $\mu_j = 0,\ \forall j$. Moreover, the overall scale of utility is not identified. Therefore, only $J-1$ scale parameters may be identified, and a natural choice of normalization is to impose that one of the θ_j is equal to 1.

The last hypothesis is that the errors are identically distributed. As the location parameter is not identified for any error term, this hypothesis is essentially an homoscedasticity hypothesis, which means that the scale parameter of the Gumbel distribution is the same for all the alternatives. As one of them has been previously set to 1, we can therefore suppose that, without loss of generality, $\theta_j = 1,\ \forall j \in 1 \ldots J$. The conditional and unconditional probabilities (4) then further simplify to:

$$(\mathrm{P}_l \mid \epsilon_l) = \prod_{j\neq l} e^{-e^{-(V_l - Vj + \epsilon_l)}} \text{ and } \mathrm{P}_l = \int_{-\infty}^{+\infty} \prod_{j\neq l} e^{-e^{-(V_l - Vj + t)}} e^{-t} e^{-e^{-t}} dt \tag{5}$$

The probabilities have then very simple, closed forms, which correspond to the logit transformation of the deterministic part of the utility.

$$P_l = \frac{e^{V_l}}{\sum_{j=1}^{J} e^{V_j}} \tag{6}$$

2.3. IIA property

If we consider the probabilities of choice for two alternatives l and m, we have $P_l = \frac{e^{V_l}}{\sum_j e^{V_j}}$ and $P_m = \frac{e^{V_m}}{\sum_j e^{V_j}}$. The ratio of these two probabilities is:

$$\frac{P_l}{P_m} = \frac{e^{V_l}}{e^{V_m}} = e^{V_l - V_m}$$

This probability ratio for the two alternatives depends only on the characteristics of these two alternatives and not on those of other alternatives. This is called the IIA property (for independence of irrelevant alternatives).

Consider for example inter urban trips between two towns, with 3 modes available. The initial situation is presented in table 1, which indicates that the initial market shares are 30% for car, 10% for plane and 60% for train. Consider now that a low-cost airline company enters the market, so that the price of plane ticket decreases from 150 to 100 euros.

	price	time	share	s1	s2
car	50	4	30%	29%	25%
plane	150	1	10%	25%	25%
train	80	2	60%	46%	50%

Table 1: Market shares

We consider in the table two scenarios of evolution of the market shares (columns s1 and s2). Both indicates that the market share of the plane increases and that the market shares of the other two modes decrease. For the first scenario (column s1), almost all the increase of plane market share is due to former train users who now take the plane. The car market share decreases very slightly (from 30 to 29%). This situation may be realistic if car trips are mainly family trips, as train and plane trips mainly trips for professional purpose. This situation cannot be predicted by a multinomial logit model, because of the IIA property; the ratio of the probabilities of choosing train and car should be identical before and after the decrease of plane price ticket. This is the case in column s2, where the relative decrease of the market shares for car and train are identical.

IIA relies on the hypothesis that the errors are identical and independent. It is not a problem by itself and may even be considered as a useful feature for a well specified model. However, this hypothesis may be in practice violated, especially if some important variables are omitted.

2.4. Interpretation

In a linear model, the coefficients can be directly considered as marginal effects of the explanatory variables on the explained variable. This is not the case for the multinomial logit model. However, meaningful results can be obtained using relevant transformations of the coefficients.

Marginal effects

The marginal effects are the derivatives of the probabilities with respect to the covariates, which can be be individual-specific (z_i) or alternative specific (x_{ij}):

$$\frac{\partial P_{il}}{\partial z_i} = P_{il}\left(\beta_l - \sum_j P_{ij}\beta_j\right)$$

$$\left\{\begin{array}{rcl} \frac{\partial P_{il}}{\partial x_{il}} &=& \gamma P_{il}(1-P_{il}) \\ \frac{\partial P_{il}}{\partial x_{ik}} &=& -\gamma P_{il}P_{ik} \end{array}\right.$$

- For an individual specific variable, the sign of the marginal effect is not necessarily the sign of the coefficient. Actually, the sign of the marginal effect is given by $\left(\beta_l - \sum_j P_{ij}\beta_j\right)$, which is positive if the coefficient for alternative l is greater than a weighted average of the coefficients for all the alternatives, the weights being the probabilities of choosing the alternatives. In this case, the sign of the marginal effect can be established with no ambiguity only for the alternatives with the lowest and the greatest coefficients.

- For an alternative-specific variable, the sign of the coefficient can be directly interpreted. The marginal effect is obtained by multiplying the coefficient by the product of two probabilities which is at most 0.25. The rule of thumb is therefore to divide the coefficient by 4 in order to have an upper bound of the marginal effect.

Note that the last equation can be rewriten: $\frac{\mathrm{d}P_{il}/P_{il}}{\mathrm{d}x_{ik}} = -\gamma P_{ik}$.
Therefore, when a characteristic of alternative k changes, the relative change of the probabilities for every alternatives except k is the same, which is a consequence of the IIA property.

Marginal rates of substitution

Coefficients are marginal utilities, which cannot be interpreted. However, ratios of coefficients are marginal rates of substitution. For example, if the observable part of utility is: $V = \beta_o + \beta_1 x_1 + \beta x_2 + \beta x_3$, join variations of x_1 and x_2 which ensure the same level of utility are such that: $dV = \beta_1 dx_1 + \beta_2 dx_2 = 0$ so that:

$$-\frac{dx_2}{dx_1}\mid_{dV=0} = \frac{\beta_1}{\beta_2}$$

For example, if x_2 is transport cost (in \$), x_1 transport time (in hours), $\beta_1 = 1.5$ and $\beta_2 = 0.2$, $\frac{\beta_1}{\beta_2} = 30$ is the marginal rate of substitution of time in terms of \$ and the value of 30 means that to reduce the travel time of one hour, the individual is willing to pay at most 30\$ more. Stated more simply, time value is 30\$ per hour.

Consumer's surplus

Consumer's surplus has a very simple expression for multinomial logit models, which was first derived by Small and Rosen (1981). The level of utility attained by an individual is $U_j = V_j + \epsilon_j$, j being the alternative chosen. The expected utility, from the searcher's point of view is then: $\mathrm{E}(\max_j U_j)$, where the expectation is taken over the values of all the error terms. Its expression is simply, up to an additive unknown constant, the log of the denominator of the logit probabilities and is often called the "log-sum":

$$\mathrm{E}(U) = \ln \sum_{j=1}^{J} e^{V_j} + C$$

If the marginal utility of income (α) is known and constant, the expected surplus is simply $\frac{\mathrm{E}(U)}{\alpha}$.

2.5. Application

Random utility models are fitted using the mlogit function. Basically, only two arguments are mandatory, `formula` and `data`, if a `mlogit.data` object (and not an ordinary `data.frame`) is provided.

ModeCanada

We first use the ModeCanada data set, which was already coerced to a `mlogit.data` object (called MC) in the previous section. The same model can then be estimated using as `data` argument a `mlogit.data` object:

```
ml.MC1 <- mlogit(choice ~ cost + freq + ovt | income | ivt, MC)
```

or a `data.frame`. In this latter case, further arguments that will be passed to mlogit.data should be indicated:

```
ml.MC1b <- mlogit(choice ~ cost + freq + ovt | income | ivt, ModeCanada,
                  subset = noalt == 4, alt.var = "alt", chid.var = "case")
```

mlogit provides two further usefull arguments:

- `reflevel` indicates which alternative is the "reference" alternative, *i.e.* the one for which the coefficients of individual specific covariates are 0,
- `alt.subset` indicates a subset of alternatives on which the estimation has to be performed; in this case, only the lines that correspond to the selected alternatives are used and all the choice situations where not selected alternatives has been chosen are removed.

