

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

5,100

Open access books available

126,000

International authors and editors

145M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



---

# The Application of Discrete Choice Models in Transport

---

Foued Aloulou

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.74955>

---

## Abstract

The discrete choice models are presented as a development and a renovation of the classical theory of choice. They have been based on the premise that the choice of economic agents is most often based on mutually exclusive alternatives or solutions, so that if the individual chooses one, he gives up the choices of the others. In this case, we speak of a discrete choice. Contrary to the microeconomic approach, discrete choice models consider that the environment that shapes the behavior of the choice of an individual is random and specific to each situation. It is influenced by a number of factors in relation to both the socioeconomic characteristics of the individual in question and the attribute being chosen and the circumstances that characterize the environment of choice. This process makes it possible to better disaggregate and personalize the behavior of economic agents and to perceive their preferences according to their motives and characteristics. The objective of this chapter is to highlight the application of these discrete choice models on the transport economy by specifying their contribution to the estimation of the transport demand and the evaluation of the severity of the accidents of the road, after having described the specificities of these models and their main characteristics and methods of application.

**Keywords:** discrete choice models, random utility, logit model, unordered multinomial logit, binary variables, disaggregated models, microeconomic analysis, probability of road accident, accident gravity, behavior of the modal choice for transport, values of time

---

## 1. Introduction

Discrete choice models are presented as a development and a renovation of the classical choice theory. They have overcome the rigidities and inadequacies of consumer behavior study by mentioning the problems of economic agent choices in a random and specific environment for each situation involving the choice between mutually exclusive alternatives.

---

These situations of choice encountered in reality do not fit with the classical assumptions of consumer theory according to which the goods are perfectly divisible and the problem of choice concerns a continuum of possibilities.

We do not seek to calculate the quantities of the various goods that an individual will need, but to determine the choice between mutually exclusive goods or alternatives from which the individual selects only one that maximizes her utility while taking into account its socioeconomic characteristics conditions and those of the alternative to be chosen.

In addition, the classical microeconomic approach considers that the chosen environment is static, stable, and transparent and that the individuals' decisions are rational and typical, so that the individual choice is deterministic and repetitive.

These hypotheses have limited the field of research in the analysis of demand and individual behavior of consumers. This demand was analyzed using an aggregated approach to macroeconomic variables.

In contrast to this approach, discrete choice models consider that the environment that shapes the individual choice behavior is random and specific to each situation. It is influenced by a number of factors in relation to both the socioeconomic characteristics of the individual in question and the attribute being chosen and the circumstances that characterize the environment of choice.

As a result, the decision-making process of economic agents, which is based on the maximization of an objective function under constraint, is represented by a description of the different characteristics, both the attributes of the alternatives to be considered and the socioeconomic characteristics of consumers as well as the environment of choice.

Each individual has an objective function that seeks to optimize it to achieve the best decision. In a random, uncertain environment, where the choice is not reproducible, this objective function is of random type, and the economic calculation is probabilistic [1].

This process allows to better disaggregate and personalize the economic agent behavior and to perceive their preferences according to their motives and characteristics.

This contribution will enable economists to detect the effect of each element determining the consumer's choice on their consumption behavior, as well as to detail and explain the function of individual and global demand.

These discrete choice models have been very successful thanks to their ability to analyze the random behavior of individuals when making a decision to choose a given solution or to appreciate the valuation of goods or actions.

They have been the subject of several theoretical developments and empirical validations. Their manipulation has become easier thanks to the availability of increasingly disaggregated data and advances in econometric techniques and software.

They were applied for the first time to estimate transport demand. They were subsequently generalized and applied to deal with all the problems of choice concerning mutually exclusive alternatives or also to assess the subjective value of an event.

The transport economy is a privileged domain of application of these probabilistic models. Indeed, the individual who travels every day must choose a particular mode of transport, such departure time or such other, such or such journey, such destination or such other, such or such movement frequency, etc.

The risk analysis of road accidents in terms of frequency and/or severity should then predict the probability that an individual with specific socioeconomic characteristics and driving in a given traffic environment is involved in a road accident and/or that the accident incurred will be of a given severity.

This type of model seeks to study the behavior of transport users regarding their choice of mode of transport or also the risk of transport and to anticipate the modifications brought by changes in the mode characteristics or socioeconomic variables of the decision-maker.

Several families of discrete choice models have been developed and applied (probit, logit, dichotomic logit, multinomial logit, conditional logit, mixed logit, nested logit, etc.), each of which is specified either by the nature of the explanatory variables selected and which characterize the alternatives and/or the individuals or by the statistic distribution law that follows the error terms or its ability to overcome the constraint of independence from irrelevant alternatives (IIA).

The aim of this chapter is to present these discrete choice models while focusing on the unordered multinomial logit model that is most used in empirical studies. This chapter will consist of two parts, the first of which will present the specificities of the multinomial logit model while reviewing its main tools for estimating and testing statistical validation and the interpretation of its coefficients. In the second part, we will try to apply this model to two studies on phenomena related to transport. The first concerns the modal choice of urban transport users for personal travel reasons in the city of Sousse (Tunisia), and the second phenomenon will deal with accidentology by trying to estimate and analyze the severity levels of road accidents in Tunisia.

The general problem of the first application therefore concerns the estimation of the urban passenger transport demand structure for the city of SOUSSE, using discrete choice models.

These models calculate, from a given observations, the probability that an individual selects a particular mode of transportation from a set of possible and mutually exclusive choices. In the second application, we seek to predict the probability of a driver's exposure to a given gravity accident. This severity may depend on three components: the driver; the vehicle, its condition of use; and the infrastructure.

These various components constitute the traffic system and determine road safety. They interact at a given time and place to explain the occurrence and severity of an accident. Several quantitative and qualitative variables can be identified and measured to describe these components. The purpose of this study is to show how and by how much these explanatory variables affect the severity of a traffic accident. The structure of the estimates is based on disaggregated data collected: on the one hand, the household-displacement survey carried out in 2004 and, on the other hand, the survey sheets proposed by the National Observatory of Circulation (Tunisia). We will then present and interpret the main results of the estimation of

these two applications. This interpretation remains a difficult exercise, especially when one addresses an uninitiated public to this type of discrete choice modeling and qualitative econometrics.

