

# Application of Discrete Choice Models for Exploring Traveller's Behaviour at Egyptian Transport Network

Ahmed M. Semeida

Civil Engineering Department, Faculty of Engineering,  
Port Said University, Port Said, Egypt, E-mail: [asmeida@eng.psu.edu.eg](mailto:asmeida@eng.psu.edu.eg)

## ABSTRACT

Mode choice models generally form a critical part of the analysis of the travel demand of any study area in any country and have played an important role in transport modelling over the last decades. These models provide a detailed representation of the complex aspects of transport demand based on strong theoretical justifications. However, this paper describes the development of a mode choice model for the Egyptian main zones based only on the behaviour of travellers. The necessary data are collected using a special questionnaire prepared by the Egyptian Ministry of Transportation in 2010 and interviewing travellers at basic taxi terminals across Egypt. The mode choice model is derived using the multinomial logit (MNL), the nested logit (NLM), and the multinomial probit (MPM). The modelling execution results show that, for all the studied Egyptian zones, the NLM models give satisfactory and better results than the MNL and MPM models. The trip cost strongly affects the choice of paid conveyance, especially when it comes to work and educational trips. This may be apparent in the majority of developing countries, where the weak economic situation negatively impacts the use of expensive transportation means. However, the overall results of the present study are valuable and confidential, suggesting a reasonable transport policy intended for users in Egypt, the main purpose of which is the use of comfortable vehicles at moderate fares. It is advisable to investigate all aspects of the problem in order to reach an extensive modelling and analysis of mode choice. This will surely help to improve the prevailing transport system in Egypt.

*Keywords: Mode choice, discrete choice model, multinomial logit, nested logit, multinomial probit.*

## 1. INTRODUCTION

The deep comprehension of traveller's behaviour in conveyance selection is considered as one of the most necessary issues for transportation authorities. The transportation planning field mightily utilises discrete choice models for that purpose. In general, the travellers specify the suitable conveyance depending on their economic ability and personal priority. Therefore, the selection of the appropriate discrete choice model of suitable performance seems to be so sensitive, Chen et al. [1]. In the discrete choice theory, the MNL concept is frequently used by many researchers, although it has several familiar restrictions, Guadagni and Little [2]. One of the most substantial restrictions is the Independence of Irrelevant Alternatives (IIA) which was explored by McFadden [3]. To overcome this restriction, an available substitution is made by the utilization of the NLM for assessment in practical applications and to defeat the previous restriction, Guadagni and Little; de Dios Ortúzar [4, 5]. The NLM concept is concerned with the collection of the identical substitution into nests inspiring a pyramidal framework of the alternatives, Train [6]. The NLM can be employed for modelling in any condition where subsets of alternatives contribute hidden interest components, Chib et al. [7]. The NLM varies from the MNL in that the error syntheses of the options do not need to have the same distributional variant, Hensher et al. [8]. Khan et al. [9] applied MNL and NLM to derive modal split models concerning passengers for Brisbane city to predict the traveller's behaviour for the South East Queensland, Australia. The produced models were derived for eight travel aims. Two private vehicle types were involved in the resulted mode choice module and include Taxi and private car. The resulting final models had statistics within the acceptable regions and were satisfactorily high. The sensitivity analyses executed on the level-of-service characteristics explained their proportional flexibility suspected for each travel mode and for a specific trip aim. Enam and Choudhury [10] specified the essential restrictions on the obtainable information accompanied with the derivation of an inclusive modal split model for the capital

of Bangladesh, Dhaka which is the 11th largest city in the world in resident's number. They suggested procedures to conquer those restrictions. The suggested programs have a tremendous probability to promote mode choice models for cities other than Bangladesh as well as other developing countries which often face scarceness of data. Chen et al. [1] employed five standard discrete choice models for mode split prediction as MNL, NLM, heteroscedastic extreme value model, MPM and, mixed multinomial logit model (MMNL). They were implemented to predict mode split for actual state involving 210 travellers between Sydney and Melbourne, Australia. The results indicated that the NLM does well in fitting identity and frequency towards replacements, while MMNL model produces the most effective procedure for the mode choice forecast and the results obtained from a single discrete choice model are not effective. Yang et al. [11] adopted the route which joins Southwest Jiaotong University district and downtown, China, as a case study. The relations between independent variables as cost, time and the dependent variable as the choice prospect of the public transport were explained. Results showed that the ability of MNL was more reliable than NLM. The researchers raised several significant standards to promote the level of service. These included built Bus Rapid Transit systems, decreased the arrival period to Metro and bus stop, reducing parking capacity to dominate the use of vehicles. In Egypt, which is considered as a developing country, this point has not paid sufficient attention due to the lack of necessary data. A research towards this direction was published by, Semeida [12] which analyses travel demand models for Port Said governorate, northeast Egypt. This research explains this issue, with details which develop the performance of the Egyptian transportation system. Recently, it is possible to apply the discrete choice models in the Egyptian provinces due to the availability of the necessary data. This aims to perform the Egyptian transportation system. The present research demonstrates the implementation of mode choice models for the eight main zones included in the Egyptian transport network considering the traveller's path from an origin location to the specified taxi terminal with different travel aims. These models utilize specific characteristics for each traveller as, travel distance,

cost, time, aim, and number of travels per year in order to calculate the ratio of trips which are classified into, paid conveyance (taxi, shared taxi, public bus and three wheel vehicle) and non-paid conveyance (private car and walk). The previous models standardisation employs three main paradigms as a multinomial logit model (MNL), nested logit model (NLM), and multinomial probit model (MPM) and analysing 4506 trips distributed on eight main zones in Egypt, and were collected in April and May 2010. These zones are indicated in Fig. 1. The details of the present study are illustrated within the scope here.