We estimate the model on the subset of three alternatives (we exclude bus whose market share is negligible in our sample) and we set car as the reference alternative. Moreover, we use a total transport time variable computed as the sum of the in and the out of vehicule time variables.

```
MC$time <- with(MC, ivt + ovt)
ml.MC1 <- mlogit(choice ~ cost + freq | income | time, MC,
                 alt.subset = c("car", "train", "air"),
                 reflevel = "car")
```

The main results of the model are computed and displayed using the summary method:

```
summary(ml.MC1)

##
## Call:
## mlogit(formula = choice ~ cost + freq | income | time, data = MC,
##     alt.subset = c("car", "train", "air"), reflevel = "car",
##     method = "nr", print.level = 0)
##
## Frequencies of alternatives:
##     car   train     air
## 0.45757 0.16721 0.37523
##
## nr method
## 6 iterations, 0h:0m:0s
## g'(-H)^-1g = 6.94E-06
## successive function values within tolerance limits
##
## Coefficients :
##                      Estimate  Std. Error  z-value  Pr(>|z|)
## train:(intercept) -0.97034440  0.26513065  -3.6599 0.0002523 ***
## air:(intercept)   -1.89856552  0.68414300  -2.7751 0.0055185 **
## cost              -0.02849715  0.00655909  -4.3447 1.395e-05 ***
## freq               0.07402902  0.00473270  15.6420 < 2.2e-16 ***
## train:income      -0.00646892  0.00310366  -2.0843 0.0371342 *
## air:income         0.02824632  0.00365435   7.7295 1.088e-14 ***
## car:time          -0.01402405  0.00138047 -10.1589 < 2.2e-16 ***
## train:time        -0.01096877  0.00081834 -13.4036 < 2.2e-16 ***
## air:time          -0.01755120  0.00399181  -4.3968 1.099e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Log-Likelihood: -1951.3
## McFadden R^2:  0.31221
## Likelihood ratio test : chisq = 1771.6 (p.value = < 2.22e-16)
```

The frequencies of the different alternatives in the sample are first indicated. Next, some information about the optimization are displayed: the Newton-Ralphson method (with analytic gradient and hessian) is used, as it is the most efficient method for this simple model for which the log-likelihood function is concave. Note that very few iterations and computing time are required to estimate this model. Follows the usual table of coefficients and some goodness of fit measures: the value of the log-likelihood function, which is compared to the value when only intercepts are introduced, which leads to the computation of the McFadden R^2 and to the likelihood ratio test.

The fitted method can be used either to obtain the probability of actual choices (`type = "outcome"`) or the probabilities for all the alternatives (`type = "probabilities"`).

```
head(fitted(ml.MC1, type = "outcome"))

##       109       110       111       112       113       114
## 0.1909475 0.3399941 0.1470527 0.3399941 0.3399941 0.2440011

head(fitted(ml.MC1, type = "probabilities"), 4)

##           car     train       air
## 109 0.4206404 0.3884120 0.1909475
## 110 0.3696476 0.2903582 0.3399941
## 111 0.4296769 0.4232704 0.1470527
## 112 0.3696476 0.2903582 0.3399941
```

Note that the log-likelihood is the sum of the log of the fitted outcome probabilities and that, as the model contains intercepts, the average fitted probabilities for every alternative equals the market shares of the alternatives in the sample.

```
sum(log(fitted(ml.MC1, type = "outcome")))

## [1] -1951.344

logLik(ml.MC1)

## 'log Lik.' -1951.344 (df=9)

apply(fitted(ml.MC1, type = "probabilities"), 2, mean)

##       car     train       air
## 0.4575659 0.1672084 0.3752257
```

Predictions can be made using the predict method. If no data is provided, predictions are made for the average of the sample on which the estimation as been performed.

```
predict(ml.MC1)

##       car     train       air
## 0.5066362 0.2116876 0.2816761
```

Assume, for example, that we wish to predict the effect of a reduction of train transport time of 20%. We first create a new data.frame simply by multiplying train transport time by 0.8 and then using the predict method with this new data.frame

```
NMC <- MC
NMC[index(NMC)$alt == "train", "time"] <-
    0.8 * NMC[index(NMC)$alt == "train", "time"]
Oprob <- fitted(ml.MC1, type = "probabilities")
Nprob <- predict(ml.MC1, newdata = NMC)
rbind(old = apply(Oprob, 2, mean),
      new = apply(Nprob, 2, mean))

##           car     train       air
## old 0.4575659 0.1672084 0.3752257
## new 0.4044736 0.2635801 0.3319462
```

If, for the first individuals in the sample, we compute the ratio of the probabilities of the air and the car mode, we obtain:

```
head(Nprob[, "air"] / Nprob[, "car"])

##       109       110       111       112       113       114
## 0.4539448 0.9197791 0.3422401 0.9197791 0.9197791 0.6021092

head(Oprob[, "air"] / Oprob[, "car"])

##       109       110       111       112       113       114
## 0.4539448 0.9197791 0.3422401 0.9197791 0.9197791 0.6021092
```

which is an illustration of the IIA property. If train time changes, it changes the probabilities of choosing air and car, but not their ratio.

We next compute the surplus for individuals of the sample induced by train time reduction. This requires the computation of the log-sum term (also called inclusive value or inclusive utility) for every choice situation, which writes:

$$\mathrm{iv}_i = \ln \sum_{j=1}^{J} e^{\beta^\top x_{ij}}$$

For this purpose, we use the logsum function, which works on a vector of `coefficients` and a `model.matrix`. The basic use of logsum consists on providing as unique argument (called `coef`) a `mlogit` object. In this case, the `model.matrix` and the `coef` are extracted from the same model.

```
ivbefore <- logsum(ml.MC1)
```

To compute the log-sum after train time reduction, we must provide a `model.matrix` which is not the one corresponding to the fitted model. This can be done using the `X` argument which is a matrix or an object from which a `model.matrix` can be extracted. This can also be done by filling the `data` (a `data.frame` or an object from which a `data.frame` can be extracted using a model.frame method), and eventually the `formula` argument (a `formula` or an object for which the formula method can be applied). If no formula is provided but if `data` is a `mlogit.data` object, the formula is extracted from it.

```
ivafter <- logsum(ml.MC1, data = NMC)
```

Surplus variation is then computed as the difference of the log-sums divided by the opposite of the cost coefficient which can be interpreted as the marginal utility of income:

```
surplus <- - (ivafter - ivbefore) / coef(ml.MC1)["cost"]
summary(surplus)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.5852  2.8439  3.8998  4.6971  5.8437 31.3912
```

Consumer's surplus variation range from 0.6 to 31 Canadian \$, with a median value of about 4\$.

Marginal effects are computed using the effects method. By default, they are computed at the sample mean, but a `data` argument can be provided. The variation of the probability and of the covariate can be either absolute or relative. This is indicated with the `type` argument which is a combination of two `a` (as absolute) and `r` (as relative). For example, `type = "ar"` means that what is measured is an absolute variation of the probability for a relative variation of the covariate.

```
effects(ml.MC1, covariate = "income", type = "ar")

##        car      train        air
## -0.1822177 -0.1509079  0.3331256
```

The results indicate that, if income has doubled, the probability of choosing `air` increases by 33 points of percentage, as the probabilities of choosing `car` and `train` decrease by 18 and 15 points of percentage.

For an alternative specific covariate, a matrix of marginal effects is displayed.

```
effects(ml.MC1, covariate = "cost", type = "rr")

##                car      train        air
## car   -0.9131273  0.9376923  0.9376923
## train  0.3358005 -1.2505014  0.3358005
## air    1.2316679  1.2316679 -3.1409703
```

The cell in the l^{th} row and the c^{th} column indicates the change of the probability of choosing alternative c when the cost of alternative l changes. As `type = "rr"`, elasticities are computed. For example, a 10% change of train cost increases the probabilities of choosing car and air by 3.36%. Note that the relative changes of the probabilities of choosing one of these two modes are equal, which is a consequence of the IIA property.

Finally, in order to compute travel time valuation, we divide the coefficients of travel times (in minutes) by the coefficient of monetary cost (in $).

```
coef(ml.MC1)[grep("time", names(coef(ml.MC1)))] /
    coef(ml.MC1)["cost"] * 60

##   car:time train:time   air:time
##   29.52728   23.09447   36.95360
```

The value of travel time ranges from 23 for train to 37 Canadian $ per hour for plane.