## 2. Presentation of the multinomial logit model

Among the discrete choice models, the multinomial logit model is the most widespread and used in many different fields. This disaggregated model seeks to study the decision of choice or the perception of the value of an event among a set of mutually exclusive alternatives.

Individual choice behavior or the perception of the value of an event is considered as a selection process between several mutually exclusive contingencies that belong to a set of eventualities. The eventuality chosen by an individual will be the one that optimizes its objective function. The decision taken will therefore result from an optimization process reflecting a rational behavior of the individual. As long as the choice of the individual is established in random circumstances that never occur identically, the modeling will be probabilistic. Nobody can correctly predict the choice of the individual, but he can estimate the probability of this choice according to the circumstances of choice and the socioeconomic characteristics of the individual as well as the technique of the alternative to choose.

The multinomial logit model will therefore allow us to estimate the probability that an individual  $i$  chooses an alternative  $j$  in given circumstances characterizing the environment of choice. This probability can be expressed as a linear (or nonlinear) function of all the variables characterizing this environment of choice ( $X_k$ ).

Formally, this probability is written according to the following expression:

$$P_{ij} = F_{ij} \left( \sum_{k=1}^K \alpha_k X_k \right) \quad (1)$$

$P_{ij}$  is the probability that an individual  $i$  establishes the choice  $j$ .

The parameters  $\alpha_k$  are unknown that we seek to estimate. They, respectively, reflect the weight of each explanatory variable ( $X_k$ ) in the determination of the probability  $P_{ij}$ .

$F_{ij}$  is a distribution function of the explanatory variables and the vector of parameters  $\alpha_k$ .

In discrete choice models, the endogenous variable we seek to explain is a qualitative and discrete variable. It illustrates the individual's choice or level of appreciation of the psychological value of a given event. This variable to be explained will take integer values that vary between 1 and  $J$  depending on the number of alternatives that make up the entire choice of the individual.

### 2.1. Specificity of the model

For a more detailed discussion, consider that an individual  $i$  of a sample  $N$  (such as  $i = 1 \dots N$ ) is in front of a set of choices (modes of transport, port of call, types of equipment, place of

residence, etc.) or belongs to a given category of population or appreciation of a psychological value of a given phenomenon (risk of accident, time value, etc.)  $j$  ( $j \in J/j = \{1,2,3...J\}$ ).

Individual  $i$  chooses the alternative  $j$  that optimizes (maximizes or minimizes) its objective function ( $S_i$ ).

The variable to be explained is expressed as follows:

$$\left\{ \begin{array}{l} Y_i = 1 \text{ if the individual } i \text{ chooses option 1, means when } S_{i1} = \underset{j=1...J}{\text{Max}} (S_{ij}) \\ Y_i = 2 \text{ if the individual } i \text{ chooses option 2 } \left( S_{i2} = \underset{j=2...J}{\text{Max}} (S_{ij}) \right) \\ Y_i = J \text{ if the individual } i \text{ chooses option } J \left( S_{iJ} = \underset{j=1...J}{\text{Max}} (S_{ij}) \right) \end{array} \right. \quad (2)$$

$Y_i$  designates the choice observed and  $S_{ij}$  the level of objective function that the choice of the alternative  $j$  gives to the individual  $i$ .

The objective function of the individual  $i$  is dependent on the socioeconomic characteristics of the individual  $i$  ( $X_{ik}$ ), on the technical ones of the option to be remembered ( $W_{jh}$ ), and on those of the environment of choice ( $E_{jm}$ ):

$$S_{ij} = S(X_{ik}, W_{jh}, E_{jm}) \quad (3)$$

It should be emphasized that these variables may be specific to each option  $j$  and/or to each individual  $i$ .

A specific variable to the individual is a variable that remains the same regardless of the option chosen by the individual, while a specific variable to the alternative  $j$  depends on the specific conditions to the choice.

As long as the objective function is random, we can break it down into two parts: one is deterministic ( $V_{ij}(X_{ik}, W_{jh}, E_{jm})$ ) and the other is random ( $\varepsilon_{ij}$ ):

$$S_{ij} = V_{ij}(X_{ik}, W_{jh}, E_{jm}) + \varepsilon_{ij} \quad (4)$$

The deterministic function ( $V_{ij}$ ) reflects the perception of an average individual of the satisfaction provided by the choice of the alternative  $j$ . It can take many forms, but the linear form is the simplest to estimate and interpret:

$$V_{ij} = \alpha_{0j} + \sum_{k=1}^K \alpha_{jk} X_{ik} + \sum_{h=1}^H \beta_h W_{jh} + \sum_{m=1}^M \mu_m E_{jm} \quad (5)$$

The arguments of this deterministic function can be quantitative as well as qualitative variables expressed in the form of a binary variable and/or polytomous.

The weighting coefficients of the explanatory variables  $\alpha_{jk}$ ,  $\beta_h$  and  $\mu_m$  reflect the relative importance of each of the explanatory variables relating, respectively, to the socioeconomic

characteristics of the individual, the attributes of the alternative, and the environment of choice, in the explanation of the objective function.

However, these coefficients cannot be directly interpreted as the impact of the absolute or relative variation of one of the explanatory variables on the probability of choosing alternative  $j$  (or belonging to a population category  $j$ ). They indicate only the variation direction of this probability but not their amplitude. If they are positive, they positively affect the probability of choice and vice versa. Moreover, the interpretation of these parameters is not identical between the explanatory variable categories [2].

$\alpha_{0j}$  is a constant that can reflect the impact of the other explanatory variables not included in the model for one reason or another and the imbalance observed in the sample between the individuals choices. Probably the individuals who opt for choice 1 will be more numerous than those opting for the second or the  $j$ th choice.

The random term of the objective function ( $\varepsilon_{ij}$ ) reflects the not observed behavior of individuals. Thus, two individuals with the same observed characteristics and faced with the same set of choices can make different decisions. It therefore implies the probabilistic nature of discrete choice models. It originates from several sources such as the measurement error on the variables or in the objective function specification, etc. [3].