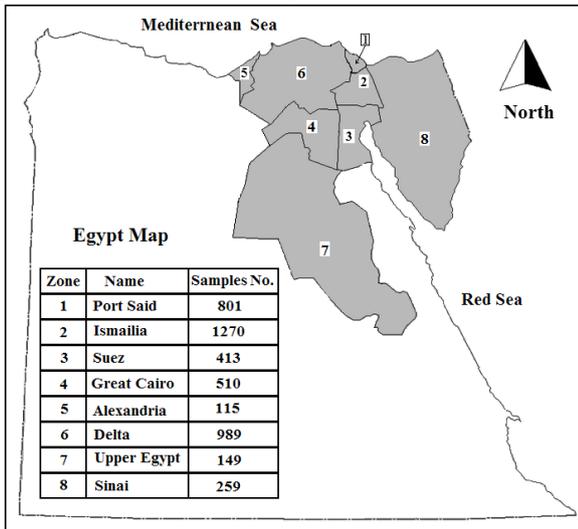


Fig. 1. Areas and code number of the zones under study on Egypt map.

## 2. DATA COLLECTION, ITS SOURCES AND PROCESSING

The studied zones cover an area of around 296,870 square kilometres (30% of Egypt area). Also, the census of these zones is around 80,000,000 residents (90% of Egypt census). These data are taken from the Egyptian Central Agency for Public Mobilization and Statics, ECAPMS [13]. The studied zones have an important position in commerce, industry, education, and tourism. However, they attract and generate yearly a huge number of personal trips from and to each other. These considerations enhance the derivation of mode choice models in these regions for the first time.

### 2.1. Data Sources

The Egyptian Ministry of transportation prepared a special questionnaire in 2010 depending on interviews with passengers at basic taxi terminals in the eight zones. The time of the interviews is from 6.00 AM to 10.00 PM in two months May and April. The results of this questionnaire consist of filling 4506 forms, distributed on the eight main zones. The number of samples at each zone is presented in Fig. 1. This questionnaire is a part of a project that has been executed in Egypt with the title (The study of preparation of inclusive transportation scheme at the national level). Two main authorities were participating in this project. First, is the Japanese International cooperation Authority (JICA), and the other is the general Authority for planning transportation projects, Ministry of Transportation, Egypt. Each form is devoted to one person and contains the following data: trip origin district to taxi terminal, trip frequency during the year, trip purpose (work, education, other), trip cost, trip time, type of conveyance, including

paid conveyance as taxi, shared taxi, public bus and three wheels vehicle or non-paid conveyance (private car and walk).

The distance between the terminal and origin district (distance on urban roads) is measured by the use of Google earth maps using the exact districts names and locations. Fig. 2 indicates the road network in Egypt.



Fig. 2. Roads network in Egypt.

### 2.2. The Used Variables

There are five independent variables that affect mode choice, including 4 numerical variables and one dummy variable as trip purpose. The trip frequency per year is calculated as follows; each form has seven choices about trip number: daily, two or three times per week, one time per week, two or three times per month, one time per month, two or three times per year, and one time per year. The traveller selects one choice of the seven; consequently, the trip frequency per year is approximately determined. There are three main purposes, the first is work, the second is education, and the third is other different purposes and includes the tour, shopping, business, and social visits. All the independent variables and their symbols are provided in Table 1. In addition, seven dependent variables that produce mode choice types are presented in the same table.

Table 1 Variables and symbols using in the present research

Variable Type	Definition	Symbol	
Independent	1- Distance between origin district and terminal in Km	D	
	2- Trip time in conveyance in minutes	Ti	
	3- Trip cost in EP	C	
	4- Trip frequency per year in trips/person/year	Tr	
Dependent	5- Trip purpose (Dummy variables)	Work Education Other	
	Paid conveyance	1- Taxi	T
		2- Shared taxi	ST
		3- Public bus	PB
Non-paid conveyance	4- Three wheel vehicle	TT	
	5- Metro (only in Great Cairo)	M	
	6- Private car	PC	
	7- Walk	W	

### 3. MODELING PROCEDURE

#### 3.1. Model Framework

Three major types of model are experiencing for the first time at the main zones in Egypt, which are; MNL, NLM, and MPM. Fig. 3 presents the paradigms of these models and also shows the seven trip conveyances involved in the choice group created for the research. The mathematical model forms comprise; three model forms illustrated as follows.

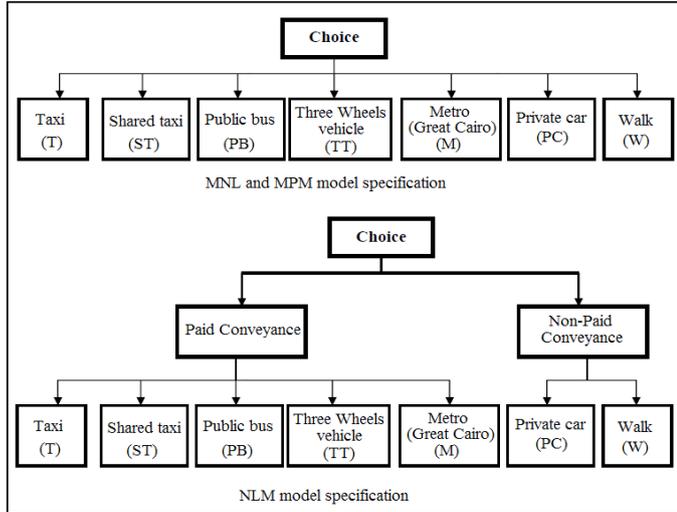


Fig. 3. A different model structure used in the present research.

##### 3.1.1. MNL model mathematical form

Domencich and McFadden [14] have pointed out that, this model is the most straightforward and workable discrete choice model. The assumption in discrete choice models is that when confronting a case option, person priority across every option can be substantiated by a benefit gauge accompanied with every option. The decision-maker is supposed to select the option that inserts the highest benefit. Anwar [15] has pointed out that, the abundant of characteristics that affect person's benefits cannot be monitored and must, therefore, be addressed as random. Thus, the benefits themselves in models are random, signifying that the choice models can grant only the prospect for which alternatives are selected, not the selection itself.