NOx

The second example is a data set used by Fowlie (2010), called NOx. She analyzed the effect of an emissions trading program (the NOx budget program which seeks to reduce the emission of nitrogen oxides) on the behavior of producers. More precisely, coal electricity plant managers may adopt one out of fifteen different technologies in order to comply to the emissions defined by the program. Some of them require high investment (the capital cost is kcost) and are very efficient to reduce emissions, some other require much less investment but are less efficient and the operating cost (denoted vcost) is then higher because pollution permits must be purchased to offset emissions exceeding their allocation.

The focus of the paper is on the effects of the regulatory environment on manager's behavior. Some firms are deregulated, whereas other are either regulated or public. Rate of returns is applied for regulated firms, which means that they perceive a "fair" rate of return on their investment. Public firms also enjoy significant cost of capital advantages.

Therefore, the main hypothesis of the paper is that public and regulated firms will adopt much more capitalistic intensive technologies than deregulated and public ones, which means that the coefficient of capital cost should take a higher negative value for deregulated firms. Capital cost is interacted with the age of the plant (measured as a deviation from the sample mean age), as firms should weight capital costs more heavily for older plants, as they have less time to recover these costs.

Multinomial logit models are estimated for the three subsamples defined by the regulatory environment. The 15 technologies are not available for every plants, the sample is therefore restricted to available technologies, using the available covariate. Three technology dummies are introduced: post for post-combustion polution control technology, cm for combustion modification technology and lnb for low NOx burners technology.

A last model is estimated for the whole sample, but the parameters are allowed to be proportional to each other. The scedasticity function is described in the fourth part of the formula, it contains here only one covariate, env. Note also that for the pooling model, the author has introduced a specific capital cost coefficient for deregulated firms[6]

```
data("NOx", package = "mlogit")
NOx$kdereg <- with(NOx, kcost * (env == "deregulated"))

NOxml <- mlogit.data(NOx, chid.var = "chid",
                     alt.var = "alt", id.var = "id")
ml.pub <- mlogit(choice ~ post + cm + lnb + vcost +
                     kcost + kcost:age | - 1,
                 subset = available & env == "public", data = NOxml)
ml.reg <- update(ml.pub, subset = available & env == "regulated")
ml.dereg <- update(ml.pub, subset = available & env == "deregulated")
ml.pool <- mlogit(choice ~ post + cm + lnb + vcost + kcost +
                      kcost:age + kdereg | - 1 | 0 | env,
                  subset = available == 1, data = NOxml, method = "bhhh")
library("texreg")
screenreg(list(Public = ml.pub, Deregulated = ml.dereg,
               Regulated = ml.reg, pooled = ml.pool))

##
## =====================================================================
##                     Public      Deregulated  Regulated    pooled
## ---------------------------------------------------------------------
## post                 -5.71 ***    -1.50 ***    -2.67 ***    -2.31 ***
##                      (1.02)       (0.22)       (0.28)       (0.21)
```

[6] Note the use of the method argument, set to bhhh. **mlogit** use its own optimisation functions, but borrows its syntax from package **maxLik** (Toomet and Henningsen 2010). The default method is bfgs, except for the basic model, for which it is nr. As the default algorithm failed to converged, we use here bhhh.

```
## cm                   -4.43 ***    -1.54 ***    -1.91 ***    -2.06 ***
##                      (0.56)       (0.19)       (0.18)       (0.16)
## lnb                  -3.96 ***    -1.55 ***    -2.21 ***    -2.03 ***
##                      (0.59)       (0.22)       (0.22)       (0.17)
## vcost                -1.56 ***    -0.19 ***    -0.28 ***    -0.31 ***
##                      (0.36)       (0.06)       (0.06)       (0.04)
## kcost                 0.04        -0.06 **      0.01         0.01
##                      (0.11)       (0.02)       (0.03)       (0.02)
## kcost:age            -0.08        -0.04 **     -0.02 *      -0.02 ***
##                      (0.04)       (0.01)       (0.01)       (0.01)
## kdereg                                                      -0.07 ***
##                                                             (0.01)
## sig.envderegulated                                           0.32 **
##                                                             (0.12)
## sig.envpublic                                               -0.33 ***
##                                                             (0.08)
## ---------------------------------------------------------------------
## AIC                 168.92       690.15       731.48      1634.22
## Log Likelihood      -78.46      -339.07      -359.74      -808.11
## Num. obs.           113          227          292          632
## =====================================================================
## *** p < 0.001, ** p < 0.01, * p < 0.05
```

Coefficients are very different on the different sub-samples defined by the regulatory environment. Note in particular that the capital cost coefficient is positive and insignificant for public and regulated firms, as it is significantly negative for deregulated firms. Errors seems to have significant larger variance for deregulated firms and lower ones for public firms compared to regulated firms. The hypothesis that the coefficients (except the `kcost` one) are identical up to a multiplicative scalar can be performed using a likelihood ratio test:

```
stat <- 2 * (logLik(ml.dereg) + logLik(ml.reg) +
             logLik(ml.pub) - logLik(ml.pool))
stat

## 'log Lik.' 61.6718 (df=6)

pchisq(stat, df = 9, lower.tail = FALSE)

## 'log Lik.' 6.377283e-10 (df=6)
```

The hypothesis is strongly rejected.

3. Logit models relaxing the iid hypothesis

In the previous section, we assumed that the error terms were *iid* (identically and independently distributed), *i.e.* uncorrelated and homoscedastic. Extensions of the basic multinomial logit model have been proposed by relaxing one of these two hypothesis while maintaining the hypothesis of Gumbell distribution.

3.1. The heteroskedastic logit model

The heteroskedastic logit model was proposed by Bhat (1995). The probability that $U_l > U_j$ is:

$$P(\epsilon_j < V_l - V_j + \epsilon_l) = e^{-e^{-\frac{(V_l - V_j + \epsilon_l)}{\theta_j}}}$$

which implies the following conditional and unconditional probabilities

$$(P_l \mid \epsilon_l) = \prod_{j\neq l} e^{-e^{-\frac{(V_l - V_j + \epsilon_l)}{\theta_j}}} \qquad (7)$$

$$\begin{aligned} P_l &= \int_{-\infty}^{+\infty} \prod_{j\neq l} \left(e^{-e^{-\frac{(V_l - V_j + t)}{\theta_j}}} \right) \frac{1}{\theta_l} e^{-\frac{t}{\theta_l}} e^{-e^{-\frac{t}{\theta_l}}} dt \\ &= \int_0^{+\infty} \left(e^{-\sum_{j\neq l} e^{-\frac{V_l - V_j - \theta_l \ln t}{\theta_j}}} \right) e^{-t} dt \end{aligned} \qquad (8)$$

There is no closed form for this integral but it can be efficiently computed using a Gauss quadrature method, and more precisely the Gauss-Laguerre quadrature method.

3.2. The nested logit model

The nested logit model was first proposed by McFadden (1978). It is a generalization of the multinomial logit model that is based on the idea that some alternatives may be joined in several groups (called nests). The error terms may then present some correlation in the same nest, whereas error terms of different nests are still uncorrelated.