The specification of the statistic distribution law of this random part makes it possible to define the definitive profile of the choice probability function ( $P_{ij}$ ). Various specifications of this law were used, but only two were mainly retained: a Weibull distribution (logit model) [4] and a multidimensional normal distribution (probit model) [5].

The individual  $i$  will choose the alternative  $j$  from a set of alternative  $J$ , if and only if  $S_{ij} > S_{il}$ . The probability of this choice is

$$P_{ij} = \Pr (S_{ij} > S_{il}) = \Pr (V_{ij} + \varepsilon_{ij} > V_{il} + \varepsilon_{il}) = \Pr (V_{ij} - V_{il} > \varepsilon_{il} - \varepsilon_{ij}) \forall j \neq l \in J. \quad (6)$$

If the error terms are independent and identically distributed according to Weibull<sup>1</sup>'s law, the probability given by the logit model is expressed by the following relation:

$$P_{ij} = \frac{\exp(V_{ij})}{\sum_{j=1}^J \exp(V_{ij})} \quad (7)$$

By respecting the laws of probability such as  $0 < P_{ij} < 1$  and  $\sum_{j=1}^J P_{ij} = 1$ , the probability associated with  $J$ th alternative does not need to be specified since it can be calculated from the rest of the calculated probabilities. This excluded alternative of the model will be considered as the reference situation that one seeks to compare it with the observed situation. The coefficients associated with this alternative  $J$  will be considered null ( $\alpha_{kJ} = \beta_{hJ} = \mu_{mJ} = 0$ ):

<sup>1</sup>A random variable follows a Weibull or double exponential law or Gumbel distribution, if its cumulative function is written:  $F(\varepsilon) = \exp(-\exp(-\varepsilon))$

$$P_{ij} = \frac{\exp(V_{ij})}{1 + \sum_{j=1}^{J-1} \exp(V_{ij})} \quad \forall j = 1 \dots J-1 \quad (8)$$

$$P_{iJ} = 1 - \sum_{j=1}^{J-1} P_{ij} = \frac{1}{1 + \sum_{j=1}^{J-1} \exp(V_{ij})} \quad (9)$$

The ratio between Eqs. (8) and (9) gives the following expression:  $\forall j = 1 \dots J-1$

$$\frac{P_{ij}}{P_{iJ}} = \exp(V_{ij}) \Rightarrow \text{Log}\left(\frac{P_{ij}}{P_{iJ}}\right) = V_{ij} = \alpha_0 + \sum_{k=1}^K \alpha_{jk} X_{ik} + \sum_{h=1}^H \beta_h W_{jh} + \sum_{m=1}^M \mu_m E_{jm} \quad (10)$$

$$\frac{\partial \text{Log}\left(\frac{P_{ij}}{P_{iJ}}\right)}{\partial X_{ik}} = \alpha_{jk} \quad (11)$$

## 2.2. Model interpretation

Unlike linear regression econometric models whose estimated coefficients can be easily interpreted as the elasticity's or the marginal impact of the explanatory variable on the variable to be explained, the interpretation of the coefficients of the logit model is more delicate.

To understand the interpretation of these coefficients, we must proceed with a reorganization of the logit model equation (Eq. (8)). It was better to express the probability of each alternative  $j$  with respect to the reference situation assumed beforehand (alternative  $J$ ). For all  $j = \{1, \dots, J-1\}$ , we must calculate the ratio between the probability of the choice of the alternative  $j$  and that of the alternative  $J$  (Eq. (10)). When only one explanatory variable varies (we go from  $X_{k0}$  to  $X_{k1}$ ), while keeping the other variables constant, we can measure its effect on the probability ratio between the observed alternative and the reference one:

$$\frac{P_{ij}(X_{k0})}{P_{ij}(X_{k1})} = \frac{P_{ij}(X_{k0})/P_{iJ}(X_{k0})}{P_{iJ}(X_{k1})/P_{iJ}(X_{k1})} = \frac{\exp(V_{ij}(X_{k0}))}{\exp(V_{ij}(X_{k1}))}$$

$$\text{Log}\left(\frac{P_{ij}(X_{k0})}{P_{ij}(X_{k1})}\right) = V_{ij}(X_{k0}) - V_{ij}(X_{k1}) = \alpha_{jk}(X_{k0} - X_{k1}) \quad (12)$$

$\alpha_{jk}$  measures the effect of changing the  $X_k$  variable from variable  $X_{k0}$  to  $X_{k1}$  on the probability of choosing alternative  $j$  rather than the reference alternative  $J$ .

When only one explanatory variable varies (e.g., we go from  $X_{k0}$  to  $X_{k1}$ ), while keeping the other variables constant, we can measure its effect on the probability ratio between the observed alternative and the reference one.

In the multinomial logit model, several categories of explanatory variables of both qualitative and quantitative orders can be integrated. The interpretation of continuous variables of a quantitative nature does not pose any problem. The exponential value of the coefficient

associated with this variable measures the unit variation impact of this explanatory variable on the probability of choosing the alternative  $j$  rather than the reference alternative  $J$ .

For qualitative variables, we distinguish between binary ones which will be coded in 0 and 1 and those polytomous which express themselves in several modalities. For example, sex as a variable characterizing the individual can be integrated in the model as a binary variable coded 0 if the individual is male and 1 if he is female. For the professional category variable of the individual, there are more than two functions. In this case, the integrated variable must take an integer from 1 to  $n$  according to the number of observed professions. For these explanatory variable categories, a reference situation must always be chosen in order to interpret their estimation coefficients. For binary variables, if the reference situation is the one relating to the code 0 (e.g., male sex), the exponential function of the associated coefficient is interpreted as the effect of the individual passing from the reference situation (0) to that observed (1) on its probability of choosing the alternative  $j$  rather than the reference alternative  $J$ .

The interpretation of these estimation coefficients becomes more difficult in the presence of a polytomous explanatory variable. These modalities present a collinearity that must be avoided by eliminating a modality and limiting itself to reasoning only according to  $n-1$  modalities which remain. The eliminated modality will be the one of reference.

In this case, the associated coefficient must be interpreted as a function of two references: one relative to the choice  $J$  and the other relating to the explanatory variable. For example, if the explanatory variable is the socioprofessional category of the individual and the reference category is the "worker," the exponential function of the coefficient associated with the "manager" variable, for example, indicates the impact of being a "manager" rather than a "worker" on the probability of choosing the alternative  $j$  rather than the reference alternative  $J$ .

