Multinomial Logit Utility Model for Tanta City Transportation System considering Ridesharing Mode

Ahmed Mohamed Elkafoury, akram elsaied Mostafa, Hafez Abbas Afify, Islam Mahmoud aboelnaga

Follow this and additional works at: https://digitalcommons.aaru.edu.jo/erjeng

med Elkafoury, akram elsaied Mostafa, Hafez Abbas Afify, Islam Mahmoud aboelnaga: Multinomial Logit Utility Model for Tanta City Transportati

Vol. 5, No. 2 – 2021   Journal of Engineering Research (ERJ)

Multinomial Logit Utility Model for Tanta City Transportation System considering Ridesharing Mode

Ahmed Elkafoury, Akram Elsaied,*, Hafez Afify and Islam Abo Elnaga
Public Works Engineering Department, Faculty of Engineering, Tanta University, Tanta City 31521, Egypt

*Corresponding Author, e-*mail:* akram.elsaied@f-eng.tanta.edu.eg

Abstract- **Recently, ridesharing has been noticed as an effective and sustainable mode of transport. In which, each passenger carries out a journey that benefits travelers and society greatly, such as reducing travel costs, reducing journey times, relieving road traffic congestions, preserving fuel and reducing air pollution. While the importance and efficiency of ridesharing, ridesharing among travelers has not been commonly utilized. This paper introduces an upgraded disaggregated utility model for a mode choice step in the four-step traffic demand model for transport of passengers in Tanta city - Gharbiya governorate. For the first time, the analysis in this paper includes ridesharing mode in addition to other three common modes; private vehicle, bus, and taxi . The aim of this research is to involve the ridesharing mode and new parameters in the mode choice analysis of the case study. This is done by including additional parameters such as travel cost and comfortable in addition to travel time. Based on surveys, the developed Multinomial logit utility models have been assessed using the McFadden pseudo R^2 values. The values demonstrated high compatibility of results with real data as pseudo R2 values ranges between 0.2 and 0.4. In addition, all utility models developed for modes are found to have P-values less than 0.05 indicating the significance of the considered utility characteristics.**

Keywords: **Ridesharing, Stated Preference, Multinomial logit model, Utility function, and McFadden pseudo**

I.   INTRODUCTION

In an era of high-technology development, energy and resources have been enormously consumed to facilitate human life. The products of such consumption have generated substantial pollution including greenhouse gas emissions contributing to global warming. Therefore, in recent decades, the sustainable development has been emphasized in every social segment. One of the basic concepts of sustainable development is to minimize the resources consumption by increasing the resources utilization. In transport segment, mass transit, a shared transport service that is available for general public used, is well-known environmentally friendly transport that increases the utilization of vehicle and reduces the traffic congestion. However, mass transit may not be able to satisfy all travel demands due to the inflexibility (i.e., fixed routes and schedules), difficult accessibility (e.g., disabled and aged people may hardly access the mass transit), inconvenience, and discomfort [1], [2]. These cause the necessity of private vehicle and/or for-hire vehicle to travel door-to-door. Traveling by using private vehicle or for-hire vehicle are certainly not an efficient travel mode comparing to mass transit. One of the key concepts for increasing the utilization and efficiency of such personal trip is sharing economy (also known as collaborative consumption). Sharing economy is a concept where people

share their personal assets and/or services to others[3]. Under the sharing economy concept, travelers can increase the utilization and efficiency of their personal trips and/or vehicle by sharing a ride and/or a vehicle to others.

Travelers sharing a ride of their personal trip to other traveler(s) who has similar itinerary is so-called ridesharing. Ridesharing has become an alternative to travel. Ridesharing compromises the advantages between mass transit and private transport. It is estimated to increase the utilization of vehicle used for private trip and reduce traffic congestion thus sacrificing appropriate levels of comfort and convenience for travelers as opposed to mass transit [4]. For some travelers, ridesharing could generate more social interaction such as travelers who share common interest. Moreover, one of the strong advantages of ridesharing transportation in travelers point of view is the reduction of travel-related expenses[5].

In the early era of ridesharing transportation, riders can find a ride by waiting at the designated point, and the driver who desires to share a ride will stop by and pick up the rider(s); this form of ridesharing is so-called slugging (also known as casual carpooling)[6]. To facilitate ridesharing transportation, the ridesharing system that provides a ride matching service for travelers to find the ridesharing partner(s) has been developed. By using ridesharing system, travelers can find ridesharing partner(s) and plan their ridesharing trip in advance. Such ride-matching can be formed in several ways such as assigning by ridesharing system, forming by travelers who have similar itinerary.