Denoting $m = 1 \ldots M$ the nests and B_m the set of alternatives belonging to nest m, the multivariate distribution of the error terms is:

$$\exp\left(-\sum_{m=1}^{M} \left(\sum_{j\in B_m} e^{-\epsilon_j/\lambda_m} \right)^{\lambda_m} \right)$$

The marginal distributions of the ϵs are still univariate extreme value, but there is now some correlation within nests. $1 - \lambda_m$ is a measure of the correlation, *i.e.* $\lambda_m = 1$ implies no correlation. In the special case where $\lambda_m = 1 \; \forall m$, the errors are *iid* Gumbel errors and the nested logit model reduce to the multinomial logit model. It can then be shown that the probability of choosing alternative j that belongs to nest l is:

$$P_j = \frac{e^{V_j/\lambda_l}\left(\sum_{k\in B_l} e^{V_k/\lambda_l}\right)^{\lambda_l - 1}}{\sum_{m=1}^{M}\left(\sum_{k\in B_m} e^{V_k/\lambda_m}\right)^{\lambda_m}}$$

and that this model is a random utility model if all the λ parameters are in the $0-1$ interval.[7]

Let us now write the deterministic part of the utility of alternative j as the sum of two terms: the first one (Z_j) being specific to the alternative and the second one (W_l) to the nest it belongs to:

$$V_j = Z_j + W_l$$

We can then rewrite the probabilities as follow:

$$\begin{aligned} P_j &= \frac{e^{(Z_j+W_l)/\lambda_l}}{\sum_{k\in B_l} e^{(Z_k+W_l)/\lambda_l}} \times \frac{\left(\sum_{k\in B_l} e^{(Z_k+W_l)/\lambda_l}\right)^{\lambda_l}}{\sum_{m=1}^{M}\left(\sum_{k\in B_m} e^{(Z_k+W_m)/\lambda_m}\right)^{\lambda_m}} \\ &= \frac{e^{Z_j/\lambda_l}}{\sum_{k\in B_l} e^{Z_k/\lambda_l}} \times \frac{\left(e^{W_l/\lambda_l}\sum_{k\in B_l} e^{Z_k/\lambda_l}\right)^{\lambda_l}}{\sum_{m=1}^{M}\left(e^{W_m/\lambda_m}\sum_{k\in B_m} e^{Z_k/\lambda_m}\right)^{\lambda_m}} \end{aligned}$$

Then denote $I_l = \ln \sum_{k\in B_l} e^{Z_k/\lambda_l}$ which is often called the log-sum, the inclusive value or the inclusive utility.[8] We then can write the probability of choosing alternative j as:

$$P_j = \frac{e^{Z_j/\lambda_l}}{\sum_{k\in B_l} e^{Z_k/\lambda_l}} \times \frac{e^{W_l+\lambda_l I_l}}{\sum_{m=1}^{M} e^{W_m+\lambda_m I_m}}$$

The first term $\mathrm{P}_{j|l}$ is the conditional probability of choosing alternative j if the nest l is chosen. It is often referred as the *lower model*. The second term P_l is the marginal probability of choosing the nest l and is referred as the *upper model*. $W_m + \lambda_m I_m$ can be interpreted as the expected utility of choosing the best alternative in m, W_m being the expected utility of choosing an alternative in this nest (whatever this alternative is) and $\lambda_m I_m$ being the expected extra utility gained by being able to choose the best alternative in the nest. The inclusive values link the two models. It is then straightforward to show that IIA applies within nests, but not for two alternatives in different nests.

[7] A slightly different version of the nested logit model (Daly 1987) is often used, but is not compatible with the random utility maximization hypothesis. Its difference with the previous expression is that the deterministic parts of the utility for each alternative is not divided by the nest elasticity. The differences between the two versions have been discussed in Koppelman and Wen (1998), Heiss (2002) and Hensher and Greene (2002).

[8] We've already encountered this expression in section 2.4.3.

A consistent but inefficient way of estimating the nested logit model is to estimate separately its two components. The coefficients of the lower model are first estimated, which enables the computation of the inclusive values I_m. The coefficients of the upper model are then estimated, using I_m as covariates. Maximizing directly the likelihood function of the nested model leads to a more efficient estimator.

3.3. Applications

ModeCanada

Bhat (1995) estimated the heteroscedastic logit model on the ModeCanada data set. Using **mlogit**, the heteroscedastic logit model is obtained by setting the `heterosc` argument to `TRUE`:

```
ml.MC <- mlogit(choice ~ freq + cost + ivt + ovt | urban + income, MC,
                reflevel = 'car', alt.subset = c("car", "train", "air"))
hl.MC <- mlogit(choice ~ freq + cost + ivt + ovt | urban + income, MC,
                reflevel = 'car', alt.subset = c("car", "train", "air"),
                heterosc = TRUE)
coef(summary(hl.MC))[11:12, ]

##           Estimate Std. Error   z-value     Pr(>|z|)
## sp.train 1.2371829  0.1104610 11.200182 0.000000e+00
## sp.air   0.5403239  0.1118353  4.831425 1.355592e-06
```

The variance of the error terms of train and air are respectively higher and lower than the variance of the error term of car (set to 1). Note that the z-values and p-values of the output are not particularly meaningful, as the hypothesis that the coefficient is zero (and not one) is tested. The homoscedascticity hypothesis can be tested using any of the three tests. A particular convenient syntax is provided in this case. For the likelihood ratio and the wald test, one can use only the fitted heteroscedastic model as argument. In this case, it is guessed that the hypothesis that the user wants to test is the homoscedasticity hypothesis.

```
lr.heter <- lrtest(hl.MC, ml.MC)
wd.heter <- waldtest(hl.MC, heterosc = FALSE)
```

or, more simply:

```
lrtest(hl.MC)
waldtest(hl.MC)
```

The Wald test can also be computed using the linearHypothesis function from the **car** package :

```
library("car")
lh.heter <- linearHypothesis(hl.MC, c('sp.air = 1', 'sp.train = 1'))
```

For the score test, we provide the constrained model as argument, which is the standard multinomial logit model and the supplementary argument which defines the unconstrained model, which is in this case `heterosc = TRUE`.

```
sc.heter <- scoretest(ml.MC, heterosc = TRUE)
```

```
sapply(list(wald = wd.heter, lh = lh.heter,
            score = sc.heter, lr = lr.heter), statpval)

##            wald     lh score    lr
## stat     25.196 25.196 9.488 6.888
## p-value   0.000  0.000 0.009 0.032
```

The homoscedasticity hypothesis is strongly rejected using the Wald test, but only a the 1 and 5% level for, respectively, the score and the likelihood ratio tests.

JapaneseFDI

Head and Mayer (2004) analyzed the choice of one of the 57 European regions belonging to 9 countries by Japenese firms to implement a new production unit.

```
data("JapaneseFDI", package = "mlogit")
jfdi <- mlogit.data(JapaneseFDI, chid.var = "firm", alt.var = "region",
                    group.var = "country")
```

Note that we've used an extra argument to mlogit.data called `group.var` which indicates the grouping variable, which will be used later to define easily the nests. There are two sets of covariates: the wage rate `wage`, the unemployment rate `unemp`, a dummy indicating that the region is eligible to European funds elig and the area area are observed at the regional level and are therefore relevant for the estimation of the lower model, whereas the social cotisation rate scrate and the corporate tax rate ctaxrate are observed at the country level and are therefore suitable for the upper model.

We first estimate a multinomial logit model:

```
ml.fdi <- mlogit(choice ~ log(wage) + unemp + elig + log(area) +
                     scrate + ctaxrate | 0, data = jfdi)
```

Note that, as the covariates are only alternative specific, the intercepts are not identified and therefore have been removed. We next estimate the lower model, which analyses the choice of a region within a given country. Therefore, for each choice situation, we estimate the choice of a region on the subset of regions of the country which has been chosen. Moreover, observations concerning Portugal and Ireland are removed as these two countries are mono-region.

```
lm.fdi <- mlogit(choice ~ log(wage) + unemp + elig + log(area) | 0,
                 data = jfdi, subset = country == choice.c &
                                  ! country %in% c("PT", "IE"))
```