We can also evaluate the impact of the variation of the explanatory variable on the comparative probability of the individual choice by the elasticity. Elasticity is defined as a percentage change in the probability of choosing alternative  $j$  rather than alternative  $J$  resulting from a 1% change in one of the characteristics of alternative  $j$  ( $W_j$ ) by keeping the other arguments of the probability function constant. The advantage of the interpretation of coefficients in terms of elasticity than unitary variation lies in the fact that the elasticity is calculated independently of the units of measurement of the explanatory variables.

The elasticity calculation constitutes a very indispensable information base for decision-makers to learn about the most influential factors in the individual behavior and determine their optimal action plan in order to achieve their goals.

The elasticity can be calculated with respect to all the arguments of the probability function. We speak of direct elasticity when it is calculated with respect to the arguments relating to the chosen alternative  $j$  and of the cross elasticity, when it is calculated with respect to the arguments relative to the other alternatives  $l \neq j$ .

This direct individual elasticity is written:

$$\ell P_j / W_j = \frac{\partial \text{Log} P_{ij}(W_{jh})}{\partial \text{Log} W_{jh}} = \beta_h W_{jh} (1 - P_{ij}) \quad (13)$$

where  $W_{jh}$  is the  $h$ th argument of the vector characterizing the alternative  $j$  ( $W_j$ ),  $\beta_h$  being its relative parameter, and  $P_{ij}$  is the probability of choice of the eventuality  $j$  by the individual  $i$ .

The cross elasticity is written:

$$\ell P_j / W_{hl} = \frac{\partial \text{Log} P_{ij}}{\partial \text{Log} W_{hl}} = -\beta_h W_{hl} P_{ij} \quad (14)$$

### 2.3. Property of independence from irrelevant alternatives

The logit multinomial model is based on a fundamental assumption but constraining in empirical studies: independence from irrelevant alternatives (IIA).

This hypothesis implies that the choice of the individual will always be the same regardless of the number of alternatives proposed, so that the probability that an individual chooses an alternative  $j$  remains constant even if other alternatives are included in the set of considered choices. This assumption imposes the independence between the alternatives, which excludes any possibility of substitution between them. It implies that the ratio of probabilities of choice between two alternatives remains unchanged following the addition or the removal of one or more alternatives from all the choices.

This property (IIA) facilitates estimation and prediction because it implies that the model can be estimated from binomial choice data or by reduced attention to choices in a limited subset of the total set of choice. Therefore, if the assumption (IIA) is verified, the model structure and the estimated parameters for the explanatory variables should remain unchanged when performing the estimate on a small subset of the set of choices.

However, this hypothesis of the logit model has been criticized by several authors, thus limiting its practical relevance. The nested logit model has been developed to overcome this property of IIA. Referring to Eq. (10), we find that the probability ratio between the two alternatives  $j$  and  $J$  does not depend on the other possible alternatives, hence the property of the independence from irrelevant alternatives (IIA).

## 3. Case studies

The multinomial logit model has been the subject of several empirical studies on the analysis of various behavioral phenomena of the individual such as the choice of modes of transport [6, 7], the choice of ports of call [8, 9], the choice of the professional function [10], the choice of place of residence [11–13], discrimination in the job market [14], the severity of road accidents [15, 16], the valuation of transport time [17, 18], etc.

In this section, two case studies will be analyzed and interpreted and treated in the case of my research work and supervision on topics particularly related to transport economics. The first study deals with the modal choice problem and the second with road accidentology. These case studies will allow us to better value the practical interest of these models of discrete choice, to account for the diversity of fields of application of these models and to present real results allowing a better understanding of the coefficient interpretation according to the qualitative and quantitative nature of the explanatory variables.

### 3.1. Modal choice study

In the first case study, we will analyze the transport behavior of transport users by estimating an unordered multimodal logit model on a sample of urban transport users from the city of Sousse (Tunisia). This study will allow us to analyze the transport demand and to identify several information about the direct and indirect elasticities of transport demand in relation to the different attributes of the modes envisaged (transport price, travel time, waiting time, etc.) and to calculate the psychological value of transport time.

The behavior of individual choice in the transport market is considered as a selection process between several modes of transport available (car, bus, metro, two-wheeled vehicle, etc.). The transport user will choose the mode that maximizes its utility.

However, this utility is unobserved. What we actually perceive is the modal choice of the user. In this context the variable to be explained will be the choice established by the transport user and not its utility.

This endogenous variable is thus discrete and qualitative which will take a limited number of integer values, whose each value illustrates a particular choice. This is the foundation of the discrete choice model.

We assume that the choice modal set is composed of three modes such as the private car, bus, and taxi ( $j = 1, 2, 3$ ).

The variable to be explained is expressed by the following system:  $\forall i = 1 \dots n$ :

$$Y_{ij} = \begin{cases} 1 & \text{if the user } i \text{ prefers the private car (PC) to other modes} \\ 2 & \text{if the user } i \text{ prefers the taxi to the other modes} \\ 3 & \text{if the user } i \text{ prefers the bus to other modes} \end{cases} \quad (15)$$

To avoid collinearity between modal choices, we eliminate the third choice (bus) while considering it as the reference situation. This reference situation will serve us to better interpret our results and evaluate the impact of changing explanatory variables on the probability of choosing the mode  $j$  (PC or taxi) rather than the bus mode.

The user  $i$  that prefers the private car to the bus mode implies that he gets more satisfaction by using the private car than the bus to get to work. This satisfaction can be systematized by a linear indirect utility function.

$Y_{ij} = 1$  (choice of the PC) if and only if  $U_i$  (PC) >  $U_i$  (bus) and  $U_i$  (PC) >  $U_i$  (taxi).

Formally, the indirect utility function  $U_{ij}$  depends on a certain number of variables relating to the attributes of the chosen transport mode ( $W_j$ ) as well as to the user's socioeconomic characteristics ( $X_i$ ).

Many explanatory variables can be integrated and tested which characterize as well the individual as the attributes of the mode to choose.