In recent decades, the development of information and communication technology (ICT) has enabled the real-time ridesharing system, also known as dynamic ridesharing system (DRS). DRS refers to a ridesharing system that enables travelers to find ridesharing partner(s) on very short notice or even en-route [7]. DRS can be developed on, for example, a smartphone application. Therefore, ridesharing transportation has become a demand responsive transport and more convenient for travelers to perform. DRS has also been developed for a commercial use such as Lfyt, and Uber. Even though ridesharing transportation has become well known to travelers, however it cannot be denied that not all travelers are willing to rideshare due to its existing disadvantages such as longer travel time caused by detour, discomfort of sharing private space[8].

This paper aims at developing an upgraded function for disaggregated utility that is utilized for the mode choice analysis in the four-step traffic demand model [2], [9] and [10]. The case study for developing the model is the transport of passengers in **Tanta city - Gharbiya governorate**. This model involves the three modes of transportation in the city in addition to the

65

ridesharing as a proposed new sustainable mode; private vehicle, public transport (Public bus and microbus), and taxi, As number of Public bus in Tanta city about 80 bus, so we consider public bus and microbus as one mode called bus.

The developed model includes not only the common utility parameters (travel time), but also additional parameters such as travel cost and comfortable conditions that influence mode choice.

II. METHODOLOGY

Figure 1 encloses the methodology of proposed model to concluding utility function, which depending on collecting data using stated preference technique then using Multinomial logit (MNL) model by using maximum likelihood estimation utilizing python BIOGEME software as the tool to find parameters of utility function.

The methodology of this research is divided into two main phases as follows:-
- **Phase one** is used for the collection data depending on stated preference technique and number of people will contribute in the survey (sampling).
- **Phase two** is used for conducting and derive the prosed model using MNL model by using maximum likelihood estimation method.
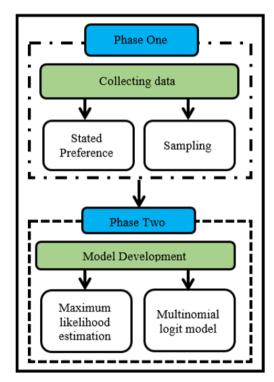

Figure 1. Methodology utilized for concluding utility function

III. STATED PREFERENCE

For the construction of utility model, thus, SP (stated preference) data is a need. However, there are various features of SP experiments, which may influence the resulting temporal meaning and the changes. The calculations are considered to be particularly sensible to the design of the SP experiment [11]. For the construction of utility model, different types of SP data is used; 1) the amount of alternatives in a package of options, 2) the sum of set of choices to be weighed (treatments), and 3) the

range and level of trading attributes. Although the SP data can be obtained by evaluating, ranking and selecting, the mentioned SC (Stated Choice) experiments provide a basis for the investigation of the relative marginal disutility of attribute variance and its possible correlations. [12]. Due to ease of calculation, SP- and RP (Revealed Preference) data are usually evaluated with standard MNL models. The MNL models, however, place such constraints as freedom from insignificant alternatives (IIA).

Valuing the attributes is likely to be underestimated in the MNL model as an aspect of unconsidered pressures on travel choices is 'forced' into the parameter estimates of the observed results if the strict, separately and identically distributed (IID). The hypothesis suggests that the time effect is more expensive since all of the unexpected features have more to do with the time than the cost of transport [13]. From an econometric standpoint, the mean of a random parameter can be greater than that of an MNL since the random logit model breaks down the unattended utility variable and normalises the parameter (via the scale parameter) depending on a portion of the unattended factor.

IV. SAMPLING

Sampling remains one of the least known fields of study in conjunction with experimental design. While it is generally easy to choose the population of samples, the purpose of a study is often much less obvious who should sample and what approach can be used to sample from the selected population. The analysis must start by identifying the sample framework for the analysis to explain the sample population. simple random samples (SRS), exogenous random stratified samples are potential sampling strategies (ESRS)[14]. Furthermore, decreases in the number of sampled respondents will result in a resulting decline in the heterogeneity of socio-demographic features and qualitative effects seen within the survey as model estimates are made, which will certainly cause issues when those effects are used in the model estimates. The optimal degree of precision of the predicted probabilities for simple random samples determines the minimum appropriate sample size N, Let p be the true choice proportion of the relevant population, a be the permissible margin of error expressed as a percentage between $\hat{p}$ and p, and γ be the confidence level for simply estimation such that Pr $(|\hat{p} - p| \leq \text{ap}) \geq \gamma$ for a given N. The minimum sample size is defined as:

$$N \geq \frac{q}{Spa^2} \left[\varphi^{-1} \left(1 - \frac{\alpha}{2} \right) \right]^2 \quad \text{.................................. (1)}$$

Where, q=1-p, p defined as true choice proportion of the relevant population, a is the allowable error, S is the number of choice task that respondent face, $\left[\varphi^{-1} \left(1 - \frac{\alpha}{2} \right) \right]^2$ is the inverse cumulative distribution function of a standard normal (Z^2) as shown in equation (1). So based on the data before the number of sample would be 600 if p=0.4 which represent the true choice proportion of relevant population of Tanta city and number of choice which each respondent faced (i.e, s = 4) at 5% confidence level

