Traffic & Transportation

A Different Nested Logit Model Structure Consistent with Random Utility Theory for Various Freight Types – A Case Study for Istanbul

Berna AKSOY[1], Mustafa GÜRSOY[2]

Original Scientific Paper
Submitted: 14 June 2023
Accepted: 13 Nov. 2023

[1] baksoy@yildiz.edu.tr, Yildiz Technical University, Civil Engineering Department, Transportation Division

[2] gursoy@yildiz.edu.tr, Yildiz Technical University, Civil Engineering Department, Transportation Division

Publisher:
Faculty of Transport and Traffic Sciences, University of Zagreb

ABSTRACT

Freight transport significantly contributes to urban traffic but is often overlooked by decision-makers compared to passenger transport. Conventional transportation modelling studies often use aggregate approaches for freight transport, undoubtedly due to the difficulty of data collection. However, the nature of freight transportation is much more complex. For this reason, examining the determinants of freight vehicle preferences with discrete approaches is crucial for the contributions that can be put forward, especially in local studies. To address this apparent gap in the study of local disaggregated approaches to freight transport, we present a discrete modelling-based methodology to investigate the factors that determine freight vehicle preferences for shippers and senders. The estimated nested logit model is constructed with the RU2 approach, the second part of random utility theory, thus avoiding the theoretical inconsistencies that arise when generic coefficients are used. As a result, the model structures provided satisfactory results compared to the literature. It was revealed that the factors affecting freight vehicle choice preferences were influenced by packaging preferences and differed according to freight groups. This local study is the first nested logit study for freight modelling in Istanbul and it is aimed to shed light on future national studies.

KEYWORDS

transportation modelling; freight modelling; random utility theory; discrete choice; logit models; nested logit.

1. INTRODUCTION

With the recent advances in transportation and communication technologies, the issue of "globalization" has been the focus of researchers, especially due to their relative economy [1]. In contrast to historical consumption behaviour, most goods consumed are produced kilometres away in modern times. Therefore, the demand for freight transportation is increasing tremendously, especially in urban areas [2, 3]. Freight transportation brings many challenges to urban mobility. Freight vehicles are designed to carry the maximum amount of material at one time, and therefore, freight vehicle sizes are large compared to passenger vehicles. In addition, they have low speeds and occupy many PCUs on the road, reducing road capacity. Due to their weight, freight vehicles are not designed for high speeds and can be a cause of traffic congestion in urban networks. Congestion and low speeds increase the travel time of all vehicles and cause economic losses and environmental pollution [4]. In order to mitigate or eliminate the negative impacts of freight transportation, understanding freight movements and managing freight traffic have been research topics for engineers [2, 3]. Freight transportation studies have remained an under-researched topic for many years. However, since the early 2000s, freight transportation has become a frequently researched topic, mainly due to increasing climate problems. If the last two decades can be divided into two parts, in the first decade, national and international levels but less detailed models were studied. Since 2013, urban-level and detailed models have become more frequent [5, 6]. Most of the existing studies adopt a deterministic and aggregated approach. However, the nature of freight movements is stochastic. In order to overcome this disadvantage of deterministic models, which react strongly to policy changes, disaggregate models have become widespread across Europe, especially in the last five years. A 2019 study [7] is Sweden's first disaggregated freight model in the national freight transport study area, modelling transport chain choice and shipment size based on individual shipment preferences. Another study [8] constructs a mixed discrete model for Europe for the first time and models the choice of freight chain and shipment size. With this increase in discrete modelling, local modelling studies with more theoretical clarity and better alignment with actual conditions are gaining importance. For this reason, nested logit models, one of the less constrained classes of discrete choice models, have been chosen for the study. Nested models are based on decision theory.

Decision makers consider different decision rules, such as dominance, satisfaction, lexicographic rules or optimisation of a scalar objective function. According to economists, for an object to have value, it must be 'useful' as a general rule. It is, therefore, expected to consider the choice behaviour of individuals/firms based on utility maximisation. When choosing one of a finite number of discrete bundles of attributes, discrete choice theory is a more appropriate basis than neoclassical consumer theory. In particular, probabilistic choice theory, which determines the probability that an individual/firm chooses any feasible alternative, provides a robust basis for analysing discrete choice situations. Rather than the constant utility approach of probabilistic choice theory, which is based on mathematical psychology, the random utility approach, more consistent with consumer theory [9], is often used in decision-making processes requiring discrete choice.

The random utility approach treats unobservable attributes, variation, measurement error and lack of observation due to using instrumental variables as random. It assumes that decision-makers always choose utility-maximising alternatives [10–13]. However, the analyst does not know the utilities with certainty and is therefore treated as random variables. From this perspective, the probability of choice alternative i is equal to the probability that the utility of alternative i, U_{in}, is greater than or equal to the utility of all other alternatives in the choice set. This can be written as follows, *Equation 1*:

$$P\left(i|C_n\right) = Pr\left[U_{in} \geq U_{jn}, all\ j \in C_n\right] \tag{1}$$

where U_{in} and U_{jn} are random variables, V_{in} and V_{jn} are deterministic (or representative) components of i and j's utility, ε_{in} and ε_{jn} are stochastic components and are called the error term, *Equation 2–3*.

$$U_{in} = V_{in} + \varepsilon_{in} \tag{2}$$

$$U_{jn} = V_{jn} + \varepsilon_{jn} \tag{3}$$

In modelling according to discrete choice theory, the type of model to be used is decided according to the distribution of the error term. If the stochastic terms are normally distributed, the model is probit; if they are distributed according to an independent extreme value distribution, the model is logit; if they are distributed according to a Cauchy distribution, the model is Arctan; and if they are distributed according to a generalised extreme value distribution, the model is nested logit [12, 14]. Logit models are frequently used in choice modelling studies due to their simple mathematical form and ease of estimation procedure [15].

In summary, the literature on freight transport modelling relies heavily on discrete choice modelling, yet the utilised variables vary between studies. With these considerations in mind, we present the following contributions in this paper:

– We propose a methodology that incorporates a discrete framework for freight transportation, and we aim that the study will make it possible to understand the advantages of specific freight vehicles. Our model structure includes variables not commonly used in freight transport models, such as packaging type and vehicle ownership.
– In order to test the effectiveness of the proposed modelling framework, we present a case study using real-world data obtained from Istanbul, Turkey.
– We discuss the details of this nested logit model with a consociate model structure for different freight types. We also provide a very explicit theoretical exposition to overcome normalisation problems often ignored when building nested models. This straightforward approach also aims to eliminate the 'non-comparability' problem, one of the biggest problems in freight transportation modelling studies.