For example, four explanatory variables characterizing the transport user such as income, sex, age, and household size and three explanatory variables characterizing the modes such as the kilometric price of the use of each mode, travel time, and access time to each mode. All variables are continuous except the sex will be expressed as a binary variable coded 0 if the user is female and 1 otherwise. The price, travel time, and access time vary for the same individual from one mode of transport to another, while the variables characterizing the user do not vary according to the mode.

With reference to Eq. (10), our model will be expressed by the following relation:

$$\begin{aligned} \text{Log}\left(\frac{P_{ij}}{P_{ij}}\right) &= U_{ij} = \alpha_{0j} + \sum_{k=1}^4 \alpha_{jk} X_{ik} + \sum_{h=1}^2 \beta_h W_{jh} \\ &= \alpha_{0j} + \alpha_{1j} R_i + \alpha_{2j} S_i + \alpha_{3j} A_i + \alpha_{4j} D_i + \beta_1 P_j + \beta_2 tp_j + \beta_3 ta_j \end{aligned} \quad (16)$$

where  $R_i$ ,  $S_i$ ,  $A_i$ ,  $D_i$ ,  $P_j$ ,  $tp_j$ , and  $ta_j$  are, respectively, income, sex, age, household size  $i$ , price, travel time, and access time by the mode  $j$ .

$\alpha_{jk}$  and  $\beta_h$  are the coefficients to estimate. The weighting coefficients relating to the socioeconomic characteristics of the users ( $\alpha_{jk}$ ) are specific to each mode of transport, while those of the attributes ( $\beta_h$ ) are constant and do not vary according to the mode or the user.

$\alpha_{0j}$  is a constant that varies from one mode to another.

The estimate of this model requires data by user-displacement couple which are collected through the household-displacement survey database dated 2004 for the city of Sousse (Tunisia). Our sample is made up of 500 households distributed homogeneously over the entire agglomeration.

We are interested to a particular aspect of displacement having a professional motive, on a path home-work that converges to the city center during the morning rush hours by bus, private car, and taxi.

**Table 1** presents the results of our estimation. It describes the estimated values of the coefficients associated to the explanatory variables; their standard error (in parenthesis), in a second column; their degree of significance in the third column; and their exponential function in the last column.

All variables are statistically significant for thresholds going from 1–10%; several indicators of quality adjustment of the model were developed to evaluate the predictive ability of the model

Variable	Coefficient	Student's T-test	Exp (coef)
Constant 1	-2.014 (0.31)	-10.66	
Constant 2	-12.82 (0.401)	-15.1	
Income 1	0.218 (0.11)	2.15	1.243
Income 2	0.853 (0.063)	32.14	1.089
Age 1	0.04 (0.013)	14.01	1.04
Age 2	-0.13 (0.0912)	-2.06	0.87
Sex			
Woman	Ref		
Man 1	-0.3341 (0.139)	-3.47	0.715
Man 2	-0.93 (0.173)	-8.7	0.394
Household size 1	0.2943 (0,112)	2.15	1.34
Household size 2	-0.25 (0.125)	-2.25	0.779
Price	-0.0167 (0,078)	-0.67	0.983
Travel time	-0.123 (0,1218)	-2.41	0.884
Access time	-0.38 (0.105)	-4.13	0.68

Standard error in parentheses:

- Number of observations = 500
- Log likelihood = - 116,6517
- Pseudo R<sup>2</sup> = 0.48.

**Table 1.** Parameter estimates of modal chose model.

(Mc Fadden's pseudo R<sup>2</sup>, Estrella indicator, Ben Akiva and Lerman indicator, etc.) [19]. According to the software used (STATA 11), only the Pseudo R<sup>2</sup> and the log likelihood are calculated. Their values show although overall; the explanatory variables selected explain at high degrees the modal choice:

$$\text{Log} \left( \frac{P_{i(j=1)}}{P_{i(j=3)}} \right) = -2.014 + 0.218 R_i + 0.04 A_i - 0.33 S_i + 0.29 D_i - 0.00167 P - 0.123 tp - 0.38 ta$$

The constant parameter illustrates the heterogeneity in the representativeness of the individual choices in our sample. This coefficient is significantly higher for the PC than the taxi, reflecting thus the higher proportion of taxi users compared to those of the PC.

We can interpret the parameter associated with an explanatory variable by fixing the other variables for a given level and varying the said variable. The exponential function of this coefficient indicates the effect of this variation on the probability of choosing the PC mode rather than the bus mode. For example, when a household's income increases by one unit, the probability of choosing the PC mode instead of the bus mode increases by 24.3% (1.243-1).

For the age variable of the users of the PC, odds ratio is 1.04. This implies that a year furthermore increased by 4% the probability of choosing the PC than the bus.

Concerning the coefficient associated with the gender variable, it is interpreted as follows: a man has 28.5% (1-0.715) and 60.6% (1-0.715) of luck less than a woman to choose, respectively, the PC and the taxi rather than the bus, everything else being equal.

The increase of the members of a household of a person increases the probability of choosing PC but brings down the probability of choosing the taxi compared to that of bus. Indeed, one more member in the family increases by 34% the probability of choosing the PC rather than the bus and decreases the probability of choosing the taxi rather than the bus of 22.1%.

In fact, by becoming a householder, we will prefer the car better than the bus thanks to its advantages of availability, flexibility, and accessibility.

For the other explanatory variables characterizing the modes of transport ( $P$ ,  $tp$ , and  $ta$ ), they negatively affect the probability of choosing both the private car and the taxi to the bus.

The estimated coefficients for these variables are, respectively,  $-0.0167$ ,  $-0.123$ , and  $-0.38$ . This implies that if the cost per kilometer of transport or the travel time or the access time to the mode of transport increases of a unit while keeping all the other variables constant, the probability of choosing the car mode compared to the bus decreases by 1.65% ( $1 - \exp(-0.016)$ ), 11.57 and 32%, respectively. The user of the car has a greater sensitivity to the transport time than the cost. This explains well the fundamental reason for the dominance of the car in the modal split, thanks to its quality of service that is better than the bus particularly in terms of access time.

So, the cost and the time of transport play a determining role in the decisions of the modal choice and affect negatively the transport demand as well as the modal sharing between the car and the bus.