V. MAXIMUM LIKELIHOOD ESTIMATION

The most common method used to measure estimators in simple and nesting logit models is the maximum probability method. Simply stated that "The maximum probability

66

med Elkafoury, akram elsaied Mostafa, Hafez Abbas Afify, Islam Mahmoud aboelnaga: Multinomial Logit Utility Model for Tanta City Transportatic

Vol. 5, No. 2 – 2021   Journal of Engineering Research (ERJ)

estimators are the parameter values that have been observed for the sample."[15].

The maximum likelihood probability estimation procedure involves two important steps [16], [17]:

1. Creating a likelihood function, which is a joint probability density function of the observed sample., and
2. Choosing parameter values that maximize the likelihood function's likelihood.

The likelihood function can be defined for a sample which can be written as:

Let 'i' for individuals, each with 'M' alternatives so the equation for likelihood function is:

$$L\ (\boldsymbol{\beta}) = \prod_{\forall i \in I} \prod_{\forall m \in M} (P_{mi}(\boldsymbol{\beta}))^{\delta_{mi}} \quad \ldots\ldots\ldots\ldots\ldots\ldots(2)$$

Where L is the probability that individual i chooses alternative M, Pmi is the probability that individual i chooses alternative M, and j is the chosen indicator (=1 if j is chosen by individual i and 0 otherwise).

Finding the first derivative of the likelihood function and equating it to zero yields the values of the parameters that maximize the likelihood function. The most common method is to maximize the logarithm of L instead of L itself. Because the logarithmic function is absolutely monotonically growing, it has no effect on the parameter estimations. As a result, the likelihood function is transformed into a log-likelihood function, which is denoted by,

$$LL(\boldsymbol{\beta}) = \log(L\ (\boldsymbol{\beta}) = \sum_{\forall i \in I} \sum_{\forall m \in M} \delta_{mi} \times \ln(P_{mi}\ (\boldsymbol{\beta}))\ldots.(3)$$

The first derivative of the likelihood function's logarithm can be represented as follows,

$$\frac{\partial(LL)}{\partial \beta_k} = \sum_{\forall i \in I} \sum_{\forall m \in M} \delta_{mi} \times \frac{1}{P_{mi}} \times \frac{\partial P_{mi}(\beta)}{\partial \beta} \forall k \ldots\ldots\ldots\ldots..(4)$$

When the second derivative is negative definite, the maximum likelihood is obtained by setting the above equation to zero and solving for the best values of the parameter vector to ensure that this is the solution for a maximum value.

Most existing estimation computer programmers estimate the coefficients that best explain the observed choices in terms of making them most likely to have occurred, given the mode choice data. Standard commercial packages like ALOGIT and BIOGEME are commonly used to estimate logit models, owing to their ability to handle complex nested logit structures, both linear and non-linear.

VI. DISCRETE CHOICE MODELS

As seen in the proportion of a partial derivative equation (4), valuing one attribute in comparison to the attribute and travel costs is relatively simple in discrete models. It is important to recognize that the reasoning for this strategy is founded on a large body of microeconomic theory that addresses how individuals allocate time and how it differs through alternatives. The utility model must be established by gathering tour operator preferences in the form of exposed preferences (RP) or specified preference data (SP). In a variety of areas, RP and/or SP data have been used to approximate readiness-to-pay interventions [18]. RP data is normally unacceptable for a richer breach of travel time. It is better defined as "dirty" in the statistical evaluation of individual choice preferences. Due to industry, infrastructure, and sampling considerations, such attribute levels cannot be found in RP (Revealed Preference) and the predictor variables (alternative attributes and effects)

can have high or extreme multi-collinearity levels [11]. On the other hand, the SP data are rich in estimating marginal willing to pay (WTP) values and are efficient [19].

In equation (5) the utility of each variable in econometric models based on random utility theory consists of an observed (deterministic) component denoted by V and a random (disturbance) component denoted by ε,

$$U\ =\ V\ +\ \boldsymbol{\varepsilon} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ (5)$$

The definite component *V* relies once again on the characteristics (*Z*) of the individual's option, the socioeconomic attributable observed (S) and the coefficient vector (β), and then on the deterministic part *V*, so the component *V* can written as:

$$V = V\ (\boldsymbol{Z}, \boldsymbol{S}, \boldsymbol{\beta})\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ (6)$$