In this study, a consociate nested logit model study was conducted for three different freight types in Istanbul. For the model study, the survey data of the Istanbul Logistics Master Plan commissioned by the Istanbul Metropolitan Municipality in 2016 were used. For this study, with 14778 surveys, the network and zone maps created within the scope of the master plan were transferred to QGIS open-source software, where the node points were connected to centroids in such a way that the closest one was assigned. Only urban shipments were considered in the study, and times and distances on the network were obtained using QUBE, a Bentley open-source software. After obtaining the data, version 4 of NLOGIT econometric software was used for all estimations.

The paper is organised as follows: Section 1 provides a detailed introduction to the research problem. Section 2 provides a comprehensive overview of the literature. Section 3 introduces the methodology and nested logit model specifications. Section 4 introduces the study area and the data. Section 5 is devoted to the empirical analysis and includes a detailed analysis of the common model and interpretation of the estimation results. Section 6 includes the evaluation of the common model for three different freight groups. Section 7 contains a discussion of the study. Section 8 is the conclusions section.

2. LITERATURE

There are two approaches to transportation modelling: aggregate and disaggregate. The aggregate approach, based on the traditional four-stage transportation model, directly models the aggregate share of decision-makers choosing an alternative. The disaggregate approach models individual preferences as a function of alternatives and individual characteristics. In freight transportation, disaggregate models are much less commonly used than in passenger transportation. The choice explicitly modelled in disaggregate freight models is the mode choice. The models used for these single discrete choices are multinomial logit, probit, nested logit, ordered generalised extreme value, cross nested logit, mixed logit and latent class logit. The multinomial logit model is the most common and widely used model [15, 16]. There are three basic assumptions underlying the MNL formulation. The first assumption is that the random components of the utilities of different alternatives are independent and identically distributed (IID) with an extreme value (or Gumbel) distribution of type I. The independence assumption means that no common unobserved factors influence the utilities of various alternatives. This assumption is violated, for example, if a decision maker assigns the same utility to all species. In such cases, the same underlying unobserved factor influences the benefits of all transit modes. The second assumption of the MNL model is that it preserves the homogeneity of responsiveness to the attributes of alternatives across individuals. More specifically, the MNL model does not allow for responsiveness (or taste) changes to an attribute (e.g. travel cost or ”rave' time in a mode choice model) due to unobservable individual characteristics. However, unobservable individual characteristics can and often will affect sentiment. Ignoring the influence of unobservable individual characteristics can lead to biased and inconsistent parameter and choice probability estimates. The third assumption of the MNL model is that the error variance–covariance structure of the alternatives is the same across individuals (i.e. the assumption of error variance–covariance homogeneity). Error variance–covariance homogeneity implies the same competitive structure across alternatives for all individuals, which is often difficult to justify. The assumption of the same variance across individuals can be violated, for example, if the transit system offers different levels of comfort (an unobservable variable) on different routes. Together, the three assumptions discussed above lead to the simple and elegant closed-form mathematical structure of MNL. However, these assumptions also confront the MNL model with the "independence of irrelevant alternatives" (IIA) property at the individual level. Therefore, relaxing the three assumptions may be necessary in many choice contexts [17].

The (NL) model represents a partial relaxation of the IID and IIA assumptions of the MNL model. While more advanced models relax the IID assumption fully (via covariances), the NL model represents an excellent advance for the analyst regarding choice studies. As with the MNL model, the NL model is relatively easy to estimate and offers the added benefit of having a closed-form solution. More advanced models, such as the multinomial probit (MNP), heteroskedastic extreme value models (HEV) and the random parameter logit (RP), also called the mixed logit (ML) model, relax the IID assumption in terms of covariances; however, they are all open-form solutions and thus require complex analytical calculations to identify changes in choice probabilities through varying levels of attributes and socio-demographic characteristics [18–20].

The nested model was first derived in 1973 by Ben-Akiva [21] as a generalisation of the joint logit model and later be formulated in different ways based on utility maximisation by Daly and Zachary in 1979 [22] Williams in 1977 [23], Ben-Akiva and Lerman in 1979 [24] and McFadden in 1978 [25], respectively, who showed that it is a special case of the generalised extreme value (GEV) model [12].

The use of the nested logit model in freight transportation is quite common. Bradley and Daly modelled shipment size and mode choice as a joint nested model with a combination of SP/RP data [26]. Jiang et al. estimated mode prediction for road, rail and combined transports from French shipper surveys with a nested model that includes shipper and firm characteristics and the distance variable but not the time and cost components [27]. De Jong et al. constructed a nested logit model for the Nord-Pas de Calais region in France based on RP/SP prediction, where the explanatory variables are transportation time, cost, reliability, flexibility and mode operation frequency [28]. Arunotayanun modelled mode choice of small truck, large truck and rail using Indonesian SP shipper surveys and French RP shipper surveys for different cross-nested structures of state-owned roads, build and operate roads, rail and combined transport [29, 16]. The Transtool3 model, an international model that partially uses aggregate and disaggregate data, uses nested models as chain choice models [8]. The TRIMODE, also an international model, uses aggregate data but partially includes different nested models to select feeder mode, main mode and vehicle type [30]. Using SP surveys, Nughoro et al. constructed a nested logit regional hinterland choice model for Indonesia-Java [31]. It is stated in various sources that the detailed analysis of freight transportation, especially at urban and regional levels, with advanced models using disaggregated data, can eliminate many problems and is becoming more and more critical [5].

3. METHODOLOGY OF NESTED MODEL

The main advantage of the nested logit model over the MNL model is that it allows for dependency and correlation between similar alternatives. According to the nested logit model, the probability ratio for each alternative within the same nest is independent of the preference of the other (i.e. the IIA assumption is satisfied within the nest) [20].

The NL model structure is in a hierarchical representation presented in *Figure 1* [20]. The nested logit model does not express a logical or behavioural relationship or choice ordering; it has only a mathematical purpose. The NLOGIT software allows for four estimation levels, but two-level nested models are the most used in the literature [20, 32]. In a two-level nested model, the bottom level consists of alternatives and the top level consists of branches. In a particular case, a branch with only one alternative is called a degenerate branch. A nested structure with two levels and degenerate branches can be seen in *Figure 1* [20, 33].