The weights of the explanatory variables can be interpreted economically as the marginal utilities of each indirect utility function argument ( $U_{ij}$ ). They indicate the effect of unitary change of each variable on the utility of the mode (PC).

$$Um_{i1P}(X_k) = \frac{\partial U_{i1}(X_k)}{\partial X_k} = \alpha_{k1}$$

If  $X_i = S_i$  is the sex variable,  $\alpha_{31} = -0.33$ ; this implies that the man is less satisfied than the woman by the use of the particular car.

If  $X_i = D_i$ ,  $\alpha_{41} = 0.29$ ; this implies that the more the household is composed of a larger number of individuals, the greater its satisfaction of the use of a private car is important. One more member in the household increases the satisfaction of PC use by 0.29 units.

The weighting coefficients related to the attribute variables of the PC mode are all negative, implying that the increase in both the cost of transport induced by the increase in the fuel price or the cost of acquisition of the PC, as well as the travel time whether it is in traffic or the search for parking caused by congestion, creates a disutility for users of the PC.

We can see that the choice probability of the PC is more sensitive to the search time of parking than the travel time and the costs of displacement. The parking search time provides a triple disutility compared to that caused by the travel time by the user of the PC:

$$\frac{Um_{i1}(ta)}{Um_{i1}(tp)} = \frac{\beta_{31}}{\beta_{21}} = 3.09 = TMS_{ta/tp}$$

The ratio of the marginal utilities of the two variables  $ta$  and  $tp$  measures the marginal rate of substitution of waiting time for travel time. The user agrees to spend 3.09 minutes more on his journey to save an extra minute to search for parking to his PC.

The ratio of marginal utilities of the two variables  $tp$  and  $P$  measures the marginal rate of substitution of money for travel time:

$$\frac{Um_{i1}(tp)}{Um_{i1}(P)} = \frac{\frac{\partial U_{i1}}{\partial tp}}{\frac{\partial U_{i1}}{\partial P}} = \frac{\beta_{21}}{\beta_{11}} = 73.65$$

The PC user agrees to pay 73.65 currency units to gain a minute in his trips the equivalent of 1.82 USD per hour. The  $TMS_{tp/P}$  measures the price of time granted by the user of the PC having given socioeconomic characteristics.

The value of time is defined as the price that the individual is willing to pay to save a unit of time given its motive for displacement and its socioeconomic characteristics.

This value is obtained by comparing the coefficient associated with the time variable and the one associated with the displacement cost variable. It corresponds to the level of disutility associated with the time spent in a given path.

From these results, it is thus possible to detect the most influential determinants on the modal choice of the transport users and consequently determine the function of the transport demand.

### 3.2. Accidentology study

Discrete choice models were also used to estimate the risk of road accidents. Several authors [15, 16, 20–22] used these models to calculate the probability of occurrence of a road accident and to detect the correlation between driver behavior, the characteristics of the traffic system, and the accident severity. They tried to model the driver's accident risk perception according to a set of factors describing the traffic system. This risk perception expresses a subjective, personal, and psychological assessment of the danger that every motorist seeks to minimize. Usually, the more this risk perception is high, the more lower the accident severity will be. And the more this risk perception is weak, the higher the probability of a serious or fatal accident is high. The risk perception will influence both the occurrence of the accident and the severity of the injuries.

These disaggregated models help to better describe and analyze the risk and severity of an accident by treating each accident separately in Ref. to its circumstances and the driver's individual behavior. The general idea is that the accident severity can be explained according

to both the socioeconomic characteristics of the driver who is the victim of a road accident, and of his driving behavior, and the circumstances of the traffic (state of the vehicle, infrastructure, and meteorology).

The objective of this case study is to analyze the severity of road accidents in Tunisia. We seek to estimate a multinomial logit model to predict the probability of a driver's exposure to a given gravity accident. The structure of the estimate is based on disaggregated data collected following the study of survey sheets proposed by the National Observatory of Circulation (Tunisia). Our sample consists of 300 randomly selected traffic accident victims from survey cards dated 2010. In our model, we defined three levels of gravity such as fatal accident, injury accident, and accident-causing material damage.

The endogenous variable is an unordered multinomial variable that will be scored from one to three to indicate the severity level of the observed accident. It will be illustrated by the following system:

$$Y_{ij} = \begin{cases} 1 & \text{if the observed accident is fatal} \\ 2 & \text{if the observed accident only causes injuries} \\ 3 & \text{if the observed accident causes only material damage} \end{cases} \quad (17)$$

The objective function of the driver is his risk perception. Each driver seeks to maximize his risk perception to better estimate the danger of the road and consequently reduce the accident severity.

$Y_{ij} = 1$  if the risk perception is minimal, so that the driver may have a serious accident.

To estimate the probability of exposure of an individual  $i$  (such as  $i = 1, 2, \dots, 300$ ) to a traffic accident of severity level  $j$  (such as  $j = 1, 2, 3$ ), it is necessary to cross the multinomial variable  $Y$  with a number of explanatory variables.

Referring to the accidentology literature, this gravity may depend on three components: the driver, the vehicle and its condition of use, and infrastructure. These various components constitute the road traffic system and determine the road safety. They interact at a given time and place to explain the occurrence and severity of an accident.

Several quantitative and qualitative variables can be identified and measured to describe these components.

We designate by  $S_{ij} = S(X_{ik}, V_{ih}, R_{jl}, E_{jm})$  the objective function of the individual  $i$ .

It is dependent on both the socioeconomic characteristics of the individual  $i$  ( $X_{ik}$ ) such as sex ( $X_1$ ), age ( $X_2$ ), householder ( $X_3$ ), vigilance level ( $X_4$ ), and seat belt wearing ( $X_5$ ); vehicle-operated characteristics ( $V_{ih}$ ) such as age ( $V_1$ ), size ( $V_2$ ), speed ( $V_3$ ), and airbag equipment ( $V_4$ ); and those of the borrowed road ( $R_{jl}$ ) such as road condition ( $R_1$ ), vision ( $R_2$ ), lighting ( $R_3$ ), position of the accident ( $R_4$ ) and the environment ( $E_{jm}$ ) such as climate ( $E_1$ ), time ( $E_2$ ), and agglomeration ( $E_3$ ) (**Table 2**).