Allow  $U = (U_1 \dots U_k)$  indicates the vector of benefits accompanied by a given group of substitution,  $K$ . This group contains  $k$  alternatives having numbers  $1, 2 \dots K$ . The benefit of every substitution to a certain decision maker can be expressed as an equation of the monitored characteristics of the alternatives and the monitored characteristics of the decision maker. Allow  $\alpha$  indicates the vector of variables which contain the previous characteristics and features. Consequently,  $U_i = U_i(\alpha)$ . To merge the impacts of hidden properties and characteristics, the benefit of every substitution is presented as a random variable containing a consistent (deterministic) compound,  $V_k(\alpha)$  and an additional random "error part",  $\zeta_i(\theta, \alpha)$ . Accordingly, (McFadden [3]) the following Eq.(1) is reached:

$$U_i(\theta, \alpha) = V_i(\theta, \alpha) + \zeta_i(\theta, \alpha) \quad \forall i \in K \quad (1)$$

The random troubles  $\zeta_i(\theta, \alpha)$  can be explained as possessing several features, such as; errors in the measure of the characteristics of the data and the sharing of negligent properties. McFadden [3] has pointed out that;  $\zeta$  is forged by a group of independent similar dispensed variables, with zero average and independent of  $\theta$  and  $\alpha$ . Therefore, (From McFadden [3]) the MNL takes the following form expressed by Eq. (2):

$$P_n(i) = \frac{e^{\beta_i X_n}}{\sum_{i=1}^k e^{\beta_i X_n}} \quad i = 1, 2, 3 \dots K \quad (2)$$

Where  $P_n(i)$  is the traveller probability  $n$  selects mode  $i$ ,  $X_n$  is a vector of commensurable properties of the travel maker  $n$ , and  $\beta_i$  is a vector of estimated coefficients of criterion maximum likelihood procedures. As stated in the introduction section, MNL has a serious property of the IIA. The IIA part convenes that for a certain motorist the proportion of the selection probabilities of each two modes is completely unconstrained by any other alternatives. The IIA feature is a result of the hypothesis that the trouble parts are alternately independent. Hausman and McFadden [16] pointed out that, the ratio of choice probabilities for alternatives  $i$  and  $j$  does not depend on the characteristics of other alternatives. (According to Hausman & McFadden 1984), Eq.(3) of IIA is:

$$\frac{P(y_n = j | x_n)}{P(y_n = i | x_n)} = \frac{e(x_{nj} \beta_j)}{e(x_{ni} \beta_i)} \quad (3)$$

Hausman and McFadden [16] explored an extensive domain of appropriate calculation trials to discover contravention of the IIA hypothesis. This includes comparisons of logit models evaluated with subsets of alternatives from the global selection group. If the IIA hypothesis takes the whole selection group, then the logit model also implements to a selection of any subset of alternatives.

##### 3.1.2. NLM mathematical form

Anwar [15] has pointed out that, the NLM framework permits assessment of percentages within chosen sub-modes, before the valuation of percentages among transport modes. In his research, such model evaluates the ratios among the paid and non-paid conveyance, before evaluation of the ratios within taxi, shared taxi, public bus, three wheel vehicle, private car and walk. This capability decreases several of the constraints of the MNL, chiefly the IIA problem. It has also been detected that the choice among sub-modes may be more sensible to travel times and costs than the choice among transport modes.

Williams [17] introduced an extensive explanation of the nested logit characteristics and considers it as an expansion of the multinomial logit model prepared to specify the correlation between alternatives and is depending on the splitting of the selection group  $C$  into different nests  $C_k$ . The utility function of every substitution contains the  $i$  expression fixed to the substitution, and a term accompanied with the nest. Take  $i \in C_k$ . Williams [17] presents the utility function as is given here below in Eq. (4):

$$U_i = V_i + \varepsilon_i + V_{C_k} + \varepsilon_{C_k} \quad (4)$$

The error expressions  $\varepsilon_i$  and  $\varepsilon_{C_k}$  are supposed to be independent. In the nested logit model, the correlated alternatives are located in a "nest", which partially ejects the IIA characteristic. Fig. 3 shows a clear paradigm of the gathering of the alternatives. The "paid

conveyance" is not obtainable as an alternative because it is just a trait for a nest and can be named as "composite alternative" and true alternatives "elemental alternatives". The following Eqs. (5-8), are after Williams [17] who presented the mathematical shape of a two nested level logit model.

$$P_n = P_{n/i} P_i \quad (5)$$

$$\text{Where, } P_{n/i} = \frac{\exp(\beta x_j / i)}{\sum_{n_i} \exp(\beta x_j / i)} \quad (6)$$

$$P_i = \frac{\exp(\gamma z_i + \theta_i I_i)}{\sum_i \exp(\gamma z_i + \theta_i I_i)} \quad (7)$$

$$\text{; and } I_i = \ln \sum_{n_i} \exp(\beta x_j / i) \quad (8)$$

Where,  $P_n$  is the ideal likelihood of choice  $n$ ;  $P_{n/i}$  is the subjunctive likelihood of selecting substitution  $n$  offered that individual has chosen the choice-set  $i$ ;  $P_i$  is the probability of choice, the choice-set  $i$ ;  $z_i$  are characteristics of the choice groups;  $I_i$  is named the comprehensive worth (log sum) of choice-set  $I$ ;  $\beta$  and  $\gamma$  are vectors of coefficients to be evaluated; and  $\theta_i$  is the coefficient of the comprehensive worth (Log sum) of choice-set  $I$ . If all comprehensive worth parameters are limited to be 1, then the NLM will be the same as MNL. The NLM is convenient with random benefit maximisation if the cases' comprehensive worth parameter ( $\theta$ ) is finite between zero and one.

### 3.1.3. MPM mathematical form

Daganzo [18] proposed that the MPM can be derived if assuming that random error terms follow the normal distribution. MPM model allows the situation that not all the random error terms are independent and identically with each other. It is the most generalised model as it fully reflects the realistic choice behaviours. The MNP model can be expressed by the following Eq. (9):

$$P_{ni} = \int I(\varepsilon_{nj} < \varepsilon_{ni} + V_{ni} - V_{nj}, \forall j \neq i) \phi(\varepsilon_n) d\varepsilon_n \quad (9)$$

Where,  $I(\cdot)$  is the index function,  $I(\cdot) = 1$  means that the decision-maker has the one with max utility chosen, otherwise will be zero,  $\phi(\varepsilon_n)$  obeys multivariate normal distribution. The previous equation has very complex integral components. It is capable of handling with heteroscedasticity problems, defining error structures of any types, as well as dealing with error terms related to time series by using panel data. The only limitation of MPM model is that all the error terms of utility functions must normally distribute, Chen et al. [1].