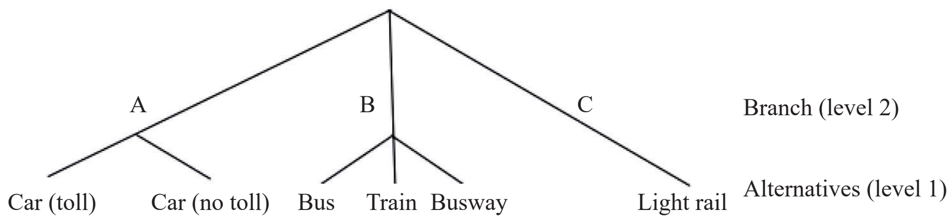


Figure 1 – A nest with two levels and degenerate branches [11]

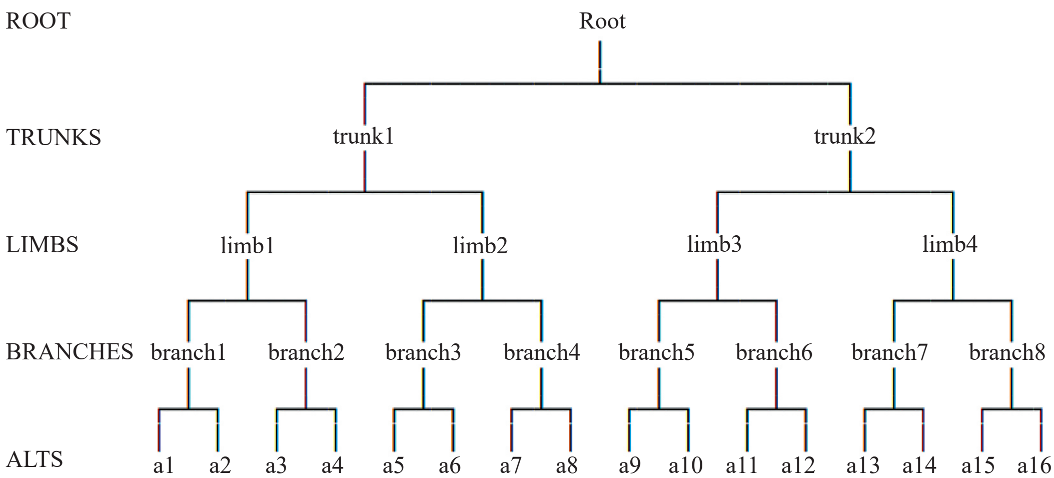


Figure 2 – A nest structure with four levels [13]

In a nested structure, different names are used for each level. The widely used NLOGIT software allows a four-level estimation. The nomenclature for this four-level estimation is given in *Figure 2*.

The probability of an alternative *i* being selected among *n* alternatives is calculated in the following general form, *Equation 4* [12].

$$P_i = \frac{\exp(\mu V_i)}{\sum_{J\in n}\exp(\mu V_j)} , \qquad i \in n \tag{4}$$

In this equation, μ is a non-negative scale parameter. It is not possible to calculate μ values for each utility function separately. In the MNL model, it is assumed that the epsilons, i.e. the errors, are equal to each other and uniformly distributed [12] and, therefore, μ=1. There are studies in the literature where the variance of the error terms or μ is fixed to 1. When μ=1, calculating different error terms is not bothered. However, in the NL model, μ can no longer be removed from the equation. There is a scale parameter that is equal for all alternatives under the same branch.

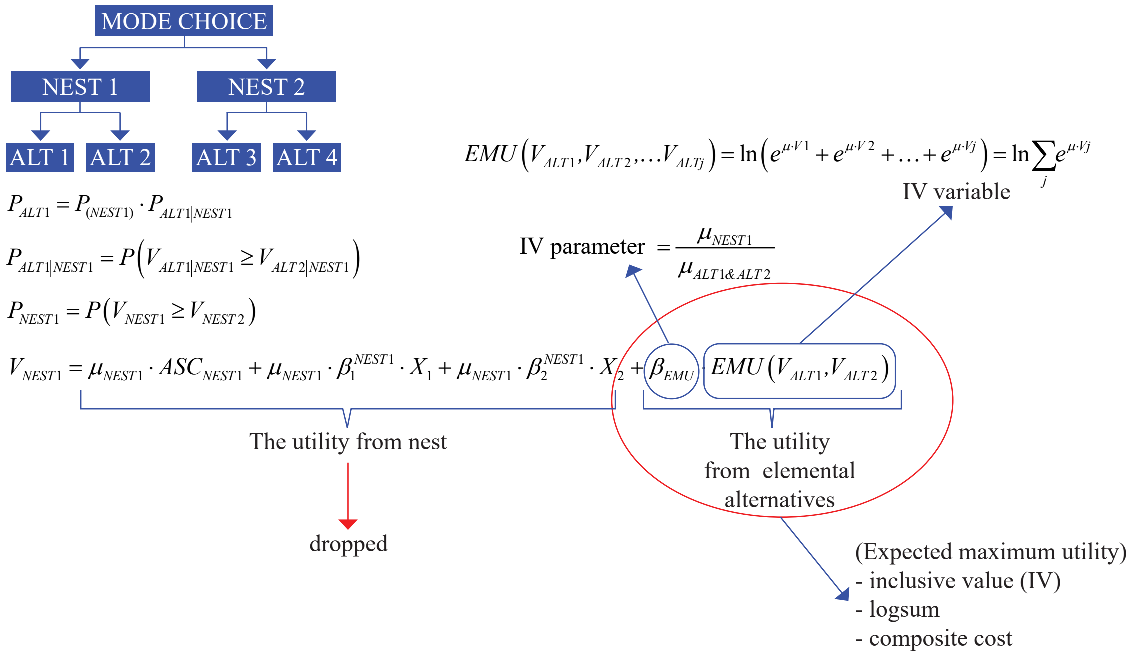
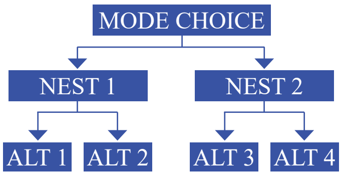


Figure 3 – Conditional probability calculations [14]

As shown in *Figure 3*, conditional probability calculations are now involved once the hierarchical structure is established. When calculating the utility function at the top level, if there are no variables affecting choice at the top level (as the nest structure does not reflect a choice behaviour, there usually has to be really conclusive evidence for a utility coming from it), the utility from the nest drops out of the equation. The utility from the elemental level forms the IV term. This term has two parts: the IV parameter and the IV variable. The ratio of the scale parameters at the nest level to the elemental level is called the IV parameter. This IV term, coined by Hensher [20], connects the two levels of the NL model and is known in the literature as expected maximum utility, inclusive value, log sum, or composite cost. A general notation for the utility functions underlying the calculations is given in *Figure 4*.