Because of the inclusion of a random component, a probabilistic assumption can be made that when an independent 'n' faces a choice set, Cn, consisting of Jn choices, the choice probability of alternative I is equal to the probability that the utility of alternative Uin, is greater than or equal to the utilities of all other alternatives in the choice set, as shown in equation (7) and equation (8)

$$Pn\ (i)\ =\ \Pr(Uin\ \geq Ujn\ , \text{for all } j\ \in Cn\ ) \ldots\ldots\ldots\ldots..(7)$$

$$Pn\ (i)\ =\ \Pr(Vin\ +\ \varepsilon in\ \geq\ Vjn\ +\ \varepsilon jn\ , \text{for all } j\ \in$$
$$Cn\ , \forall j\ \neq\ i) \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(8)$$

Ben Akiva assumed [15] IID (Gumbel distribution) for, the MNL model will predict the likelihood that a person will select i as shown in the following equation.

$$Pin = \frac{e^{v_{in}}}{\sum_{j \in C_n} e^{v_{in}}} \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..\ldots\ldots...\ (9)$$

This model is useful for modelling choice behavior and can be estimated using the maximum likelihood technique. This model, however, has several limitations. The most severe of these is the IIA property, which states that changing one alternative's attributes changes the probabilities of the other alternatives proportionally. This substitution pattern may not be applicable in all circumstances. Second, in a choice experiment, the coefficients of all attributes are assumed to be the same for all respondents, whereas in reality, there may be significant variability in how different individuals respond to attributes.
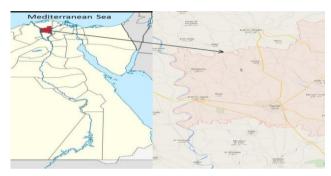
VII. CASE STUDY

Gharbia Governorate is divided into 8 administrative centers (Tanta- Almahalla Alkobra - Kafr El Zayat - Basyoun - Qutor – Elsanta – Samanod - Zefta). Among which, Tanta is considered the biggest administrative center. Tanta city is the capital of Gharbia Governorate and locates in the middle of Delta Nile. Tanta is 94 km north of Cairo and 130 km southeast of Alexandria. Its Geographic coordinates are (30° 47' 28N, 30° 59' 53E). It is the cotton-ginning center and the main railroad hub of the Nile Delta. The biggest commercial city in the delta. The biggest university in the delta locates in Tanta. Most and Biggest common roads in Tanta are Al-Galaa Street, Al- Bahr (algeish) street, Al-Nahaas Street and Saeed Street.

For the purpose of this paper, Tanta city was administratively divided in the research methodology into 14 Transportation Analysis Zones (TAZs) as shown as in figure 2 and codes of every zone shown in table 1 [20]

Table 1. Coding system for Tanta city administrative sections

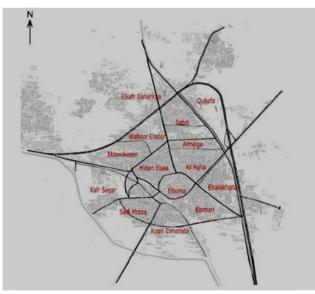| Zone code	TAZ
1	Quhafa
2	Waboor Elnoor
3	Ali Agha
4	Almalga
5	Midan Elsaa
6	Eldawaween
7	Elborsa
8	Elkafr Elsharkya
9	Sabri
10	Elsalakhana
11	Elemari
12	Segar
13	Elmahata
14	Sedi Marzoq

Figure 2. Location and Administrative sections of Tanta city

VIII. DATA COLLECTION AND QUESTINNAIRE DIESIGN

To formulate rational and real improvement strategy for transportation system in Tanta city, it is essential to understand how users value different attributes of travel and reflect real situations and attitudes of Tanta people towards transportation system. In this concern, Survey types were created to gather information about trip characteristics, respondents' socioeconomic characteristics, and reported preference "option" from an option collection. The aim of a survey questionnaire is to collect the respondent's specified preferences "option." To gather data from respondents, face-to-face interviews with paper and pencil were performed. The respondent was simply asked which mode would you choose for travel as shown in figure 3. The respondent has right to choose just one of four choices. Three attributes were considered for the design of choice sets.
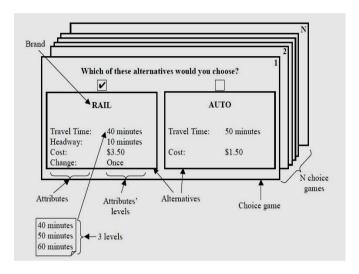

Figure 3. Example of choice-based questionnaire

The development of alternatives through complete or fractional factor design is standard practise. However, using both of these combinations in SP experiments is neither necessary nor realistic.[21].

The statistical design to create conceptual alternatives and games offered to respondents is the central part of the defined choice technique. An experimental design is typically 'orthogonal,' meaning that the characteristics given to respondents vary independently of one another. Consequently, the influence of each attribute level on responses can be more effectively isolated. This avoids the problem of 'multi-collinearity' between attributes, which is typical with exposed preference data [22].