We next use the fitted lower model in order to compute the inclusive value, at the country level:

$$\mathrm{iv}_{ig} = \ln \sum_{j \in B_g} e^{\beta^\top x_{ij}}$$

where B_g is the set of regions for country g. When a grouping variable is provided in the mlogit.data function, inclusive values are by default computed for every group g (global inclusive values are obtained by setting the `type` argument to `"global"`). By default, `output` is set to `"chid"` and the results is a vector (if `type = "global"`) or a matrix (if `type = "region"`) with row number equal to the number of choice situations. If `output` is set to `"obs"`, a vector of length equal to the number of lines of the data in long format is returned. The following code indicates different ways to use the logsum function:

```
lmformula <- formula(lm.fdi)
head(logsum(ml.fdi, data = jfdi, formula = lmformula, type = "group"), 2)

##          BE       DE       ES       FR       IE       IT       NL
## 3 3.595818 5.415838 3.593702 5.153709 1.933707 5.051387 4.077845
## 4 4.113243 5.765190 4.445012 5.383095 1.960462 5.687569 4.490379
##          PT       UK
## 3 2.702028 4.900622
## 4 3.200124 5.378561

head(logsum(ml.fdi, data = jfdi, formula = lmformula, type = "global"))

##        3        4        5        7        8        9
## 6.736116 7.182139 7.121855 7.084245 7.133368 7.133368

head(logsum(ml.fdi, data = jfdi, formula = lmformula, output = "obs"))

## [1] 3.595818 3.595818 3.595818 3.595818 5.415838 5.415838

head(logsum(ml.fdi, data = jfdi, formula = lmformula,
        type = "global", output = "obs"))

## [1] 6.736116 6.736116 6.736116 6.736116 6.736116 6.736116
```

To add the inclusive values in the original data.frame, we use output = "obs" and the type argument can be omitted as its default value is "group":

```
JapaneseFDI$iv <- logsum(lm.fdi, data = jfdi, formula = lmformula,
                         output = "obs")
```

We next select the relevant variables for the estimation of the upper model, select unique lines in order to keep only one observation for every choice situation / country combination and finally we coerce the response (choice.c) to a logical for the chosen country.

```
JapaneseFDI.c <- subset(JapaneseFDI,
                        select = c("firm", "country", "choice.c",
                                   "scrate", "ctaxrate", "iv"))
JapaneseFDI.c <- unique(JapaneseFDI.c)
JapaneseFDI.c$choice.c <- with(JapaneseFDI.c, choice.c == country)
```

Finally, we estimate the upper model, using the previously computed inclusive value as a covariate.

```
jfdi.c <- mlogit.data(JapaneseFDI.c, choice = "choice.c",
                      alt.var = "country", chid.var = "firm",
                      shape = "long")
um.fdi <- mlogit(choice.c ~ scrate + ctaxrate + iv | 0, data = jfdi.c)
```

If one wants to obtain different iv coefficients for different countries, the iv covariate should be introduced in the 3th part of the formula and the coefficients for the two mono-region countries (Ireland and Portugal) should be set to 1, using the constPar argument.

```
um2.fdi <- mlogit(choice.c ~ scrate + ctaxrate | 0 | iv, data = jfdi.c,
                  constPar = c("iv:PT" = 1, "iv:IE" = 1))
```

We next estimate the full-information maximum likelihood nested model. It is obtained by adding a nests argument to the mlogit function. This should be a named list of alternatives (here regions), the names being the nests (here the countries). More simply, if a group variable has been indicated while using mlogit.data, nests can be a boolean.

Two flavors of nested models can be estimated, using the un.nest.el argument which is a boolean. If TRUE, one imposes that the coefficient associated with the inclusive utility is the same for every nest, which means that the degree of correlation inside each nest is the same. If FALSE, a different coefficient is estimated for every nest.

	mult. logit	lower model	upper model	upper model	nested logit	nested logit
log(wage)	0.47	1.21^{*}			0.46	0.77^{**}
	(0.25)	(0.48)			(0.25)	(0.25)
unemp	-8.90^{***}	-9.77^{***}			-7.62^{***}	-6.95^{**}
	(1.69)	(2.39)			(1.60)	(2.28)
elig	-0.25	-0.89^{**}			-0.34	-0.18
	(0.21)	(0.33)			(0.20)	(0.18)
log(area)	0.31^{***}	0.29^{***}			0.29^{***}	0.15^{**}
	(0.05)	(0.06)			(0.05)	(0.05)
scrate	-2.26^{***}		-2.49^{***}	0.15	-2.44^{***}	-0.88
	(0.38)		(0.38)	(1.72)	(0.38)	(0.86)
ctaxrate	-4.82^{***}		-3.69^{***}	-1.78	-4.13^{***}	-2.35
	(0.59)		(0.60)	(1.32)	(0.66)	(1.25)
iv			0.66^{***}		0.85^{***}	
			(0.06)		(0.08)	
iv:BE				0.72^{***}		0.48^{*}
				(0.08)		(0.19)
iv:DE				0.72^{***}		0.52^{**}
				(0.08)		(0.17)
iv:ES				0.86^{***}		0.77^{***}
				(0.05)		(0.20)
iv:FR				0.75^{***}		0.67^{***}
				(0.04)		(0.09)
iv:IT				0.62^{***}		0.25^{**}
				(0.05)		(0.08)
iv:NL				0.69^{***}		0.18
				(0.06)		(0.10)
iv:UK				0.87^{***}		0.86^{***}
				(0.06)		(0.10)
AIC	3469.13	1738.04	1746.58	1713.86	3467.96	3437.01
Log Likelihood	-1728.57	-865.02	-870.29	-845.93	-1726.98	-1703.50
Num. obs.	452	421	452	452	452	452

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 2: Nested logit models for the choice by Japanese firms of a european region

```r
nl.fdi <- mlogit(choice ~ log(wage) + unemp + elig + log(area) +
                     scrate + ctaxrate | 0, data = jfdi,
                 nests = TRUE, un.nest.el = TRUE)
nl2.fdi <- update(nl.fdi, un.nest.el = FALSE,
                  constPar = c('iv:PT' = 1, 'iv:IE' = 1))
```

The results of the fitted models are presented in table 2 using the **texreg** package. For the nested logit models, two tests are of particular interest:

- the test of no nests, which means that all the nest elasticities are equal to 1,
- the test of unique nest elasticities, which means that all the nest elasticities are equal to each other.

For the test of no nests, the nested model is provided as the unique argument for the lrtes and the waldtest function. For the scoretest, the constrained model (*i.e.* the multinomial logit model) is provided as the first argument and the second argument is `nests`, which describes the nesting structure that one wants to test.

```
lr.nest <- lrtest(nl2.fdi)
wd.nest <- waldtest(nl2.fdi)
sc.nest <- scoretest(ml.fdi, nests = TRUE,
                     constPar = c('iv:PT' = 1, 'iv:IE' = 1))
```

The Wald test can also be performed using the linearHypothesis function:

```
lh.nest <- linearHypothesis(nl2.fdi,
                            c("iv:BE = 1", "iv:DE = 1", "iv:ES = 1",
                              "iv:FR = 1", "iv:IT = 1", "iv:NL = 1",
                              "iv:UK = 1"))
```

```
sapply(list(wald = wd.nest, lh = lh.nest,
            score = sc.nest, lr = lr.nest), statpval)

##            wald      lh score     lr
## stat    208.407 208.407 60.28 50.122
## p-value   0.000   0.000  0.00  0.000
```

The three tests reject the null hypothesis of no correlation. We next test the hypothesis that all the nest elasticities are equal.

```
lr.unest <- lrtest(nl2.fdi, nl.fdi)
wd.unest <- waldtest(nl2.fdi, un.nest.el = TRUE)
sc.unest <- scoretest(ml.fdi, nests = TRUE, un.nest.el = FALSE,
                      constPar = c('iv:IE' = 1, 'iv:PT' = 1))
lh.unest <- linearHypothesis(nl2.fdi,
                             c("iv:BE = iv:DE", "iv:BE = iv:ES",
                               "iv:BE = iv:FR", "iv:BE = iv:IT",
                               "iv:BE = iv:NL", "iv:BE = iv:UK"))
```

```
sapply(list(wald = wd.unest, lh = lh.unest,
            score = sc.unest, lr = lr.unest), statpval)

##           wald     lh score     lr
## stat    73.535 73.535 60.28 46.954
## p-value  0.000  0.000  0.00  0.000
```

Once again, the three tests strongly reject the hypothesis.