Since it is impossible to calculate the elemental and nest levels and values in the IV parameter separately, these values are normalised. The value of the IV parameter should be 0<μ≤1. This implies a non-zero correlation between nests and is suitable for a nested logit structure. A value greater than one or less than 0 is inconsistent with the theoretical basis of the NL model and the NL model is rejected. When the value of the IV parameter is 0, there is a perfect correlation between pairs of alternatives under nests, and selection is deterministic. In contrast, when the value of the IV parameter is equal to 1, the model collapses to MNL and remains at the elemental level. A decreasing value of the IV parameter indicates increased substitution between alternatives in the nest [15]. While evaluating the model, the boundary conditions of 0 and 1 are checked by calculating the



$$V_{ALT1|NEST1} / \mu_{ALT1\&ALT2} = ASC_{ALT1|NEST1} + \beta_1^{ALT1|NEST1} \cdot X_1 + \beta_2^{ALT1|NEST1} \cdot X_2$$

$$V_{ALT2|NEST1} / \mu_{ALT1\&ALT2} = ASC_{ALT2|NEST1} + \beta_1^{ALT2|NEST1} \cdot X_1 + \beta_2^{ALT2|NEST1} \cdot X_2$$

$$V_{ALT3|NEST2} / \mu_{ALT3\&ALT4} = ASC_{ALT3|NEST2} + \beta_1^{ALT3|NEST2} \cdot X_1 + \beta_2^{ALT3|NEST2} \cdot X_2$$

$$V_{ALT4|NEST2} / \mu_{ALT3\&ALT4} = ASC_{ALT4|NEST2} + \beta_1^{ALT4|NEST2} \cdot X_1 + \beta_2^{ALT4|NEST2} \cdot X_2$$

$$V_{NEST1} = \mu_{NEST1} / \mu_{ALT1\&ALT2} \cdot IV_{NEST1}$$

$$V_{NEST2} = \mu_{NEST2} / \mu_{ALT3\&ALT4} \cdot IV_{NEST2}$$

Figure 4 – A general illustration of the utility functions based on the calculations [14]

Wald A and Wald R statistics for the IV parameter to be significant. When Wald R is greater than +1.945 or less than -1.945, the IV parameter is statistically different from zero. The same check is made for the Wald A value, and a value greater than +1.945 or less than -1.945 indicates that the IV parameter is statistically significant. The Wald A formula is given below, *Equation 5* [15, 20, 34].

$$Wald_A = \frac{IV - 1}{Standart\ Error} \tag{5}$$

An unfortunate feature of the NL model is that it tries to estimate more parameters than can be estimated with the amount of information available. The general logic of the NL model is to normalise the scale parameters at one level to 1 and estimate all other parameters according to these normalisations. While it is possible to normalise scale parameters at any level of an NL model, it is expected to normalise all scale parameters at the lowest level of the tree (RU1) or all scale parameters at the highest level (RU2). According to one study [18], RU1 and RU2 approaches are equivalent if all parameters used in a nested model are alternative-specific. Most studies on this topic do not mention whether the estimated models are based on RU1 or RU2. This makes it difficult to compare the results of the studies. Specifying the level at which the scale parameters are normalised would greatly help to improve the research [20, 33]. Normalisation using RU2 is preferable to RU1 because RU2 guarantees that the model under any specification of all parameters of the model (i.e. whether generic or alternative-specific) is consistent with global utility maximisation under random utility theory without having to add a dummy node level below level 1. More details can be found in [20, 33].

Silberhorn et al. [35] and Ortuzar and Willumsen [36] mention two different nested structure specifications: UMNL (Utility Maximisation Nested Model) and NNNL (Non-Normalised Nested Model). Mac Fadden's UMNL model is consistent with utility maximisation theory if the conditions $0<\mu\leq 1$ are satisfied for all values of μ. The main difference between Daly's NNNL model and UMNL model formulations is that the NNNL model excludes the inverse of the log sum parameters in the utility functions of the alternatives in each nest [37]. In model estimation, the nested logit feature implemented in the software must be considered. If only alternative-specific coefficients are present in the model estimated with software that includes NNNL, the model coefficients can be estimated by rescaling the nested coefficients. However, when a generic coefficient enters the model, if the NNNL model does not impose restrictions on the scale parameters, the model will not be consistent with random utility theory. Software such as SAS and ALOGIT can only estimate NNNL, while software such as STATA, LIMDEP and GAUSS can estimate both specifications [35]. A very detailed evaluation of UMNL and NNNL is available. In these evaluations, [35] it is proved that an RU2 estimation in UMNL software is consistent with random utility theory. When using THE UMNL software, it is necessary to distinguish between RU1 and RU2. In this study, NLOGIT4, a LIMDEP software, was used and RU2 normalisation was performed. There are two estimation procedures for nested models: sequential estimation and simultaneous estimation. Sequential estimation estimates the NL model starting from the lower levels.

Simultaneous estimation estimates the NL model simultaneously. Here, the estimator is called Full Information Maximum Likelihood (FIML) and is superior to sequential estimation. NLOGIT4 uses the FIML estimator in its estimation procedure [38].

4. DATA AND STUDY AREA

Within the scope of the study, commodity flow surveys conducted by Istanbul Metropolitan Municipality in 2016 for the Istanbul Logistics Master Plan were used. For this study commissioned by the Municipality, the surveys were designed to estimate the logistics mobility of the companies receiving and providing services in Istanbul province, districts and sectors. The total sample size was 71,992 firms due to the sampling design designed by TSI (Turkish Statistical Institute). A stratified two-stage cluster sampling method was used. Districts and sectors were used as external stratification criteria. The first stage sampling unit is the districts; the second stage sampling unit is the workplaces systematically selected from each cluster as a sector. The framework used in sampling the research is the workplace registry records obtained by TSI from the Republic of Turkey Ministry of Treasury and Finance. This study excludes sectors that do not affect the Istanbul logistics network. As a result, 520,324 companies were determined as the population size. For this population size, 71,992 firms were surveyed. All of the surveys were conducted face-to-face and included surveys for each district in a distribution that will ensure the whole sample rate, so it represents the situation of each district [39]. From a total of 152,488 surveys for ten freight types, 14,778 surveys for three freight types were cleaned and used via SPSS data. In this context, only urban freight transportation was included in the study. The boundaries of the study area are shown in *Figure 5*.

Figure 5 – Study area

The surveys include data such as vehicle ownership, whether it contains hazardous materials, type of packaging, shipment size, origin-destination points for the movement, type of shipment and type of vehicle in which the shipment is being shipped. In addition to this information, time and cost variables, which are most used in the modelling literature, were also needed. For this purpose, zone and network maps prepared within the scope of the Istanbul Logistics Master Plan were utilised. The linear function used to calculate the costs is given in *Equation 6*.

$$TTC = TYC + DP + LC + LUC + RM + FC \quad (6)$$

where: TTC – total transportation cost, TYC – tyre cost, DP – depreciation cost, LC – labour cost, LUC – loading and unloading cost, RM – repair and maintenance cost, FC – fuel cost.

In the cost calculations, fuel consumption values were taken from the factory catalogues of famous brands for five different vehicle types, and the vehicle cost was determined for each vehicle. Depreciation costs based on a 7-year lifespan, tire costs assuming a tire change every 300,000 km, labour costs based on 2016 minimum wage and legal working hours, loading and unloading costs, and repair and maintenance costs twice the tire costs were calculated. In order to obtain the fuel consumption, which constitutes the most oversized item in the cost, the distances between O-D points were needed. QGIS open-source software was used for this calculation. In the maps prepared for the logistics master plan, Istanbul province was divided into 573 zones, and nodes were first assigned to the road network map. Then, centroids were added to connect the nodes to the closest ones. These maps were transferred from QGIS to CUBE, one of BENTLEY's open-source software, and a distance matrix of 573×573 was obtained. The shortest routes according to the distance were selected, and the times for these routes were generated.