## 3.2. Calibration of Models

The results of modelling are obtained by using XLSTAT software package [19]. The outputs include the following main estimation results:

-Assessment of individual model parameters:

This involves parameter denotations, coefficient values, standard errors of these values and the resulting Wald Chi-square statistic ( $\chi^2$ ) for every variable. The significance level selected for this study is 10%. This reflects a moderately restrictive approach to the selection of the predictive variables during modelling. This can be

achieved if the value of  $\chi^2$  is equal to, or more than, 2.706. If  $\chi^2$  value is greater than 3.842 this indicates the same at a higher 5%, significance level, Semeida [20]

- Assessment of goodness-of-fit of model:

Testing the overall goodness-of-fit is expressed by Eq. (10), from Agresti [21]:

$$2(LL(\beta) - LL(0)) \quad (10)$$

Where,  $LL(\beta)$  = the model log-likelihood of convergence; and

$LL(0)$  = the model log-likelihood with only the constant term (i.e.

With all parameters are set to zero). In addition, log-likelihood ratio index ( $\rho^2$ ) is used which can by Eq. (11):

$$\rho^2 = 1 - (LL(\beta) / LL(0)) \quad (11)$$

Where, the value of rho-squared mensuration places between 0 and 1 (this is the same as the  $R^2$  measure for linear regression models). Zero value explains that the model is worse than the original one, on the contrary, one value gives an ideal model; in which, each selection is predicted truly.

## 4. DATA ANALYSIS AND RESULTS

The analysis includes eight models for the selected zones as mentioned before. Three main types of models are applied for the first time in each of the Egyptian zones as; MNL, NLM, and MPM. The results are presented in Tables 2, 3, and 4, respectively.

### 4.1. MNL Results

Table 2 presents the estimated parameters journey to taxi terminal. The considerable characteristics are located in dark cells. The examination of the variables of the model for every zone, the determined coefficients for the trip distance, cost and time variables (level of service characteristics) has the predictable negative signs and they appropriately passed the 0.05 and 0.1 level of significance. For zones 1, 2, 4, and 8, the trip distance, time, and cost are effective. The trip time and cost are significant for zone 3 only. In addition, the trip distance and cost are significant for zones 5 and 6. Finally, the trip distance and time are significant for zone 7 only. In terms of trip frequency, it is significant only for zone 8. For trip purpose attributes (dummy variable), only education purpose is effective for two zones 2 and 8.

Reviewing the goodness of fit measures shows that, the four models of the zones 3,5 and 7 are acceptable. Considering the statistical measures of these models, the acquired log likelihood values when converging ( $LL(\beta)$ ) are -292, -16.5, -12.8, and -104.53, respectively. The overall fit of these models is rational according to a log likelihood ratio index ( $\rho^2$ ) of 0.21, 0.826, 0.527, and 0.342 that mean all variables produced in every model are statistically considerable. Also, the number of correct samples does not exceed 80 % for the previous zones. Therefore, these models still need improvement and promotion.

### 4.2. NLM Results

The results of NLM are presented in Table 3. In this model, the significant attributes are indicated in shaded cells. Reviewing the results of the model for each zone indicates that the best accuracy is verified with the MNL model. So, the use of NLM procedure is reliable and has a suitable guidance in the present case.

Table 2 Results of MNL for the eight zones under research

Variables	Coefficient & $\chi^2$							
	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	Zone 6	Zone 7	Zone 8
Constant (T)	-0.029 (0.003)	-1.973 (87.64)**	1.382 (30.61)**	1.235 (12.86)**	9.97 (3.97)**	0.188 (1.158)	--	-2.11 (14.16)**
Constant (ST)	-3.689 (39.67)**	-6.11 (243)**	-3.31 (97.02)**	-1.078 (10.18)**	0.412 (0.02)	-4.04 (266.8)**	Ref.	Ref.
Constant (PB)	--	--	--	--	--	-4.89 (306)**	--	--
Constant (TT)	--	-0.09 (0.187)	--	--	--	2.134 (119.5)**	-1.04 (0.48)	--
Constant (M)	--	--	--	Ref.	--	--	--	--
Constant (PC)	Ref.	Ref.	Ref.	--	Ref.	Ref.	--	--
Constant (W)	6.055 (103.94)**	1.23 (36.4)**	4.49 (126.3)**	3.123 (63.64)**	--	2.56 (159.7)**	2.89 (2.76)*	0.1 (0.04)
D	-1.103 (81.13)**	-0.65 (288.2)**	0.001 (0.013)	-0.145 (11.11)**	-0.182 (3)*	-0.16 (64.5)**	-0.933 (8.45)**	-0.35 (41.71)**
Ti	-0.47 (52.04)**	-0.222 (126)**	-0.05 (16.23)**	-0.104 (22.03)**	-0.007 (0.016)	-0.01 (0.88)	-0.305 (4.31)**	-0.074 (13.4)**
C	-1.399 (240.25)**	-0.366 (36.64)**	-0.478 (102.5)**	-0.211 (35.4)**	-2.721 (12.22)**	-0.6 (207)**	-0.104 (0.09)	-0.27 (11.17)**
Tr	--	--	--	--	-0.001 (1.39)	--	0.298 (0.59)	-0.007 (2.95)*
Wr	0.2 (0.324)	--	--	-0.39 (1.97)	--	--	0.291 (0.039)	0.64 (2.13)
E	0.628 (2.625)	-0.5 (10.82)**	-0.48 (2.44)	--	1.279 (0.62)	0.256 (2.6)	--	2.03 (3.92)**
O	--	--	-0.255 (0.226)	0.5 (2.4)	-2.52 (0.66)	0.04 (0.84)	--	--
<b>Goodness of fit measures</b>								
$LL(\beta)$	-355.44	-1402	-292	-527	-16.5	-1103	-12.8	-104.53
$LL(0)$	-616	-1815	-370	-622	-93.85	-1316	-27.04	-158.76
$\rho^2$	0.424	0.227	0.21	0.153	0.826	0.162	0.527	0.342
% correct samples	47.33	38.15	41.21	28.16	78.69	30.42	63.69	71.68

\*\* More than 95 % significance level; \* More than 90 % significance level

Shaded cells indicate the variables that are in existence in the final model, (R.) presents a reference mode

### 4.3. MPM Results

The same steps are done for MPM, Table 4. Investigating the results of the model for each zone indicates that the accuracy is less than that in the MNL model. Hence, the use of MPM procedure is not accurate and misleading in this case.