4. The random parameters (or mixed) logit model

4.1. Derivation of the model

A mixed logit model or random parameters logit model is a logit model for which the parameters are assumed to vary from one individual to another. It is therefore a model that takes the heterogeneity of the population into account.

The probabilities

For the standard logit model, the probability that individual i choose alternative j is:

$$P_{il} = \frac{e^{\beta' x_{il}}}{\sum_j e^{\beta' x_{ij}}}$$

Suppose now that the coefficients are individual-specific. The probabilities are then:

$$P_{il} = \frac{e^{\beta_i' x_{il}}}{\sum_j e^{\beta_i' x_{ij}}}$$

The mixed logit model consists on considering the β_i's as random draws from a distribution whose parameters are estimated. The probability that individual i choose alternative l, for a given value of β_i is:

$$P_{il} \mid \beta_i = \frac{e^{\beta_i' x_{il}}}{\sum_j e^{\beta_i' x_{ij}}}$$

To get the unconditional probability, we have to integrate out this conditional probability, using the density function of β. Suppose that $V_{il} = \alpha + \beta_i x_{il}$, *i.e.* there is only one individual-specific coefficient and that the density of β_i is $f(\beta, \theta)$, θ being the vector of the parameters of the distribution of β. The unconditional probability is then:

$$P_{il} = \mathrm{E}(P_{il} \mid \beta_i) = \int_\beta (P_{il} \mid \beta) f(\beta, \theta) d\beta = \int_\beta \frac{e^{\beta_i' x_{il}}}{\sum_j e^{\beta_i' x_{ij}}} f(\beta, \theta) d\beta$$

which is a one-dimensional integral that can be efficiently estimated by quadrature methods. If $V_{il} = \beta_i^\top x_{il}$ where β_i is a vector of length K and $f(\beta, \theta)$ is the joint density of the K individual-specific coefficients, the unconditional probability is:

$$P_{il} = \mathrm{E}(P_{il} \mid \beta_i) = \int_{\beta_1} \int_{\beta_2} \ldots \int_{\beta_K} (P_{il} \mid \beta) f(\beta, \theta) d\beta_1 d\beta_2 \ldots d\beta_K$$

This is a K-dimensional integral which cannot easily be estimated by quadrature methods. The only practical method is then to use simulations. More precisely, R draws of the parameters are taken from the distribution of β, the probability is computed for every draw and the unconditional probability, which is the expected value of the conditional probabilities is estimated by the average of the R probabilities.

Individual parameters

The expected value of a random coefficient (E(β)) is simply estimated by the mean of the R draws on its distribution : $\bar{\beta} = \sum_{r=1}^{R} \beta_r$. Individual parameters are obtained by first computing the probabilities of the observed choice of i for every value of β_r:

$$P_{ir} = \frac{\sum_j y_{ij} e^{\beta_r' x_{ij}}}{\sum_j e^{\beta_r' x_{ij}}}$$

where y_{ij} is a dummy equal to one if i has chosen alternative j. The expected value of the parameter for an individual is then estimated by using these probabilities to weight the R β values:

$$\hat{\beta}_i = \frac{\sum_r P_{ir} \beta_r}{\sum_r P_{ir}}$$

Panel data

If there are repeated observations for the same individuals, this longitudinal dimension of the data can be taken into account in the mixed logit model, assuming that the random parameters of individual i are the same for all his choice situations. Denoting y_{itl} a dummy equal to 1 if i choose alternative l for the t^{th} choice situation, the probability of the observed choice is:

$$P_{it} = \prod_j \frac{\sum_j y_{itj} e^{\beta_i x_{itl}}}{\sum_j e^{\beta_i x_{itj}}}$$

The joint probability for the T observations of individual i is then:

$$P_i = \prod_t \prod_j \frac{\sum_j y_{itj} e^{\beta_i x_{itj}}}{\sum_j e^{\beta_i x_{itj}}}$$

and the log-likelihood is simply $\sum_i \ln P_i$.

4.2. Application

The random parameter logit model is estimated by providing a `rpar` argument to mlogit. This argument is a named vector, the names being the random coefficients and the values the name of the law of distribution. Currently, the normal (`"n"`), log-normal (`"ln"`), zero-censored normal (`"cn"`), uniform (`"u"`) and triangular (`"t"`) distributions are available. For these distributions, two parameters are estimated which are, for normal related distributions, the mean and the standard-deviation of the underlying normal distribution and for the uniform and triangular distribution, the mean and the half range of the distribution. For these last two distributions, zero-bounded variants are also provided (`"zbt"` and `"zbu"`). These two distributions are defined by only one parameter (the mean) and their definition domain varies from 0 to twice the mean.

It's often the case that we are willing to impose that the distribution of a random parameter takes only positive or negative values. For example, the price coefficient should be negative for every individual. In this case, `"zbt"` and `"zbu"` can be used. The use of `"ln"` and `"cn"` can also be relevant but, in this case, if only negative values are expected, one should consider the distribution of the opposite of the random price coefficient. This can easily be done using the `opposite` argument of `mlogit.data`. For example, if `opposite = c("price", "time")` is used, `mlogit.data` returns the opposite of the two variables, so that the corresponding coefficients should now be positive.

`R` is the number of draws, `halton` indicates whether halton draws (see Train 2003, chapter 9) should be used (`NA` indicates that default halton draws are used), `panel` is a boolean which indicates if one wishes to use the panel data version of the log-likelihood.

Correlations between random parameters can be introduced only for normal-related distributed random parameters, using the `correlation` argument. If `TRUE`, all the normal-related random parameters are correlated. The `correlation` argument can also be a character vector indicating the random parameters that one wishes to be correlated.

Train

We first use the Train data set, previously coerced to a `mlogit.data` object called Tr. We first estimate the multinomial model: both alternatives being virtual train trips, it is relevant to use only generic coefficients and to remove the intercept:

```
Train.ml <- mlogit(choice ~ price + time + change + comfort | - 1, Tr)
coef(summary(Train.ml))

##           Estimate  Std. Error   z-value     Pr(>|z|)
## price   0.06735804 0.003393252 19.850585 0.000000e+00
## time    1.72055142 0.160351702 10.729861 0.000000e+00
## change  0.32634094 0.059489152  5.485722 4.117843e-08
## comfort 0.94572555 0.064945464 14.561842 0.000000e+00
```

All the coefficients are highly significant and have the predicted positive sign (remind than an increase in the variable comfort implies using a less comfortable class). The coefficients can't be directly interpreted, but dividing them by the price coefficient, we get monetary values:

```
coef(Train.ml)[- 1] / coef(Train.ml)[1]

##     time   change  comfort
## 25.54337  4.84487 14.04028
```

We obtain the value of 26 euros for an hour of traveling, 5 euros for a change and 14 euros to travel in a more comfortable class.

We then estimate a model with three random parameters, time, change and comfort. We first estimate the uncorrelated mixed logit model:

```
Train.mxlu <- mlogit(choice ~ price + time + change + comfort | - 1, Tr,
                     panel = TRUE, rpar = c(time = "n", change = "n",
                                            comfort = "n"), R = 100,
                     correlation = FALSE, halton = NA, method = "bhhh")
names(coef(Train.mxlu))

## [1] "price"      "time"       "change"     "comfort"
## [5] "sd.time"    "sd.change"  "sd.comfort"
```

Compared to the multinomial logit model, there are now three more coefficients which are the standard deviations of the distribution of the three random parameters. The correlated model is obtained by setting the `correlation` argument to `TRUE`.

```
Train.mxlc <- update(Train.mxlu, correlation = TRUE)
names(coef(Train.mxlc))

##  [1] "price"           "time"            "change"
##  [4] "comfort"         "time.time"       "time.change"
##  [7] "time.comfort"    "change.change"   "change.comfort"
## [10] "comfort.comfort"
```

There are now 6 parameters which are the elements of the Choleski decomposition of the covariance matrix of the three random parameters.