Despite the mathematical benefits of complete factorial designs, they are only practical for minor problems with a small number of characteristics, stages, or both. It is clear that if we have three levels and four alternative we will have 81 ($3^4 = 81$) is the solution to a very small problem involving four attributes of three levels each.

As a result, it is important to limit the number of possible variations (number of scenarios). One solution is fractional factorial architecture, which has been listed in several publications [22] as the most widely used solution to reduce the number of scenarios. The concept of fractional factorial architecture stems from the study of interactions. Not only are the relationships between key effects orthogonal in the complete factorial design, but they are also orthogonal between interactions. On the other hand, except for key results, certain associations are neglected in the fractional factorial architecture.

68

med Elkafoury, akram elsaied Mostafa, Hafez Abbas Afify, Islam Mahmoud aboelnaga: Multinomial Logit Utility Model for Tanta City Transportatic

Vol. 5, No. 2 – 2021   Journal of Engineering Research (ERJ)

As a result, the fractional factorial orthogonal technique is used to delete some of the option sets while maintaining much of the statistical properties of the complete factorial design. Fractional and orthogonal key effects are designed and successfully reduced by removing such higher-order combinations[23]. Fractional factorial orthogonal design is used to produce and generate the alternatives.

Therefore, the planning of option sets considers characteristics such as inconvenience (comfortable level), travel time and travel expenses. Each trait has two levels of description. Levels are determined after consultations with consultants and trip participants. The qualitative attribute of discomfort is described and coded (see Table 2). Table 3 shows the characteristics and levels used in the analysis.

Table 2. Representation of discomfort attribute

Public Bus	
Condition of travel	Dummy value
Standing	0
Seating	1
Private car, Taxi and Ridesharing	
Condition of travel	Dummy value
No AC (Air Conditioner)	0
AC (Air Conditioner)	1

Table 3. Attributes and Levels of attributes

Attributes	Level 1	Level 2
Travel Time	Actual	+50%
Travel Cost	Actual	+50%

By using only fractional factorial orthogonal key effects, it is assumed that all interaction effects are marginal. [21], [22], [24] to reduce respondents' confusion and exhaustion. The following table from (engineering statistics handbook) show how we can get fractional factorial factors (see table 4). Building full factorial design required generation of eight situation (two level with three alternative), so based on fraction factorial theory only four situation choice are introduced per each respondent to reduce confusion and fatigue and make respondent choose in good way.

Table 4. Illustration of fractional factorial design

Number of Factors,k	Design specification	Number of Runs N
3	2^{3-1}	4
4	2^{4-1}	8
5	2^{5-1}	16
5	2^{5-2}	8
6	2^{6-1}	32
6	2^{6-2}	16

Paper and pencil interview method survey is used to present the four situation as shown in table 5. In this survey, the respondent is asked to choose only on choice from (A, B, C or D), where A represent Bus, B represent Taxi, C represent Private car and D represent ridesharing mode.

Collecting data are coded in dummy variable and fed into Biogeme software in order to covert descriptive data into numerical data and estimation variables of utility function as show in table 6.

The database includes information on the path, the journey, the socioeconomic characteristics of the respondent and ultimately the preferences of the respondent in the form of "choices." Road features include route length, number of bus stations, fare structure and timetable. Travel features include origin, destination, intent, travel time and fare charged. Age, gender, education, profession and income are the social and economic characteristics of the respondent. Preference data are obtained by choices in which respondents select an alternative from the three options.

IX. DESCRIPTIVE STATISTICS OF COLLECTED DATABASE

Summary statistics of the information about trip purposes and socioeconomic details such as gender, age, professional situation, working hours, and education level and car ownership forming the database for the present work are given in Table 7 and figure 4. The sample consists of about 68% of male respondents. About 54% of the sample is in the age group more than 20 years, 46% less than 20 years. About 35% of the respondents are still working, 22% unemployed, 15% retired and 28% are students. We also survey about working hours per week, which divided into 78% working less than 35 hours per week. About 26% high education level and 61% with medium education level and 13% with low education level. Regarding the private transport ownership, only 25% of the respondents has car and 29% of the surveyed population has bicycle.