The obtained results showed that all the variables retained are statistically and theoretically significant and explain at different degrees the severity of an accident.

Explanatory variables	Coefficients	
	Fatal accident	Injury accident
Constant	0.2472(0.006)***	2.5440(0.026)**
Sex (X <sub>1</sub> )	0.6233 (0.082)*	0.6549 (0.060)*
Driver's age (X <sub>2</sub> )	0.3968 (0.030)**	-0.2925 (0.020)**
Householder (X <sub>3</sub> )	-0.1323 (0.744)	0.7078 (0.402)
Vigilance level (X <sub>4</sub> )	0.3375 (0.084)*	0.4374 (0.086)*
Seat belt wearing (X <sub>5</sub> )	0.9995 (0.001)***	-0.5509 (0.062)*
Vehicle age (V <sub>1</sub> )	0.6950(0.040)**	-0.4719(0.050)**
Vehicle size (V <sub>2</sub> )	0.7906 (0.010)***	-0.5346 (0.006)***
Speed (V <sub>3</sub> )	0.1888 (0.049)**	-0.1671 (0.075)*
Airbag equipment (V <sub>4</sub> )	0.1022 (0.002)***	-0.6098 (0.020)**
Road condition (R <sub>1</sub> )	-0.4400 (0.089)*	0.4722 (0.093)*
Vision (R <sub>2</sub> )	0.5127 (0.013)**	-0.5983 (0.004)***
Lighting (R <sub>3</sub> )	-0.7897 (0.080)*	0.8874 (0.089)*
Position of the accident (R <sub>4</sub> )	-0.1362 (0.060)*	0.2368 (0.068)*
Climate (E <sub>1</sub> )	-0.2630 (0.050)**	0.2367 (0.049)**
Time (E <sub>2</sub> )	-0.9784 (0.009)***	-0.6947 (0.010)***
Agglomeration (E <sub>3</sub> )	-0.7772 (0.012)**	0.1086 (0.015)**

Standard error in parentheses.

\*\*\*Significant to only one of 1%; \*\*significant to only one of 5%; \*significant to only one of 10%.

The incorporeal accident is the reference category:

- Number of observations = 300
- Log likelihood = - 195.969
- Pseudo R<sup>2</sup> = 0.405.

**Table 2.** Parameter estimates of accidentology study.

Referring to Eq. (10), the estimated value of the weighting coefficient of the sex variable (X<sub>1</sub>) corresponds to the ratio of relative probabilities as follows:

$$\log \left( \frac{P(Y = 1; X_1 = 1)/P(Y = 3; X_1 = 1)}{P(Y = 1; X_1 = 0)/P(Y = 3; X_1 = 0)} \right) = 0.62$$

$$\Rightarrow \frac{P(Y = 1; X_1 = 1)/P(Y = 3; X_1 = 1)}{P(Y = 1; X_1 = 0)/P(Y = 3; X_1 = 0)} = \exp(0.62) = 1.86$$

The sign of the coefficient is positive. It implies that the gender variable has a positive effect on the probability of being a victim of a fatal bodily injury rather than an intangible accident. We can interpret this coefficient as follows: a man has a probability of 86% to be the victim of a fatal accident rather than an intangible accident. This probability rate is 92.5% in the case of an injury accident.

For the vigilance variable ( $X_4$ ), the ratio of relative probabilities is equal to 1.4 ( $\exp(0.33)$ ). This implies that driving without concentration (zero vigilance: alcohol, sleeping while driving, etc.) increases the probability to be the victim of a fatal accident of 40% compared to an intangible accident and increases the probability of being injured in an accident by 55% compared to an intangible accident. Tiredness, driving drowsiness, and alcohol are the factors that increase road insecurity and the probability of having more and more serious accidents. These factors are particularly related to the irresponsible behavior of the driver.

According to the weighting coefficient of the seat belt-wearing variable ( $X_5$ ), the nonuse of the seat belt increases the probability of going from an accident with material damage to a fatal accident of 170%. However, the coefficient of this same variable ( $X_5$ ) relative to the injury accident alternative is negative. This implies that not wearing a seat belt reduces the probability of being an injury accident victim in relation to an intangible accident. Not wearing a seat belt does not prevent injury accident, but it reduces the risk of a fatal accident. Therefore, not wearing a seatbelt is a key factor in the explanation of fatal traffic accidents.

For the age variable ( $X_2$ ), 1 more year in the driver's age reduces the probability of being a fatal accident victim rather than an intangible accident of 48%. The fatal accident risk decreases with the increase of the driver's age. For the variable speed ( $V_3$ ), its coefficient is 0.188 in the event of a fatal accident and  $-0.167$  in the event of an injury accident. These coefficients are interpreted as follows: any increase in the circulation speed of 1 Km/h causes an increase in the probability of the fatal accident risk compared to an intangible accident of 20% ( $\exp(0.133) - 1$ ) and a decrease in the injury accident risk of 18% compared to the reference situation. So speed is a risk factor whose excess increases the accident severity.

Concerning the variable airbag equipment ( $V_4$ ), it positively affects the probability of occurrence of a fatal accident, but negatively that of an injury accident. In other words, a car not equipped with an airbag increases the probability of a fatal accident compared to an injury accident by 10.7%, while it decreases the probability of an injury accident by 45%.

With regard to the infrastructure characteristic vector (road condition, position of the accident, lighting, and vision at the time of the accident), it represents one of the elements that contributes to the explanation of the probability of a fatal accident risk.

All other things being equal, the driver reduces the probability of a fatal accident compared to an intangible accident by 12.7% when it avoids overtaking on a straight-line trajectory, by 35.6% when he takes a good quality road and by 54% when the road is illuminated and the vision is clear.

In terms of environmental factors, we find that the climate ( $E_1$ ), the time, and the location of the accident negatively affect the probability of occurrence of a fatal accident compared to an intangible accident. In other words, driving in an environment characterized by a normal, sunny day and in an agglomeration zone reduces the probability of a fatal accident in relation to an intangible accident by 30% compared to the driving in the rain, by 60% compared to night driving, and by 54% compared to an out agglomeration driving.