Table 1 – Number of observations and descriptions for three different freight types

Freight type	Main type	Chosen vehicle type					Total	Description
		Trailer	Truck	Light duty vehicle	Panelvan	Motorcycle		
		# of obs.	# of obs.	# of obs.	# of obs.	# of obs.	# of obs.	
Freight type 8	Computers, electronics and optical products	48	106	663	712	16	1545	Computers, electronic and optical products, electrical equipment
Freight type 4	Textiles, apparel, leather and related products	121	612	1746	2226	11	4716	Textiles, apparel, leather and related products
Freight type 10	Other manufactured goods	758	1552	2735	3390	82	8517	Except food, mineral products, rubber, chemicals, paper products, base metals and furniture

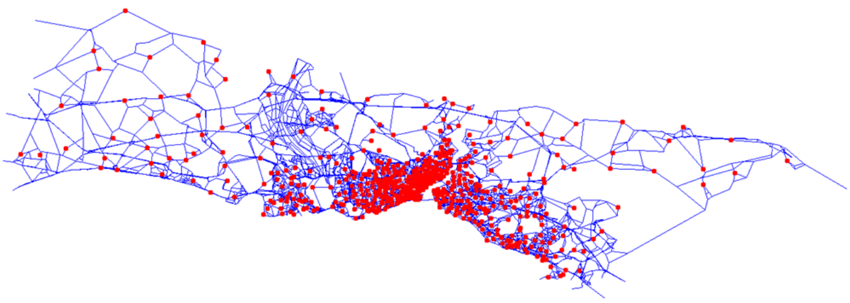


Figure 6 – Network map (QUBE)

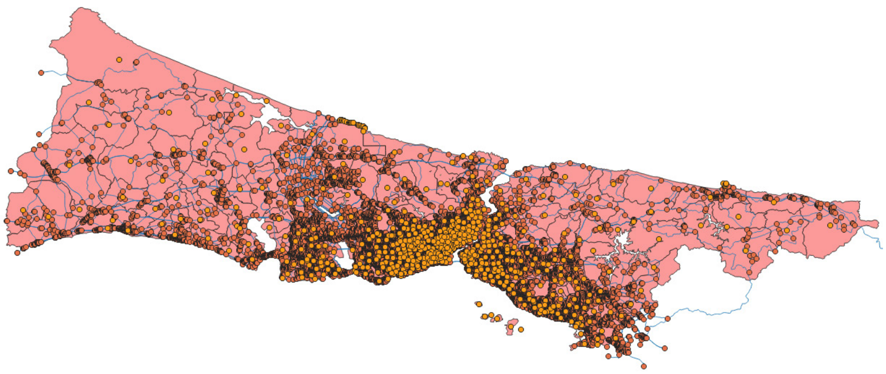


Figure 7 – Nodes and centroids connected to the network map

When obtaining the speeds for a route, the free flow speeds for that route were used, and for routes where free flow speeds were not available, calculations were made according to the 50 km/h urban speed constraint. Intraregional time and distance calculations (e.g. travel from zone 17 to zone 17) equal half the minimum time and distance from that zone to other zones. Fatih Sultan Mehmet Bridge, 15 Temmuz Şehitler Bridge and the Avrasya Tüneli (Eurasia Tunnel Project), which are prohibited for heavy vehicles, were excluded from the network by writing an additional constraint function. The number of observations and detailed descriptions of the three freight types and two views of the network maps where the centroids are connected are given in *Table 1*, *Figures 6 and 7*, respectively.

The flow chart of the study is given in *Figure 8*. The best model was sought according to this flow.