Table 5. The attributes and corresponding levels for each alternative

	Bus				Taxi				Private Car				RideSharing				Choice
	Travel time (Min.)	Travel Cost (L.E.)	Comfort		Travel time (Min.)	Travel Cost (L.E.)	Comfort		Travel time (Min.)	Travel Cost (L.E.)	Comfort		Travel time (Min.)	Travel Cost (L.E.)	Comfort		A or B or C or
A	15	2	Seating	B	10	15	AC	C	7	10	AC	D	10	5	AC		
A	22.5	2	Standing	B	15	15	No AC	C	10.5	10	No AC	D	15	5	No AC		
A	15	3	Standing	B	10	22.5	No AC	C	7	15	No AC	D	10	7.5	No AC		
A	22.5	3	Seating	B	15	22.5	AC	C	10.5	15	AC	D	15	7.5	AC		

69

Table 6. Paper and Pencil questionnaire

| | Bus | | | | Taxi | | | | Private Car | | | | RideSharing | | | | Choice |
	Travel time (Min.)	Travel Cost (L.E.)	Comfort		Travel time (Min.)	Travel Cost (L.E.)	Comfort		Travel time (Min.)	Travel Cost (L.E.)	Comfort		Travel time (Min.)	Travel Cost (L.E.)	Comfort	A or B or C or D
A	1	1	1	B	1	1	1	C	1	1	1	D	1	1	1	
A	0	1	0	B	0	1	0	C	0	1	0	D	0	1	0	
A	0	0	0	B	0	0	0	C	0	0	0	D	0	0	0	
A	1	0	1	B	1	0	1	C	1	0	1	D	1	0	1	

Table 7. Characteristics of the individuals in the dataset

		Frequency	Percentage (%)
Gender	Male	408	68
	Female	192	32
Age	Age < 20	276	46
	Age >20	324	54
Professional situation	Employed	210	35
	Unemployed	132	22
	Retired	90	15
	Student	168	28
Working hours	Working hrs.<35 hr. per week	468	78
	Working hrs.> 35 hr. per week	132	22
Education level	Low	72	12
	Medium	366	61
	High	156	26
Ownership	Car	150	25
	Bicycle	174	29
	N/A	276	46

X. ESTIMATION OF UTILITY FUNCTION PARAMETERS

During the survey, 600 samples of preference data from the research region were obtained in order to construct the utility equation. The SP selection data was coded and fed into the BIOGEME software for estimation using MNL model which estimating maximum likelihood of parametric models in general, with a focus on discrete choice models. Pandas, a Python Data Analysis Library package, is required. On the other hand ALOGIT is a powerful piece of software for power users. It handles large models easily and quickly, with virtually no limit on problem size, and provides a set of features and analysis capabilities that leading-edge modelers have found useful for more than 30 years. For simplicity and free fees, BIOGEME is used to estimate our model.

XI. RESULTS AND DISCUSSION

The results of this research work is the utility model that considers the ridesharing mode and attributes in Tanta City. Overall, the utility function of any mode were considered to have the form:

$$U_i = \text{ASC\_MODE}_i + \sum_{j=1}^{j} \beta_j * Z \ldots\ldots\ldots\ldots\ldots\ldots\ldots(10)$$

Where

Z: is the value of the utility characteristic (factor)

β: is the estimated parameter for the utility characteristic j

med Elkafoury, akram elsaied Mostafa, Hafez Abbas Afify, Islam Mahmoud aboelnaga: Multinomial Logit Utility Model for Tanta City Transportatio

Vol. 5, No. 2 – 2021   Journal of Engineering Research (ERJ)

GENDER

AGE

PROFESSIONAL SITUATION

WORKING HOURS

EDUCATION LEVEL

OWNERSHIP

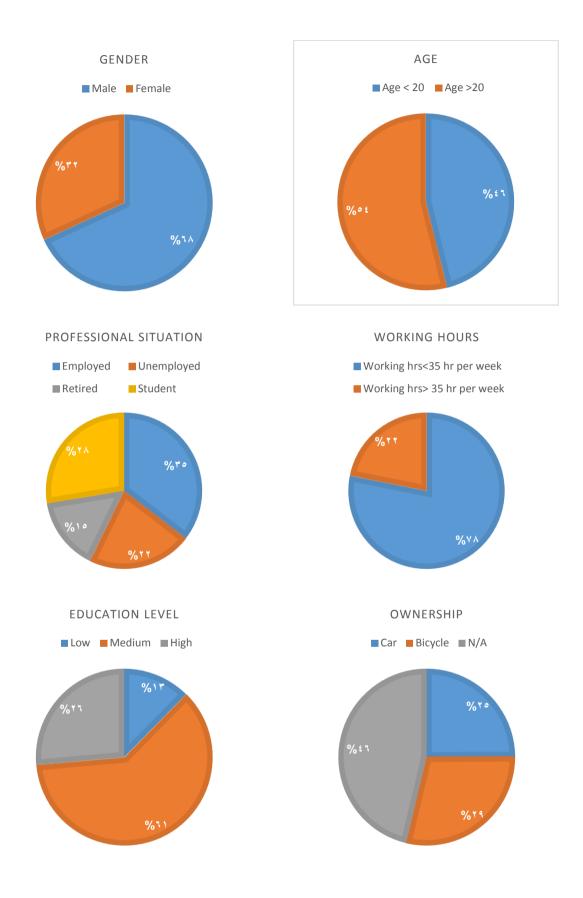**Figure 4. Charts of characteristics of the individuals in the dataset.**