## 5. DISCUSSION OF RESULTS

Reviewing of the reached results shows that these are different, depending on the type of model applied on each of the eight studied Egyptian zones. This may be expected since these zones have different demographic, transportation, and finance features. On a comparative basis, the NLM gives the best results for the studied zones, which for simplicity are discussed and explained for each zone respectively as follows:

Zone 1 (Port Said): The census of this zone is 603787 residents [13]. According to the same source, the zone area is 1351 Km<sup>2</sup>. Thus, the residents of this zone estimate only 0.75 % of the total census and 0.14 % of the total area of Egypt. The average personal income in Egyptian pounds (EP) per year in this zone is 26181

[13]. In addition, four modes are used as follows: T, ST, PC, and W. According to results in Table 3, the significant variable is mode cost for paid conveyance (T and ST). Also, trip time is effective for walking choice.

Zone 2 (Ismailia): The census of this zone is 942832 residents [13]. Also, the zone area is 5066 Km<sup>2</sup>. Thus, the residents of this zone estimate only 1.13 % of the total census and 1.71 % of the total area of Egypt. The average personal income in Egyptian pounds (EP) per year in this zone is 26312 [13]. Moreover, five modes are used as follows: T, ST, TT, PC, and W. Table 3 shows that trip cost is significant for paid conveyance (T and TT). Trip distance and time are effective for paid conveyance (TT) and the two non-paid conveyance. The trip purpose other is effective only for mode T. The probability of using T increases with this trip purpose.

Zone 3 (Suez): The census of this zone is 651848 residents [13]. The zone area is 10000 Km<sup>2</sup>. Thus, the residents of this zone estimate only 0.79 % of the total census and 3.38 % of the total area of Egypt. The average personal income in Egyptian pounds (EP) per year in this zone is

Table 3 Results of NLM for the eight zones under research

		Coefficient & $\chi^2$							
Mode	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	Zone 6	Zone 7	Zone 8	
T	Constant	-17.45 (0.59)	-5.9 (31.78)	-4.9 (0.5)	-2.97 (2.3)	-16.6 (0.2)	-2 (12.6)	-1.3 (0.23)	-2.67 (4.6)
	D	-0.2(1.21)	-0.16 (2.09)	-0.14 (0.7)	-2.37 (28.7)	-2.8 (0.1)	-0.14 (9.8)	-3.5 (1.2)	-0.26 (18)
	Ti	-0.43 (0.72)	-0.18 (4.2)	-0.2 (0.9)	-1.56 (36.4)	-0.9 (1.7)	-0.43 (63.8)	-0.27 (5.6)	-0.01 (0.4)
	C	-35.9 (9.45)	-4.7 (118)	-24.6 (166)	-4.82 (34.7)	-25 (5.3)	-3.76 (133)	-2 (10.4)	-2.1 (19.9)
	Tr	1.55 (0.88)	0.113 (0.3)	-0.41 (0.1)	-0.015 (5.9)	0.19 (0.44)	-0.57 (1.03)	-0.3 (0.9)	0.02 (1.6)
	Wr	1.8 (0.4)	--	--	0.1 (0.3)	--	--	-29 (0.8)	-0.62 (0.3)
	E	2.08 (0.3)	0.52 (0.49)	0.12 (0.3)	--	-5.5 (1.1)	-0.5 (0.9)	--	-0.56 (0.8)
	O	3 (0.2)	1.25 (3.12)	0.8 (0.1)	-2.5 (0.88)	1.3 (0.1)	0.3 (0.23)	--	--
ST	Constant	-3.08 (0.07)		-3.6 (0.28)	-0.79 (1.01)	1.31 (0.2)			
	D	-1.24 (0.07)		-0.24 (0.5)	-2.2 (26.1)	-1.1 (0.32)			
	Ti	-0.08 (0.07)		0.04 (0.3)	-1.12 (26)	-0.04 (0.2)			
	C	-26 (6.04)	Control Mode	-23.5 (151)	-2.36 (17.3)	-18.6 (3.6)	Control Mode	Control Mode	Control Mode
	Tr	1.23 (0.3)		0.78 (0.12)	0.002 (1.3)	-0.02 (21)			
	Wr	-0.78 (0.14)		--	-0.14 (0.6)	--			
	E	--		-0.8 (0.2)	--	-2.4 (1.7)			
	O	0.53 (0.03)		-0.5 (0.3)	-0.99 (1.9)	1.4 (0.1)			
Paid Conveyance PB	Constant				3.34 (0.5)	-4.2 (65.9)			
	D				-0.05(0.6)	-0.04 (1.8)			
	Ti				-0.01 (0.1)	-0.01 (0.1)			
	C				-17.9 (3.3)	-0.31 (1.9)			
	Tr				-0.06 (0.3)	-1.78 (7.5)			
	Wr				--	--			
	E				-4.5 (1.12)	-1.26 (8.5)			
	O				1.44 (0.1)	1.8 (7.17)			
TT	Constant		-2.3 (4.43)			-1.7 (7)	6.17 (2.8)		
	D		-4.3 (89.6)			-1.02 (66.3)	-31 (1.03)		
	Ti		-1.65 (68.4)			-0.14 (9.8)	-0.17 (8.3)		
	C		-3.4 (57.1)			-3.5 (107)	-1.33 (7.5)		
	Tr		1.07 (1.16)			0.12 (0.9)	-0.03 (0.6)		
	Wr		--			--	2.2 (0.7)		
	E		0.7 (0.47)			-2.02 (4.7)	--		
	O		0.67 (0.61)			0.65 (0.81)	--		
M	Constant				Control Mode				
	D								
	Ti								
	C								
	Tr								
	Wr								
	E								
	O								
Non-paid Conveyance PC	Constant		3.78 (4.5)			1.98 (3.9)			
	D		-0.64 (4.14)			-0.29 (4.2)			
	Ti		-0.39 (33.7)			-0.14 (1.4)			
	Tr	Control Mode	1.23 (0.83)	Control Mode		0.6 (0.22)			
	Wr		--			--			
	E		-0.7 (0.2)			-0.56 (0.8)			
	O		-1.25 (0.8)			0.9 (0.4)			
	W	Constant	3.19 (1.4)	-2.23 (0.9)	-1.1 (0.9)	1.5 (0.1)		4.25 (10.3)	9.75 (0.23)
D		-3.35 (0.81)	-7.21 (52.6)	-4 (3.86)	-3.8 (3.9)		-1.11 (12.7)	-0.26 (5.8)	-1.46 (2.8)
Ti		-0.29 (12.25)	-2.24 (29.75)	-0.09 (3.4)	-1.33 (3.2)		-0.13 (0.55)	-0.14 (4.1)	-0.29 (1.5)
Tr		1.29 (1.15)	1.63 (0.6)	1.12 (10.8)	0.01 (1.3)		1.53 (1.6)	0.03 (0.2)	0.04 (2.1)
Wr		-0.94 (0.26)	--	--	0.86 (0.2)		--	--	-6.5 (1.5)
E		0.58 (0.1)	-0.15 (0.21)	-0.61 (0.56)	--		-0.2 (0.3)	--	-8.8 (0.5)
O		-9.6 (0.6)	1.78 (0.74)	-2 (0.4)	0.1 (0.3)		0.4 (0.8)	-12.2 (0.8)	--
<b>Goodness of fit measures</b>									
$LL(\beta)$	-23.53	-136.79	-92.91	-125.9	-13.6	-318.4	-22.57	-24.01	
$LL(0)$	-619.3	-1815	-369.63	-497.6	-114.4	-1316	-51.23	-158.76	
$\rho^2$	0.962	0.925	0.749	0.747	0.881	0.758	0.559	0.849	
% correct samples	98.75	98.35	92.48	85.35	96.52	92.31	83.67	94.38	