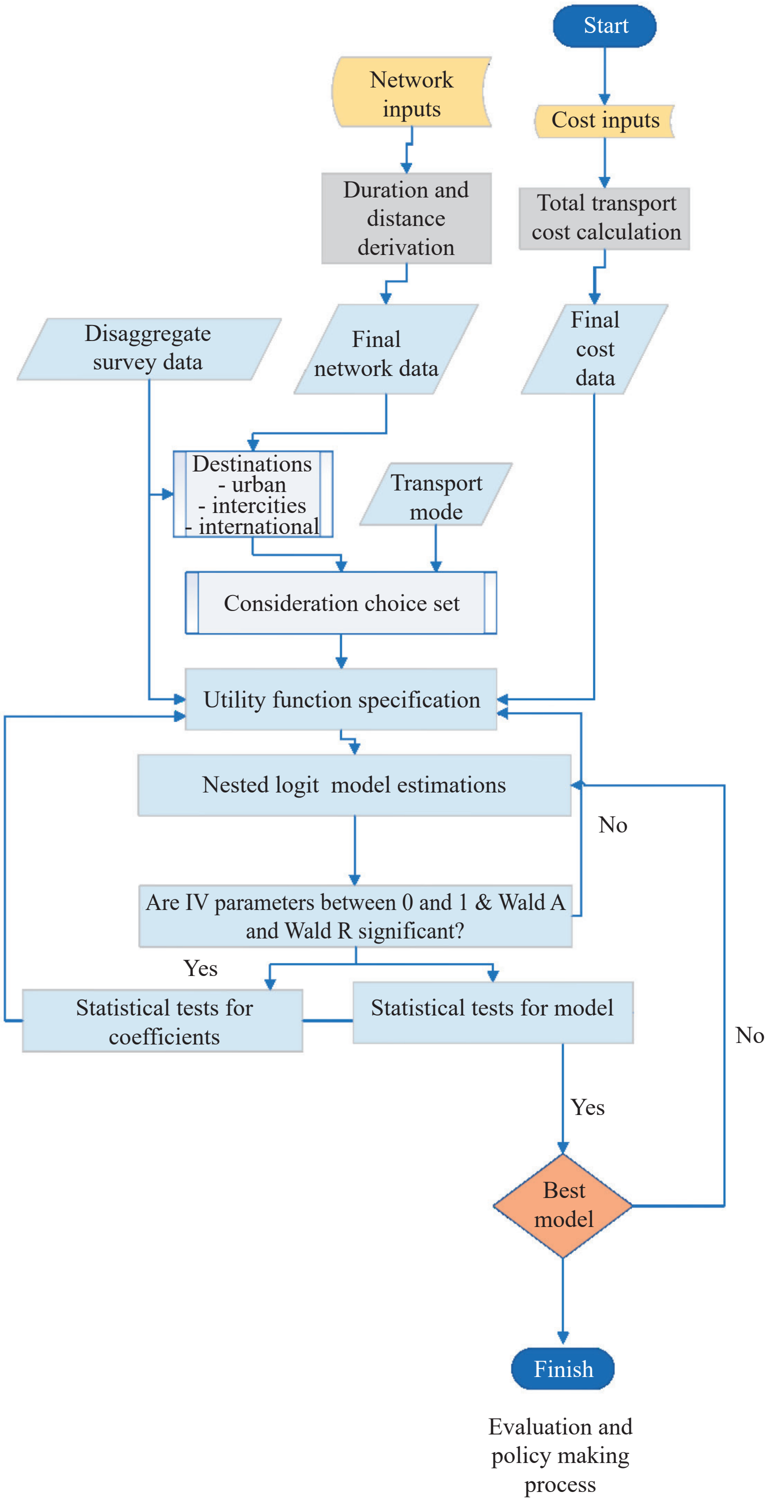


Figure 8 – Flowchart of the study

5. EMPIRICAL ANALYSIS AND ESTIMATIONS

A representative nest structure for the model designed within the scope of the study is given in *Figure 9*. The independent variables used in the model and their explanations are given in *Table 2*. Accordingly, five different vehicle types are considered under two different branches. The distinction between large and small freight vehicles is primarily heuristic. The conditions under which a nested structure can perform cannot be predicted at the beginning of the modelling phase. Here, an aggregation was made for 54 different vehicle groups [39]. A wide variety of vehicles, such as refrigerated vehicles, taut liners and ordinary trucks, were grouped according to their wheelbase and weight and grouped all vehicles into five different classes. The variables entered into the model are vehicle ownership, shipment size, whether the transportation is carried out within the same zone, cost, dry bulk freight and parcel freight. A correlation check was made between the model variables in the survey data, and variables with a pairwise correlation greater than 0.5 were not included in the model. Variables that give convergence errors when written into functions were also not used in the model.

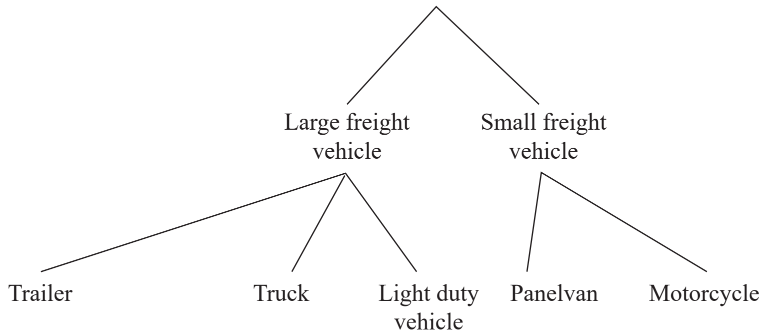


Figure 9 – Nest structure investigated in the study

Table 2 – Variables in the model and their explanations

Independent variable name	Variable abbreviation	Explanation	Variable type
Car ownership	CO	1 = own property 0 = shipping by rental car	Alternative specific
Shipment size	SIZE	Weight [kg]	Generic
Intrazonal shipping	INTRA	1=Yes 0=No	Generic
Cost	COST	Value (₺)	Alternative specific
Dry bulk freight	PTC	1=Yes 0=No	Generic
Parcel freight	PTE	1=Yes 0=No	Generic

6. INTERPRETATION OF RESULTS

When evaluating the coefficients for nested models, it should be remembered that probability calculations are conditional. Therefore, the coefficients are not a precise indicator for the interpretation of the model but a guide for the interpreter. Whether the coefficients are positive or negative is more important than the absolute value of the coefficients in interpreting the results. When evaluating nested models, it is more important that nested models calculate probability calculations more accurately and realistically rather than interpreting the magnitude of the coefficients. Therefore, they contain meaningful information for policymaking, especially scenario analysis. The coefficients, t-statistic values and IV parameter values obtained as a result of the models are given in *Table 3*.

A consociate model structure for the three different freight groups was estimated with NLOGIT4 software, which is based on the RU2 approach to UMNL estimation, thus eliminating the possibility that the model may be inconsistent with utility theory despite the generic coefficients used in the model. The IV parameters of all three estimated models remain within the acceptable range of 0–1 and are significantly smaller than 1 (see Wald A and Wald R statistics). Therefore, the MNL model is rejected and the NL models are accepted. Although it is not very accurate to evaluate the variable coefficients of the three models in a proportionally common way due to the different IV parameters, almost all variables preferred in the model are statistically significant at 95% and above. This indicates that there are more infrequent non-significant variables in the

Table 3 – Coefficients and t-statistic values of the three models

	Freight type 4		Freight type 8		Freight type 10	
Variable	Coef.	t-stat.	Coef.	t-stat.	Coef.	t-stat.
Trailer						
SIZE	0.0009	18.1370*	0.0004	8.8560*	0.0008	31.9600*
INTRA (1=Yes)	-0.2847	-2.7510*	0.5228	3.1370*	-0.0143	-0.1980
CO (1=Own Property)	-1.0386	-3.3000*	-2.2059	-3.7480*	-0.9957	-8.3120*
PTE (1=Yes)	0.0036	0.1200	-0.2297	-2.4410*	0.1363	2.9320*
PTC (1=Yes)	-0.3707	-0.7010	1.2508	3.4540*	0.1851	1.2290
CONSTANT	-4.0874	-26.1740*	-2.8632	-12.9840*	-2.5338	-24.6980*
Truck						
SIZE	0.0009	18.1370*	0.0004	8.8560*	0.0008	31.9600*
INTRA (1=Yes)	-0.2847	-2.7510*	0.5228	3.1370*	-0.0143	-0.1980
CO (1=Own Property)	-0.4985	-2.7840*	-0.0354	-0.1410	-1.2336	-11.5380*
PTE (1=Yes)	0.0036	0.1200	-0.2297	-2.4410*	0.1363	2.9320*
PTC (1=Yes)	-0.3707	-0.7010	1.2508	3.4540*	0.1851	1.2290
CONSTANT	-2.6975	-22.2930*	-2.5698	-12.4140*	-1.6975	-17.9690*
Light duty vehicle						
INTRA (1=Yes)	-0.2847	-2.7510*	0.5228	3.1370*	-0.0143	-0.1980
PTE (1=Yes)	0.0036	0.1200	-0.2297	-2.4410*	0.1363	2.9320*
PTC (1=Yes)	-0.3707	-0.7010	1.2508	3.4540*	0.1851	1.2290
CO (1=Own Property)	-0.3590	-4.2530*	-0.2123	-1.5390	-0.0696	-0.9120
CONSTANT	-0.1127	-1.7430	0.0798	0.5760	0.2368	3.0940*
Panelvan						
COST	-0.0224	-5.8170*	-0.0249	-5.7410*	-0.0154	-6.9580*
CO (1=Own Property)	0.0257	0.8010	0.1227	1.5140	-0.1876	-3.3950*
PTE (1=Yes)	0.0036	0.1200	-0.2297	-2.4410*	0.1363	2.9320*
CONSTANT	0.2060	9.4070*	0.5097	5.5400*	0.7246	12.8730*
Motorcycle						
COST	-0.0681	-5.2970*	-0.1159	-5.4370*	-0.0721	-6.2450*
# of Observations	4716		1545		8517	
LL (β) estimated model	-4283.215		-1480.557		-8476.092	
LL (β*)base	-7542.912		-2473.090		-13852.64	
ρ^2	0.432		0.401		0.388	
IV Parameter	0.044		0.143		0.197	
Standard error	0.0065		0.0127		0.00765	
Wald A	-147.08		-67.480		-104.967	
Wald R	6.730		11.245		25.802	
Chi-square (-2LL)	6519.395		1985.066		10753.10	
Critical chi-square	24.996		24.996		24.996	
Degrees of freedom	15		15		15	