## 4. Conclusion

Discrete choice models are a valuable tool for analyzing the behavior of individuals when faced with a choice between mutually exclusive alternatives. They are based on the logic of economic rationality which aims at optimizing an objective function while taking into account both the socioeconomic characteristics of individuals and the technical-economic characteristics of the alternative to be chosen, as well as the uncertainty of the environment where the choice reigns.

This objective function is conditional, discrete, and random. It is discrete because the problem of choice is no longer a continuum of possibilities but rather mutually exclusive alternatives, so that if the individual chooses a given alternative, he must renounce others. It is random in that the individual in question does not have perfect knowledge of the value of his objective function dependent on a given choice. This function is not observable. What is known is the choice of the user and not the value of this function. The objective function is conditional because it formalizes the satisfaction of the individual under the condition that he has already chosen the preferred alternative.

The multinomial logit model is the most used in empirical studies. It has the advantage of being able to treat the individual choice between a multitude of options and seeks to estimate the probability of having chosen a given alternative that better meets the requirements of the individual and the specific conditions characterizing the environment of choice.

It predicts the effects of modifying one of the characteristics of the alternative to choose or the individual's socioeconomic variables on the probability of making a relative decision of choice.

It allows better analysis of economic phenomena in relation to human behavior as a decision-making unit such as transport demand, accidentology, and valuation of nonmarket goods (transport time, membership of a given category population, etc.).

The objective of this chapter was to provide the reader with some essential elements for putting this multinomial logit model into practice by presenting in a first part its specificities and the interpretation of its estimated coefficients. In a second part, we tried to apply this model on two cases in relation to transport, one on modal choice and the other on accidentology.

Based on the results of these two applications, several pieces of information can be deduced which may be of great practical interest to individuals and public authorities involved in the transport. They constitute an important information base which guides these economic actors to the best choices of preventive actions and the orientation of the transport policy as well as in the matter of investment, pricing, road safety, etc. They offer us the possibility to calculate a specific time value to each individual according to their socioeconomic characteristics, their modal choice, and the conditions of travel (reason for travel, zone origin destination, time of departure, etc.), to propose the best preventive actions to accidentology, etc.

## Author details

Foued Aloulou

Address all correspondence to: [aloulouf@yahoo.fr](mailto:aloulouf@yahoo.fr)

Higher Institute of Transportation and Logistics, University of Sousse, Tunisia

## References

- [1] De Palma A, Thisse J-F. Les modèles de choix discrets. *Annales d'Economie et de Statistique*. 1987;**9**:151-191
- [2] Essafi CA. Les modèles logit polytomique non ordonnés: théorie et applications, Institut National de la Statistique et des Etudes Economiques; Série des Documents de Travail de la Direction des Statistiques Démographiques et sociales: Unité Méthodes Statistiques; Série des Documents de Travail Méthodologie Statistique; 2004 ;0301. Available from: [http://master.is.free.fr/Ancien\\_site/Documents/modele\\_logit.pdf](http://master.is.free.fr/Ancien_site/Documents/modele_logit.pdf)
- [3] Manski CF. *The Analysis of Qualitative Choice*, PhD Dissertation, MIT; 1973
- [4] Mc Fadden D. *Conditional Logit Analysis of Qualitative Choice Behaviors*. New York: *Frontiers in Econometrics*; 1974
- [5] Hausman J, Wise D. A conditional Probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences. *Econometrica*. 1978;**46**(2): 403-426
- [6] McFadden D. The measurement of urban travel demand. *Journal of Public Economics*. 1974;**3**:303-328
- [7] Ben-Akiva M. Structure of passenger travel demand models. *Transportation research board record*. 1973; 526, Washington DC
- [8] Tongzon JL. Port choice and freight forwarders. *Transportation Research Part E*. 2009; **45**(1):186-195
- [9] Malchow M, Kanafani A. Disaggregate analysis of port selection. *Transportation Research Part E*. 2004;**40**(4):317-338
- [10] Boskin MJ. A conditional Logit model of occupational choice. *Journal of Political Economy*. 1974;**82**(2)
- [11] Bhat CR, Guo J. A mixed spatially correlated logit model: Formulation and application to residential choice modeling. *Transportation Research Part B*. 2004;**38**:47-168

- [12] De Palma A, Picard N, Waddell P. Discrete choice models with capacity constraints : An empirical analysis of the housing market of the greater Paris region. *Journal of Urban Economics*. 2007;**62**:204-230
- [13] Gabriel SA, Rosenthal SS. Household location and race: Estimates of a multinomial Logit model. *The Review of Economics and Statistics*. 1989;**71**(2):240-249
- [14] Schmidt P, Strauss RP. The prediction of occupation using multiple Logit models. *International Economic Review*. 1975;**16**(2)
- [15] Aloulou F, sana N. Analyse microéconométrique des accidents routiers en Tunisie. *revue économique*. 2016;**67**(6):1211-1232
- [16] Eluru N, Bhat CR. A joint econometric analysis of Seat Belt use and crash-related injury severity. *Accident Analysis and Prevention*. 2007;**39**(5):1037-1049
- [17] Blayac T, Causse A. Value of travel time; a theoretical legitimization of some nonlinear representative utility in discrete choice models. *Transportation Research Part B*. 2001;**35**:391-400
- [18] De Lapparent M. Individual demand for travel modes and valuation of time attributes within the regular journey to work framework. *L'actualité économique; Revue d'analyse économique*. 2003:42
- [19] Mc Fadden D, Hausman J. Specification test for the multinomial logit model. *Econometrica*. 1984;**52**(5):1219-1240
- [20] De Lapparent M. Willingness to use safety belt and levels of injury in car accidents. *Accident Analysis and Prevention*. 2008;**40**(3):1023-1032
- [21] Boyer M, Dionne G, et Vanasse C. Infractions au code de sécurité routière, infractions au code criminel et accidents automobiles. Publication CRT, 583, Centre de recherche sur les transports, Université de Montréal. 1988
- [22] Abdel Aty MA, et Radwan AE. Modeling traffic accident occurrence and involvement. *Accident Analysis and Prevention*. 2000;**32**(5):633-642