Table 4 Results of MPM for the eight zones under research

Variables	Coefficient & $\chi^2$							
	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	Zone 6	Zone 7	Zone 8
Constant (T)	-0.508 (3.747)*	-9.52 (67.89)**	0.725 (28.65)**	0.543 (7.5)**	--	0.113 (1.44)	--	-0.523 (1.104)
Constant (ST)	-2.273 (65.07)**	-2.68 (328)**	-1.808 (122.7)**	-0.765 (14.84)**	--	-2.18 (351)**	--	Ref.
Constant (PB)	--	--	--	--	--	-2.57 (427)**	--	--
Constant (TT)	--	0.153 (1.79)	--	--	--	1.17 (141)**	--	--
Constant (M)	--	--	--	R.	--	--	--	--
Constant (PC)	R.	R.	R.	--	--	R.	--	--
Constant (W)	2.476 (84.73)**	0.741 (42.17)**	2.15 (162.2)**	1.474 (51.2)**	--	1.38 (192.4)**	--	0.389 (0.6)
D	-0.616 (125.41)**	-0.269 (337)**	0.002 (0.146)	-0.05 (6.34)**	--	-0.07 (69.5)**	--	--
Ti	-0.259 (76.26)**	-0.093 (96.62)**	-0.022 (10.61)**	-0.045 (21.4)**	--	-0.002 (0.2)	--	-0.08 (29.2)**
C	-1.23 (291.47)**	-0.171 (35.87)**	-0.212 (101.4)**	-0.119 (45.7)**	--	-0.295 (197)**	--	-0.07 (2.83)*
Tr	--	--	--	--	--	--	--	0.004 (3.23)*
Wr	0.084 (0.223)	--	--	-0.234 (2.11)	--	--	--	0.22 (0.97)
E	0.415 (4.366)**	-0.325 (13.08)**	-0.255 (2.36)	--	--	0.118 (1.77)	--	1.07 (5.68)**
O	--	-0.01 (0.01)	-0.114 (0.161)	0.354 (3.37)*	--	0.044 (0.17)	--	--
<b>Goodness of fit measures</b>								
$LL(\beta)$	-380.286	-1462	-308	-539	--	-1131	--	-128.9
$LL(0)$	-836.106	-2057	-477	-709	--	-1640	--	-169.6
$\rho^2$	0.545	0.289	0.353	0.239	--	0.311	--	0.24
% correct samples	41.96	26.61	34.56	20.22	--	35.23	--	53.05

28969 [13]. Also, four modes are used as follows: T, ST, PC, and Walking. Table 3 explains that trip cost is effective for paid conveyance. In addition, trip time, distance and frequency are effective for non-paid conveyance. The probability of walking increases with more trip frequency in the year.

Zone 4 (Great Cairo): The census of this zone is 16,916,979 residents. The zone area is 15034 Km<sup>2</sup>. In addition, the residents of this zone estimate 20 % of the total census and 2 % of the total area of Egypt. The average personal income in Egyptian pounds (EP) per year in this zone is 21733 [13]. There are four modes that used in this zone as T, ST, M, and Walking. Trip distance and time are effective for all modes. Also, the trip cost is significant for paid conveyance. Trip frequency is significant only for T mode. The probability of using taxi decreased with more trip frequency in the year.

Zone 5 (Alexandria): The census of this zone is 5 million residents. The zone area is 2679 Km<sup>2</sup>. The residents of this zone estimate 5.88 % of the total census and 0.71 % of the total area of Egypt. The average personal income in Egyptian pounds (EP) per year in this zone is 25474 [13]. The modes are used in this zone are T, ST, BP, and PC. The results in Table 3 show that trip cost is the only significant for paid conveyance.