The summary method supplies the usual table of coefficients, and also some statistics about the random parameters. Random parameters may be extracted using the function rpar which take as first argument a `mlogit` object, as second argument `par` the parameter(s) to be extracted and as third argument `norm` the coefficient (if any) that should be used for normalization. This is usually the coefficient of the price (taken as a non random parameter), so that the effects can be interpreted as monetary values. This function returns a `rpar` object, and several methods/functions (summary, mean, med for the median and stdev for the standard deviation) are provided to describe it:

```
time.value <- rpar(Train.mxlc, "time", norm = "price")
summary(time.value)

##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
##      -Inf  8.753119 33.367588 33.367588 57.982056       Inf
```

In case of correlated random parameters further functions are provided to analyze the correlation of the coefficients:

```
cor.mlogit(Train.mxlc)

##                time      change   comfort
## time     1.00000000 -0.02956296 0.3695645
## change  -0.02956296  1.00000000 0.2489270
## comfort  0.36956453  0.24892701 1.0000000

cov.mlogit(Train.mxlc)

##               time     change  comfort
## time    28.6460389 -0.2787999 5.557933
## change  -0.2787999  3.1047367 1.232467
## comfort  5.5579334  1.2324667 7.895535

stdev(Train.mxlc)

##     time   change  comfort
## 5.352199 1.762026 2.809899
```

As the `change` attribute seems to be weakly correlated with the two other random parameters, the correlation can be restricted to the `time` and `comfort` attributes by filling the `correlation` argument with a character vector.

```
Train.mxlc2 <- update(Train.mxlc, correlation = c("time", "comfort"))
cor.mlogit(Train.mxlc2)

##              time   comfort
## time    1.0000000 0.3909467
## comfort 0.3909467 1.0000000
```

The presence of random coefficients and their correlation can be investigated using any of the three tests. Actually, three nested models can be considered, a model with no random effects, a model with random but uncorrelated effects and a model with random and correlated effects. We first present the three tests of no correlated random effects:

```
lr.mxc <- lrtest(Train.mxlc, Train.ml)
wd.mxc <- waldtest(Train.mxlc)
lh.mxc <- linearHypothesis(Train.mxlc,
                           c("time.time = 0", "time.change =  0",
                             "time.comfort = 0", "change.change = 0",
                             "change.comfort = 0", "comfort.comfort = 0"))
sc.mxc <- scoretest(Train.ml,
                    rpar = c(time = "n", change = "n", comfort = "n"),
                    R = 100, correlation = TRUE, halton = NA,
                    panel = TRUE)
sapply(list(wald = wd.mxc, lh = lh.mxc,
            score = sc.mxc, lr = lr.mxc), statpval)

##            wald      lh   score      lr
## stat    288.287 288.287 208.765 388.057
## p-value   0.000   0.000   0.000   0.000
```

The hypothesis of no correlated random parameters is strongly rejected. We then present the three tests of no correlation, the existence of random parameters being maintained.

```
lr.corr <- lrtest(Train.mxlc, Train.mxlu)
lh.corr <- linearHypothesis(Train.mxlc, c("time.change = 0",
                                          "time.comfort = 0",
                                          "change.comfort = 0"))
wd.corr <- waldtest(Train.mxlc, correlation = FALSE)
sc.corr <- scoretest(Train.mxlu, correlation = TRUE)
sapply(list(wald = wd.corr, lh = lh.corr,
            score = sc.corr, lr = lr.corr),  statpval)

##            wald      lh  score     lr
## stat    103.195 103.195 10.483 42.621
## p-value   0.000   0.000  0.015  0.000
```

The hypothesis of no correlation is strongly reject with the Wald and the likelihood ratio test, only at the 1% level for the score test.

RiskyTransport

The second example is a study by León and Miguel (2017) who consider a mode-choice model for transit from Freetown's airport (Sierra-Leone) to downtown. Four alternatives are available: ferry, helicopter, water-taxi and hovercraft. A striking characteristic of their study is that all these alternatives experienced fatal accidents in recent years, so that the fatality risk is non-negligible and differs much from an alternative to another.

For example, the probabilities of dying using the water taxi and the helicopter are respectively of 2.55 and 18.41 out of 100,000 trips. This feature enables the authors to estimate the value of a statistical life. For an individual i, the utility of choosing alternative j is:

$$U_{ij} = \beta_{il}(1 - p_j) + \beta_{ic}(c_j + w_i t_j) + \epsilon_{ij}$$

where p_j is the probability of dying while using alternative j, c_j and t_j the monetary cost and the transport time of alternative j and w_i the wage rate of individual i (which is supposed to be his valuation of transportation time). $C_{ij} = c_j + w_i t_j$ is therefore the individual specific generalized cost for alternative j. β_{il} and β_{ic} are the (individual specific) marginal utility of surviving and of expense. The value of the statistical life (VSL) is then defined by:

$$\mathrm{VSL}_i = -\frac{\beta_{il}}{\beta_c} = \frac{\Delta C_{ij}}{\Delta(1 - p_j)}$$

The two covariates of interest are cost (the generalized cost in $PPP) and `risk` (mortality per 100,000 trips). The `risk` variable being purely alternative specific, intercepts for the alternatives cannot therefore be estimated. To avoid endogeneity problems, the authors introduce as covariates marks the individuals gave to 5 attributes of the alternatives: comfort, noise level, crowdedness, convenience and transfer location. We first estimate a multinomial logit model.

```r
data("RiskyTransport", package = "mlogit")
RT <- mlogit.data(RiskyTransport, shape = "long", choice = "choice",
                  chid.var = "chid", alt.var = "mode", id.var = "id")
ml.rt <- mlogit(choice ~ cost + risk  + seats + noise + crowdness +
                    convloc + clientele | 0, data = RT, weights = weight)
```

Note the use of the `weights` argument in order to set weights to the observations, as done in the original study. The ratio of the coefficients of risk and of cost is 9.84 (hundred of thousands of $), which means that the estimated value of the statistical life is a bit less than one million $. We next consider a mixed logit model. The coefficients of `cost` and `risk` are assumed to be random, following a zero-bounded triangular distribution.

```r
mx.rt <- mlogit(choice ~ cost + risk  + seats + noise + crowdness +
                    convloc + clientele | 0, data = RT, weights = weight,
                rpar = c(cost = 'zbt', risk = 'zbt'),
                R = 100, halton = NA, panel = TRUE)
```

```
screenreg(list('mult.logit' = ml.rt, 'mixed logit' = mx.rt), digits = 4)

##
## ==============================================
##                 mult.logit      mixed logit
## ----------------------------------------------
## cost               -0.0095 ***     -0.0187 ***
##                    (0.0011)        (0.0013)
## risk               -0.0939 ***     -0.1030 ***
##                    (0.0110)        (0.0159)
## seats               0.1517          0.1085
##                    (0.2443)        (0.2333)
## noise              -0.0290          0.1423
##                    (0.2654)        (0.2289)
## crowdness          -0.9186 ***     -0.7157 **
##                    (0.2445)        (0.2225)
## convloc            -0.3772         -0.1497
##                    (0.2016)        (0.1971)
## clientele          -0.2567         -0.3314
##                    (0.2651)        (0.2541)
## ----------------------------------------------
## AIC              3250.7472       3177.2501
## Log Likelihood  -1618.3736      -1581.6251
## Num. obs.        1793            1793
## ==============================================
## *** p < 0.001, ** p < 0.01, * p < 0.05
```

Individual-level parameters can be extracted using the fitted method, with the type argument set to `parameters`.

```
indpar <- fitted(mx.rt, type = "parameters")
```

We can then compute the VSL for every individual and analyse their distribution, using quantiles and plotting the empirical density of VSL for African and non-African travelers (as done in León and Miguel 2017, table 4, p.219 and figure 5, p.223).