**Meaningful at 95% confidence interval. Statistically, variables that are significant in confidence intervals of 95% and above are marked with an asterisk.*

model (it is common in logit modelling to remove non-significant variables when there is no evidence that they improve the model). When an individual evaluation is made in all three models, as the freight size increases, the probability of choosing larger vehicles increases, which is expected. Especially in the FT4 textile and leather products group, this increase is higher (0.009), whereas it is more significant for the FT10 freight group (t-test 31.9600). This is also expected since the FT10 freight group is mixed. The generic intrazonal shipping coefficient was insignificant only for the FT10 model. The fact that the transportation is within the zone decreases the probability of choice for all types for the FT4 textile products group (-0.2847). In contrast, it has an increasing effect on the probability of selection for all types for the FT8 freight group (0.5228). Intra-zonal transports generally imply closer distances and the smallest possible vehicle size choice. Since it is used as a generic coefficient here, FT4 has more spatial dispersion in the distribution of FT4 and FT8 freight groups. It can be inferred that the mobility of the FT8 device sector is more localised, while the textile sector has a more dispersed market in Istanbul. This indicates that this coefficient can be used as an alternative specific coefficient in future studies. When the vehicle ownership variable is evaluated, it is statistically significant at a 95% confidence interval in all models for trailer type (-1.0386, -2.2059, -0.9957), and it is seen that the probability of transportation by trailer decreases as vehicle ownership increases in all models, and this decrease is predominantly higher for the FT8 model (-2.2059). As vehicle ownership increases, the probability of choosing the truck type (-0.4985, -1.2336) decreases significantly in the FT4 and FT10 freight groups at a 95% confidence interval, while for the LDV type, this is meaningful only in the FT4 freight type (-0.3590). Accordingly, firms do not tend to carry out transportation with their own vehicles, especially when they prefer LDV vehicles for the FT4 freight type. For the FT10 freight type, it can be said that when vehicle ownership increases, the probability of choosing the panelvan decreases at a 95% confidence interval (-0.1876). The fact is that the parcel freight shipment is statistically significant for FT8 and FT10 freight groups. While it decreases the probability of choice for all vehicle types for FT8, it increases the probability of choice for all types for FT10. This indicates that for the FT10 freight type, this packaging type is frequently preferred by companies. The dry bulk coefficient is significant only for type FT8 and increases the probability of choice for all modes, indicating that dry bulk is the dominant packaging type for FT8. The cost variable is statistically significant for all three models. It has a decreasing effect on the selection probabilities for the motorcycle and panelvan modes in all three models, which indicates that the model works correctly and is in line with the results accepted in the literature [15, 40]. Although the Mac Fadden Pseudo R^2 (ρ^2) values for the nested models are not as significant as in the MNL models, it is seen that the obtained models give relatively high results when compared to the base model. Since the Chi-square values of the models are higher than the critical values in the Critical Chi-square evaluations made according to the degrees of freedom, it can be said that the models are advanced.

7. DISCUSSION

In essence, the value of nested models lies in the detail they provide and how they alleviate the MNL model's limitations. Therefore, the consociate nested model structure for the three different freight types is valuable in detailing and understanding the factors affecting mode choice for freight transportation in urban areas. When looking at the actual and model estimation probability values in *Table 4*, it is seen that the three nested models that were validated using a consociate model structure could estimate the probability values quite closely. In particular, it is seen that the FT10 freight type gives minor differences between the actual situation and the model for truck, LDV and panelvan types (18.53% and 18.22%; 33.58% and 32.11%; 38.45% and 39.80%) while the FT4 type gives the minor differences between the actual situation and the model for trailer and motorcycle (2.64% and 2.57%; 0.04% and 0.023%). Although all three models worked successfully when an evaluation is made for the best and closest results, the best model is the FT10 model, and the weakest model is the FT8 model. The main reason for the relatively significant differences in all models for the motorcycle mode is that it is sometimes difficult to associate choice with the change of variables when the number of data is limited. These all revealed that packaging types, vehicle ownership, intra-zone transportation situations, freight sizes and total costs are influential factors in the choice of freight vehicle type. The estimation procedure used is clearly presented, which allows comparisons to be made with future studies. In addition, various scenario analyses will be possible in case of changes in the numerical variables (except dummies) used (e.g. increase in dimensions or changes in transportation costs). This study has not yet presented a scenario analysis. Scenario analyses with nested models are meaningful and valuable for decision-makers, mainly because they reveal how the mode choice may change when a change is made in a traffic component. This study has a theoretical underpinning and future studies will focus mainly on policymaking.

Table 4 – Actual and model probability results

Freight type	Result	Mode choice (%)				
		Trailer	Truck	Light duty vehicle	Panelvan	Motorcycle
FT4	Model	2.64	13.33	39.03	44.97	0.04
	Actual	2.57	12.98	37.02	47.20	0.23
FT8	Model	3.26	7.23	46.10	43.06	0.36
	Actual	3.11	6.86	42.91	46.08	1.04
FT10	Model	9.05	18.53	33.58	38.45	0.40
	Actual	8.90	18.22	32.11	39.80	0.96

8. CONCLUSION

In this study, we estimate a nested logit model for the province of Istanbul using an extensive data set and a very explicit estimation procedure. Using a nested model provides computational simplicity and relaxes the assumptions of MNL models. In developing societies, transportation planning studies are often conducted at the aggregate level. This is mainly due to the difficulty of obtaining and working with disaggregated data. However, the use of nested logit models, which stand out among disaggregate models in terms of realistic results and computational simplicity, can benefit technical decision-makers in local administrations. Especially in solving the traffic congestion problem, which is an essential problem for local administrations in a large metropolis like Istanbul, the information provided by the models to technical decision-makers will make it possible to reveal how vehicle types can be substituted in various price policy applications and how the vehicle composition in traffic will change. This study, considered to be valuable for its theoretical clarity and the extensive survey data it contains, will shed light on future studies on locally-based freight transportation modelling.