Zone 6 (Delta): The census of this zone is nearly 35 million residents. The zone area is 27312 Km<sup>2</sup>. The residents of this zone estimate 42 % of the total census and 3 % of the total area of Egypt. The average personal income in Egyptian pounds (EP) per year in this zone is 16295 [13]. There are six modes that used in this zone as T, ST, PB, TT, PC, and Walking. According to results in Table 3, trip distance is significant for modes T and TT (paid conveyance), and non-paid conveyance. Also, trip time is effective for T and TT. Trip cost is significant for only TT mode. Trip frequency is effective for only PB mode. The probability of PB decreases with more trip frequency per year. In addition, the educational trip purpose is effective for two modes PB and TT. The probability of using the two modes decreases with this trip purpose.

Zone 7 (Upper Egypt): The census of this zone is nearly 24 million residents. The zone area is 107426 Km<sup>2</sup>. The residents of this zone estimate 28.9 % of the total census and 10.7 % of the total area of Egypt. The average personal income in Egyptian pounds (EP) per year in this zone is 12419 [13]. There are four modes that used in this zone as T, ST, TT, and Walking. Trip time is significant for all modes. Also, trip distance is effective for only walking choice. Trip cost is significant with the paid conveyance.

Zone 8 (Sinai): The census of this zone is nearly 461208 residents. The zone area is 56000 Km<sup>2</sup>. The residents of this zone estimate 0.55 % of the total census and 6 % of the total area of Egypt. The average personal income in Egyptian pounds (EP) per year in this zone is 25693 [13]. The modes used in this zone are T, ST, and Walking. The results in Table 3 indicate that trip distance is effective for the three modes. Also, the trip cost is significant for paid conveyance. All the previous results are rational.

Finally, Fig. 4 clarifies the comparison between the three models in terms of  $\rho^2$  and percentage of correct samples. Overview of this figure assures the success of NLM over MNL and MPM for all the studied Egyptian zones.

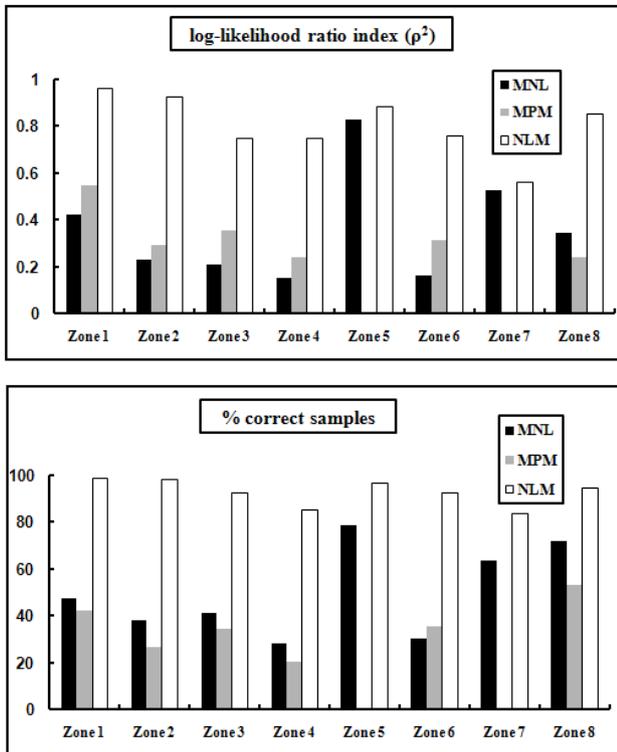


Fig. 4. Comparisons between MNL, MPM, and NLM

## 6. CONCLUSIONS

The present research is applied for implementation of mode choice models in the main zones in Egypt considering the traveller's path from an origin location to the specified taxi terminal with various trip purposes. The substantial conclusions of this research are listed as follows:

. The NLM models gave better confidential results than the MNL and MPM models in terms of the mode type prediction for all the studied zones. Investigation of the parameters of this model for each zone, including; the estimated coefficients for the trip distance, cost, and time variables (level of service attributes), shows expected negative signs and they satisfactorily pass the 0.05 and 0.1 level of significance.

. Generally for all zones, the trip cost is so important factor which strongly affects the choice of paid conveyance. The travellers prefer to select the conveyance with lower cost, especially in work and educational trips. Economic conditions of

the people, especially in developing countries force them for this choice.

. For entertainment trips, the probability of using Taxi highly increases because these trips have a special status as these are usually performed by people having a good economic condition and prefer riding a comfortable transportation means. It is clear that the repetition of these trips during the year does not exceed five or six times, however, increasing the probability of this choice.

. For all trip purposes, the probability of using taxi decreases with more trip frequency in the year. This confirms that the weak economic situation negatively affects the use of the transportation means with expensive cost

. As a rule for all zones, trip time and distance are effective for walking choice. If the time and distance do not exceed 5 minutes and 300 m respectively, the traveller prefers to choose walking. In addition, the probability of walking increases with more trip frequency in the year, which means that the economic conditions highly control this choice.

. For zones that have a relatively wide area such as Delta, Upper Egypt, and Sinai, the trip distance and time is effective. In this case, the travellers prefer riding a shared taxi rather than the taxi because of its lower cost than the taxi.

. In view of the purpose of the educational trip, it is effective for two modes: public bus and three wheels vehicle. The probability of using the two modes decreases with this trip purpose. These two conveyances are the most uncomfortable for travellers, especially for educational trips.

Finally, it is recommended that future papers should be conducted to extend all aspects of the problem in order to reach more extensive modelling and analysis of mode choice. This will surely improve the transportation system in Egypt.