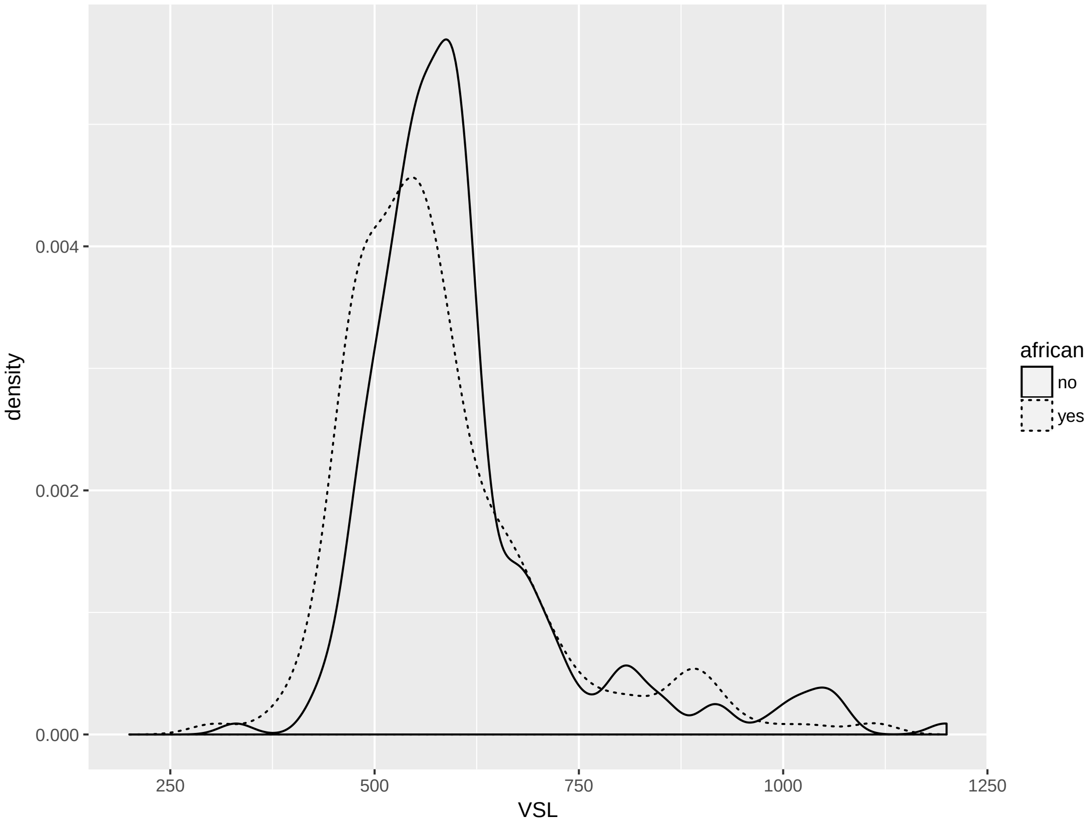
```
indpar$VSL <- with(indpar, risk / cost * 100)
quantile(indpar$VSL, c(0.025, 0.975))

##      2.5%     97.5%
##  432.4132 1054.3365

mean(indpar$VSL)
```

```
## [1] 608.9317

library("ggplot2")
indpar <- merge(unique(subset(RT, select = c("id", "african"))), indpar)
ggplot(indpar) + geom_density(aes(x = VSL, linetype = african)) +
    scale_x_continuous(limits = c(200, 1200))
```



5. Conclusions

mlogit estimates a large set of random utility models with a unified and friendly interface. Some of these models haven't been presented in this article, namely the rank-ordered logit model, the overlapping nested logit model, the paired combinatorial logit model and the multinomial probit model. Moreover, it provides usefull functions and methods which compute and return useful results, like predicted probabilities, inclusive values, marginal effects and consumer surplus.

References

Ben-Akiva M, Bolduc D, Bradley M (1993). "Estimation of travel choice models with randomly distributed values of time." *Transportation Research Record*, **1413**, 88–97.

Bhat C (1995). "A heterocedastic extreme value model of intercity travel mode choice." *Transportation Research B*, **29**(6), 471–483.

Daly A (1987). "Estimating 'tree' logit models." *Transportation Research B*, pp. 251–267.

Forinash CV, Koppleman FS (1993). "Application and interpretation of nested logit models and intercity mode choice." *Transportation Record*, **1413**, 98–106.

Fowlie M (2010). "Emissions Trading, Electricity Restructuring, and Investment in Pollution Abatement." *American Economic Review*, **100**(3), 837–69. doi:10.1257/aer.100.3.837. URL http://www.aeaweb.org/articles?id=10.1257/aer.100.3.837.

Fox J, Weisberg S (2010). *An R Companion to Applied Regression*. Second edition. Sage, Thousand Oaks CA. URL http://socserv.socsci.mcmaster.ca/jfox/Books/Companion.

Hasan A, Wang Z, Mahani A (2016). "Fast Estimation of Multinomial Logit Models: R Package mnlogit." *Journal of Statistical Software, Articles*, **75**(3), 1–24. ISSN 1548-7660. doi:10.18637/jss.v075.i03. URL https://www.jstatsoft.org/v075/i03.

Head K, Mayer T (2004). "Market Potential and the Location of Japanese Investment in the European Union." *The Review of Economics and Statistics*, **86**(4), 959–972. doi:10.1162/0034653043125257. https://doi.org/10.1162/0034653043125257, URL https://doi.org/10.1162/0034653043125257.

Heiss F (2002). "Structural choice analysis with nested logit models." *The Stata Journal*, **2**(3), 227–252.

Hensher DA, Greene WH (2002). "Specification and estimation of the nested logit model: alternative normalisations." *Transportation Research Part B*, **36**, 1–17.

Koppelman FS, Wen CH (1998). "Alternative nested logit models: structure, properties and estimation." *Transportation Research B*, **32**(5), 289–298.

Koppelman FS, Wen CH (2000). "The paired combinatorial logit model: properties, estimation and application." *Transportation Research B*, **34**, 75–89.

León G, Miguel E (2017). "Risky Transportation Choices and the Value of a Statistical Life." *American Economic Journal: Applied Economics*, **9**(1), 202–28. doi:10.1257/app.20160140. URL http://www.aeaweb.org/articles?id=10.1257/app.20160140.

McFadden D (1974). "The measurment of urban travel demand." *Journal of public economics*, **3**, 303–328.

McFadden D (1978). "Spatial interaction theory and planning models." In A Karlqvist (ed.), *Modeling the choice of residential location*, pp. 75–96. North-Holland, Amsterdam.

Meijer E, Rouwendal J (2006). "Measuring welfare effects in models with random coefficients." *Journal of Applied Econometrics*, **21**(2), 227–244. doi:10.1002/jae.841. https://onlinelibrary.wiley.com/doi/pdf/10.1002/jae.841, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.841.

Sarrias M, Daziano R (2017). "Multinomial Logit Models with Continuous and Discrete Individual Heterogeneity in R: The gmnl Package." *Journal of Statistical Software, Articles*, **79**(2), 1–46. ISSN 1548-7660. doi:10.18637/jss.v079.i02. URL https://www.jstatsoft.org/v079/i02.

Small KA, Rosen HS (1981). "Applied welfare economics with discrete choice models." *Econometrica*, **49**, 105–130.

Swait J, Louviere J (1993). "The role of the scale parameter in the estimation and use of multinomial logit models." *Journal of Marketing Research*, **30**.

Toomet O, Henningsen A (2010). *maxLik: Maximum Likelihood Estimation*. R package version 0.8-0, URL http://CRAN.R-project.org/package=maxLik.

Train KE (2003). *Discrete choice methods with simulation*. Cambridge University Press, Cambridge, UK.

Zeileis A, Croissant Y (2010). "Extended Model Formulas in R: Multiple Parts and Multiple Responses." *Journal of Statistical Software*, **34**(1), 1–13. URL http://www.jstatsoft.org/v34/i01/.

Zeileis A, Hothorn T (2002). "Diagnostic Checking in Regression Relationships." *R News*, **2**(3), 7–10. URL http://CRAN.R-project.org/doc/Rnews/.

Affiliation:

Yves Croissant
Faculté de Droit et d'Economie
Université de la Réunion
15, avenue René Cassin
BP 7151
F-97715 Saint-Denis Messag Cedex 9
Telephone: +33/262/938446
E-mail: yves.croissant@univ-reunion.fr