71

As the main objective of the paper, and according to the signs of the coefficients, statistical importance of the coefficients, and predictability of the model, the following utility model was chosen for ridesharing mode and can be written as follows:

$$U = -2.47 - 0.932 \times \text{Travel}_{\text{Cost}} - 0.657 \times \text{Travel}_{\text{Time}} + 2.541 \times \text{Comfort} \quad (12)$$

According to Koppelman and Bhat [25] there exist three methods to review the estimation results for the multinomial logit model:

1) Formal test (statistical tests),

2) Overall goodness of fit (rho square value),

3) Informal test (signs of estimated parameters). Joel Horowitz [26] concluded that the formal test (statistical tests and goodness of fit) is the most powerful judgment on logit model, whereas the test based on extrapolation (informal test) is the weakest.

Table 8 encloses the estimated parameters of utility functions for each mode resulting from the MNL modeling performed by BIOGEME software. First, the signs of the factor of the considered parameter are found to follow the engineering expectations and sense and in accordance with the actual scenario of the study. It is found that as the travel monetary cost increases the utility of the mode consequently decreases. In addition, as the comfort level of the mode increases the utility increases in response.

On the other hand, sign and value of the coefficients of travel time and comfort reflect the real choices and socioeconomic characteristics of public transport users in Tanta city. This is due to predominantly low-income portion of population with negligible car ownerships. This portion of people cares for travel monetary cost regardless of the travel time.

Table 8. The MNL model estimation results.

Name	Value	P-value
ASC_BUS	-1.72	0.0041
ASC_CAR	-2.13	0.0016
ASC_TAXI	0.86	0.0009
ASC_RIDESHARING	-2.47	0.0057
β_Cost	-0.932	0.0014
β_Time	-0.657	0.031
β_Comfort	2.541	0.0011
Number of samples = 600		
Initial log likelihood: -1011.490		
Final log likelihood: -2690.661		
R-square for the model: 0.376		

In addition, and to assess the developed models, P-value is considered as an important indicator of trust. Thus, it is important to measure this value for different involved

parameters. As shown in Table 8, the p-value indicates that the parameters involved in the model present good representativeness of the real case. This is because all P-values are less than 0.05 [21].

More, the developed model was assessed using the McFadden Pseudo (R-squared) value. This value is a probability ratio that forms a statistic value which is commonly used to assess how well models complement the results (i.e., how well the model performs compared with a model in which all the parameters are zero). This value can be estimated as follows:

$$R^2 = 1 - \frac{\text{Final log likelihood}}{\text{Initial log likelihood}} \quad (11)$$

The model is considered to have a good fit if it has a pseudo R^2 value that ranges between 0.2 and 0.4 [24].

XII. CONCLUSION

As a preference model, the paper presents the estimation of ridesharing utility function parameters for implementation in Tanta district. The Maximum Likelihood approach was used to estimate ridesharing mode option model parameters, which allows for comparison of various types of utility functions for other modes of transportation based on final log-likelihood. The McFadden pseudo R^2 value between 0.2 and 0.4 shows that the model fit is optimal for determining the best function. In addition, all utility models developed for modes are found to have P-values less than 0.05 indicating the significance of the considered utility characteristics.

XIII. FUTURE WORK

In the future work, Policy and sensitivity analysis based on the developed model is needed for forecasting and policy analysis in case of considering the ridesharing mode.

XIV. REFERENCES

[1] Sun, Y., Chen, Z., & Zhang, L. (2020). Nonprofit peer-to-peer ridesharing optimization. Transportation Research Part E: Logistics And Transportation Review, 142,102053.https://doi.org/10.1016/j.tre.2020.102053

[2] Chiu Chuen, O., Karim, M., & Yusoff, S. (2014). Mode Choice between Private and Public Transport in Klang Valley, Malaysia. The Scientific World Journal, 2014,1-14.https://doi.org/10.1155/2014/394587

[3] Furuhata, M., Dessouky, M., Ordóñez, F., Brunet, M., Wang, X., & Koenig, S. (2013). Ridesharing: The state-of-the-art and future directions. Transportation ResearchPartB:Methodological,57,2846.https://doi.org/10.1016/j.trb.2013.08.012

[4] Erdoğan, S., Cirillo, C., & Tremblay, J. (2014). Ridesharing as a Green Commute Alternative: A Campus Case Study. International Journal Of SustainableTransportation, 9(5),377388.https://doi.org/10.1080/15568318.2013.800619

[5] Santi, P., Resta, G., Szell, M., Sobolevsky, S., Strogatz, S., & Ratti, C. (2014). Quantifying the benefits of vehicle pooling with shareability networks. Proceedings Of The National Academy Of Sciences, 111(37),1329013294.https://doi.org/10.1073/pnas.1403657111

[6] Oliphant, Marc. (2008) "The Native Slugs of Northern Virginia." Master Science. Urban Reg. Plan. Va. Tech Blacksbg.