## 7. REFERENCES

- [1] Chen, X.; Liu, X.; and Li, F. "Comparative study on mode split discrete choice models.", *Journal of Modern Transportation*, 21(4), 266-272, 2013. DOI: 10.1007/s40534-013-0028-5
- [2] Guadagni, PM.; and Little, JDC. "A logit model of brand choice calibrated on scanner data.", *Marketing Science*, 2(3), 203-238, 1983. DOI: 10.1287/mksc.1070.0331
- [3] McFadden, D. "Conditional Logit Analysis of Qualitative Choice Behavior.", In: P. Zarembka (ed.) *Frontiers in Econometrics*, New York: Academic Press, 1974.
- [4] Guadagni, PM.; and Little, JDC. "When and what to buy: a nested logit model of coffee purchase.", *Forecasting*, 17(3), 303-326, 1998. DOI:10.1002/(SICI)1099-131X(199806/07)17:3/4<303::AID-FOR698>3.0.CO;2-6
- [5] Dios Ortúzar, J. "On the development of the nested logit model.", *Transportation Research Part B METHODOLOGICAL*, 35(2), 213-216, 2001. DOI: 10.1016/S0191-2615(00)00033-3
- [6] Train, KE. "Discrete choice methods with simulation.", Cambridge: Cambridge University Press, 2003.
- [7] Chib, S; Seetharaman, PB.; and Strijnev, A. "Model of brand choice with a no-purchase option calibrated to scanner-panel data.", *Marketing Research*, 41(2), 184-196, 2004. DOI: 10.1509/jmkr.41.2.184.28674
- [8] Hensher, DA.; Rose, JM.; and Greene, WH. "Applied choice analysis.", Cambridge: Cambridge University Press, 2005.
- [9] Khan, O.; Kruger, J.; and Trivedi, T. "Developing passenger mode choice models for Brisbane to reflect observed travel

- behavior from the South East Queensland Travel Survey.”, Proceedings 30th Australasian Transport Research Forum. Melbourne, Australia, 1-17, 2007.
- [10] Enam, A.; and Choudhury, CF. “Methodological issues in developing mode choice models for Dhaka city.”, Transportation Research Record, 2239, 84-92, 2011. DOI: 10.3141/2239-10
- [11] Yang, F.; Chen, L.; Cheng, Y.; Yao, Z.; and Zhang, X.” Urban Public Transport Choice Behavior Analysis and Service Improvement Policy-making: A Case Study from the Metropolitan City, Chengdu, China.”, Applied Statistics, 42(4), 806-816, 2014. DOI:10.1080/02664763.2014.986438
- [12] Semeida, AM. “Derivation of Travel Demand Forecasting Models for Low Population Areas: the Case of Port Said Governorate, North East Egypt.”, Journal of Traffic and Transportation Engineering (English Edition), 1(3), 196-208, 2014. DOI:10.1016/S2095-7564(15)30103-3
- [13] ECAPMS. “Egyptian Central Agency for Public Mobilization and Statics”, Cairo, Egypt, 2010. <http://www.capmas.gov.eg/database.aspx>
- [14] Domencich, T.; and McFadden, D. “Urban Travel Demand: A Behavioural Analysis.”, North-Holland, Amsterdam: Berkeley, 1975.
- [15] Anwar, AM. “Determinants of Modal Choice Behavior Using Nested Logit Model: A Journey to Work Trip in Dhaka City”, S. Jahan and A. Kalam, ed. Dhaka Metropolitan Area and its Planning: Problems, Issues and Policies, 2012.(URL: [http://www.bip.org.bd/SharingFiles/journal\\_book/20140427154119.pdf](http://www.bip.org.bd/SharingFiles/journal_book/20140427154119.pdf))
- [16] Hausman, J.; and McFadden, D. “Specification tests for the multinomial logit model.”. Econometrica, 52, 1219-1240, 1984. DOI: 10.2307/1910997
- [17] Williams, HCWL. “On the formation of travel demand models and economic evaluation measures of user benefit.” Environment and Planning, 9(A), 285-344, 1977. DOI: 10.1068/a090285
- [18] Daganzo, CF. “Multinomial probit: the theory and its application to demand forecasting.”, New York: Academic Press, 1979.
- [19] XLSTAT 4.06 Inc., Version 2014.4.06 copyright Addinsoft 1995-2014 for Windows.
- [20] Semeida, AM. “Analysis and evaluation of road safety in Egypt using conventional and non-conventional modeling techniques.”, PhD dissertation, Port-Said University, Port-Said, Egypt, 2011.
- [21] Agresti, A. “An Introduction to Categorical Data Analysis”, (2nd Edition). Hoboken, New Jersey: Wiley, 2007.

### ملخص البحث باللغة العربية

#### تطبيق نماذج الاختيار المنفصل لاستكشاف سلوك المسافرين في شبكة النقل المصرية

تشكل نماذج اختيار وسائل النقل بوجه عام جزءاً حاسماً في تحليل الطلب على السفر في أي منطقة دراسة في أي دولة وشكلت دوراً هاماً في نمذجة النقل على مدى العقود الماضية. وتوفر هذه النماذج تمثيلاً مفصلاً للجوانب المعقدة في الطلب على النقل على أساس مبررات نظرية قوية. البحث الحالي يقوم بوصف وتطوير نماذج اختيار وسائل النقل و ذلك لمناطق رئيسية في مصر بالاعتماد فقط على سلوك المسافرين. و قد تم تجميع البيانات اللازمة باستخدام استبيان خاص أعدته وزارة النقل المصرية في عام 2010 يتضمن مقابلة المسافرين في محطات سيارات الأجرة الأساسية في جميع محافظات مصر. يتم اشتقاق نموذج اختيار وسائل النقل بتطبيق ثلاثة نماذج هي: the multinomial probit (MPM), the nested logit (NLM), the multinomial logit (MNL). تبين نتائج تنفيذ النمذجة باستخدام النماذج السابقة أن نموذج NLM لجميع المناطق المصرية تحت الدراسة يعطي نتائج مرضية وأفضل من النموذجين الآخرين. تؤثر تكلفة الرحلة بشدة على اختيار المسافر لوسيلة النقل المدفوعة الأجر، وخاصة عندما يتعلق الأمر بالعمل والرحلات التعليمية. قد يكون هذا واضحاً في معظم البلدان النامية، حيث يؤثر الوضع الاقتصادي الضعيف سلباً على استخدام وسائل النقل المكلفة. لذا، فإن النتائج العامة لهذه الدراسة قيمة وثرية، مما يشير إلى سياسة النقل المعقولة الموجهة للمسافرين على الطرق المصرية، والغرض الرئيسي منها هو استخدام المركبات المريحة نوعاً ما بأقل تكلفة للمسافر. من المفضل التحقيق في جميع جوانب المشكلة من أجل التوصل إلى نمذجة شاملة لتحليل نماذج اختيار وسائل النقل مما قد يساعد على تحسين نظام النقل السائد في مصر.