[7] Agatz, N., Erera, A., Savelsbergh, M., & Wang, X. (2010). Sustainable passenger transportation: Dynamic ride-sharing. Erasmus

72

med Elkafoury, akram elsaied Mostafa, Hafez Abbas Afify, Islam Mahmoud aboelnaga: Multinomial Logit Utility Model for Tanta City Transportatic

Vol. 5, No. 2 – 2021   Journal of Engineering Research (ERJ)

Research Institute of Management - ERIM

[8]   Lin, Y., Li, W., Qiu, F., & Xu, H. (2012). Research on Optimization of Vehicle Routing Problem for Ride-sharing Taxi. Procedia - Social And Behavioral Sciences, 502.doi:10.1016/j.sbspro.2012.04.122

[9]   Wei, B., Saberi, M., Zhang, F., Liu, W., & Waller, S. (2020). Modeling and managing ridesharing in a multi-modal network with an aggregate traffic representation: A doubly dynamical approach. Transportation Research Part C: Emerging Technologies, 117, 102670. doi: 10.1016/j.trc.2020.102670

[10]   Sanchez-Anguix, V., Chao, K., Novais, P., Boissier, O., & Julian, V. (2021). Social and intelligent applications for future cities: Current advances. Future Generation Computer Systems, 114, 181-184. doi: 10.1016/j.future.2020.07.055

[11]   Hensher, D. A. (2001). Measurement of the valuation of travel time savings. Journal of Transport Economics and Policy (JTEP), 35(1), 71-98

[12]   Louviere, J. J., Hensher, D. A., & Swait, J. D. (2000). Stated choice methods: analysis and applications. Cambridge university press.

[13]   Jara-Díaz, S. R. (1998). Time and income in travel demand: towards a microeconomic activity framework. Garling, T., T. Laitia, and K. Westin, K.(eds.) Theoretical foundations of travel choice modeling. Pergamon.

[14]   Rose, J., & Bliemer, M. (2013). Sample size requirements for stated choice experiments. Transportation, 40(5), 1021-1041. doi: 10.1007/s11116-013-9451-z

[15]   Ben-Akiva, M., McFadden, D., Abe, M., Böckenholt, U., Bolduc, D., Gopinath, D., ... & Steinberg, D. (1997). Modeling methods for discrete choice analysis. Marketing Letters, 8(3), 273-286.

[16]   Toulouse, M. (2014). On Optimization Algorithms for Maximum Likelihood Estimation.Interuniversity Research Center on Enterprise Networks, logistics and transportation conference

[17]   Ramanujam, V., & Balakrishnan, H. (2011, June). Estimation of maximum-likelihood discrete-choice models of the runway configuration selection process. In Proceedings of the 2011 American Control Conference (pp. 2160-2167). IEEE

[18]   Kumar, C., Basu, D., & Maitra, B. (2004). Modeling Generalized Cost of Travel for Rural Bus Users: A Case Study. Journal Of Public Transportation, 7(2), 59-72. doi: 10.5038/2375-0901.7.2.4

[19]   Hensher, D., & Sullivan, C. (2003). Willingness to pay for road curviness and road type. Transportation Research Part D: Transport And Environment, 8(2), 139-155. doi: 10.1016/s1361-9209(02)00038-x

[20]   Elkafoury, Ahmed & Ali, Mohamed & Afify, Hafez. (2012). Transportation Planning Process for Urban Areas - Case Study Tanta City. MSC thesis, Tanta University

[21]   Hensher, D. (1994). Stated preference analysis of travel choices: the state of practice. Transportation, 21(2), 107-133. doi: 10.1007/bf01098788

[22]   Pearmain, D. (1991). Stated preference techniques. Richmond: Steer Davies Gleave.

[23]   Ilzarbe, L., Álvarez, M. J., Viles, E., & Tanco, M. (2008). Practical applications of design of experiments in the field of engineering: a bibliographical review. Quality and Reliability Engineering International, 24(4), 417-428.

[24]   Louviere, J. (1992). Experimental choice analysis: Introduction and overview. Journal Of Business Research, 24(2), 89-95. doi: 10.1016/0148-2963(92)90054-f.

[25]   A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models (2006), Frank S. Koppelman and Chandra Bhat, technical report, U.S. Department of Transportation Federal Transit Administration.

[26]   Horowitz, J. (1981). Identification and diagnosis of specification errors in the multinomial logit model. Transportation Research Part B: Methodological, 15(5), 345-360. doi: 10.1016/0191- 2615(81)90019-9.