REFERENCES

[1] Gokasar I, Şahin O. Evaluation of the travel behaviors and attitudes of the passengers towards the BRT line in Istanbul. *Journal of Transportation Systems*. 2020;(1):16-27. DOI: 10.5281/zenodo.3696656.
[2] Chankaew N, et al. Freight traffic analytics from national truck GPS data in Thailand. *Transportation Research Procedia*. 2018;34:123-130. DOI: 10.1016/j.trpro.2018.11.023.
[3] Kiba-Janiak M. Urban freight transport in city strategic planning. *Research in Transportation Business & Management*. 2017;24:4-16. DOI: 10.1016/j.rtbm.2017.05.003.
[4] Vullapu SS, Jain J, Tarafdar AK. Streamlining freight transport through planning interventions in Vijayawada city. In: Chatterjee U, Bandyopadhyay N, Setiawati MD, Sarkar S. (eds) *Urban Commons, Future Smart Cities and Sustainability*. Springer Geography. Springer, Cham; 2023. DOI: 10.1007/978-3-031-24767-5_37.
[5] De Jong GC, de Bok M, Thoen S. Seven fat years or seven lean years for freight transport modeling? Developments since 2013. *Journal of Transport Economics and Policy*. 2021;55(2):124-140.
[6] De Jong GC, et al. Recent developments in national and international freight transport models within Europe. *Transportation*. 2013;40:347-371. DOI: 10.1007/s11116-012-9422-9.
[7] Abate M, et al. A disaggregate stochastic freight transport model for Sweden. *Transportation*. 2019;46:671-696. DOI: 10.1007/s11116-018-9856-9.
[8] Jensen AF, et al. A disaggregate freight transport chain choice model for Europe. *Transportation Research Part E*. 2019;121:43-62. DOI: 10.1016/j.tre.2018.10.004.
[9] Manski CF. The structure of random utility models. *Theor Decis 8*. 1977;229-254. DOI: 10.1007/BF00133443.
[10] Luce RD, Suppes P. Preference, utility, and subjective probability. In: Luce RD, Bush RR, Galanter E. (eds) *Handbook of Mathematical Psychology*. Vol. III. New York: Wiley; 1965. p. 252-410.
[11] Tezcan HO, Öğüt KS, Çidimal B. A multinomial logit car use model for a megacity of the developing world: Istanbul. *Transportation Planning and Technology*. 2011;34(8):759-776. DOI: 10.1080/03081060.2011.613585.
[12] Ben-Akiva M, Lerman SR. Discrete choice analysis: Theory and application to travel demand. Cambridge, MA: MIT press; 1985.
[13] Manski CF. *The analysis of qualitative choice*. PhD thesis. MIT; June 1973.
[14] Domenchic T, McFadden DL. *Urban travel demand - A behavioral analysis*. Chapter 4. North-Holland Publishing Co.; 1975.

[15] Koppelman FS, Bhat C. *A self-instructing course in mode choice modeling: Multinomial and nested logit models*. U.S. Department of Transportation; 2006.
[16] Tavasszy L, de Jong GC. *Modeling freight transport*. Elsevier; 2014. DOI: 10.1016/C2012-0-06032-2.
[17] Bhat C. Flexible model structures for discrete choice analysis. In: Hensher DA, Button KJ. (eds) *Handbook of Transport Modelling*. Oxford, UK: Elsevier Science Ltd; 2000. p. 71-89.
[18] Louviere JJ, Hensher DA, Swait JD. *Stated choice methods: Analysis and applications*. Cambridge, UK: Cambridge University Press; 2000.
[19] Kenneth T. *Discrete choice methods with simulation*. Cambridge: Cambridge University Press; 2003.
[20] Hensher DA, Rose JM, Greene WH. *Applied choice analysis: A primer*. Cambridge, UK: Cambridge University Press; 2005.
[21] Ben-Akiva M. *Structure of passenger travel demand models*. PhD thesis. Department of Civil Engineering, MIT; 1973.
[22] Daly A, Zachary S. Improved multiple choice models. In: Hensher D, Dalvi Q. (eds) *Identifying and measuring the determinants of model choice*. Teakfield, London; 1979.
[23] Williams HCWL. On the formation of travel demand models and economic evaluation measures of user benefit. *Envir. and Planning*. 1977;9:285-344.
[24] Ben-Akiva M, Lerman S. Disaggregate travel and mobility choice models and measures of accessibility. In: Hensher D, Stopher P. (eds) *Behavioral travel modelling*. Croom Helm, London; 1979.
[25] McFadden D. Modelling the choice of residential location. In: Karlquist A, et al. (eds) *Spatial interaction theory and residential location*. North Holland, Amsterdam; 1978. p. 75-96.
[26] Bradley MA, Daly AJ. Estimation of logit choice models using mixed stated preference and revealed preference information. In: Stopher PR, Lee-Gosselin M. (eds) *Understanding travel behavior in an era of change*. Amsterdam: Elsevier; 1997.
[27] Jiang F, Johnson P, Calzada C. Freight demand characteristics and mode choice: An analysis of the results of modeling with disaggregate revealed preference data. *Journal of Transportation and Statistics*. 1999;2(2):149-158.
[28] De Jong GC, Vellay C, *Houée* M. A joint SP/RP model of freight shipments from the region Nord-Pas de Calais. *Proceedings of the AET European Transport Conference, 10-12 Sep. 2001, Cambridge, UK*. 2001. 15 p.
[29] Arunotayanun K. *Modelling freight supplier behavior and response*. Ph.D. thesis. Centre for Transport Studies, Imperial College; 2009.
[30] Martino A, et al. TRIMODE – Integrated transport model for Europe. *Proceedings of 7th Transport Research Arena TRA 2018, 16-19 Apr. 2018, Vienna, Austria*. 2018.
[31] Nugroho M, de Jong GC, Whiteing AE. Port and inland mode choice from the exporters and forwarders perspectives: Case study – Java, Indonesia. *Research in Transportation Business and Management*. 2016. DOI: 10.1016/j.rtbm.2016.03.010.
[32] Grene WH. *NLOGIT6 reference guide*; 2016.
[33] Hensher DA, Greene WH. Specification and estimation of nested logit model: Alternative normalisations. *Transportation Research Part B*. 2002;36(1):1-17. DOI: 10.1016/S0191-2615(00)00035-7.
[34] Bierlaire M. A theoretical analysis of the cross-nested logit model. *Annals of Operations Research*. 2006;144:287-300. DOI: 10.1007/s10479-006-0015-x.
[35] Silberhorn N, Boztug Y, Hildebrandt L. Estimation with the nested logit model: Specifications and software particularities. *OR Spectrum*. 2008;30(4):635-653. DOI: 10.1007/s00291-007-0109-0.
[36] Ortúzar JD, Willumsen LG. *Modelling transport*. 4th Ed. John Wiley & Sons; 2011. DOI: 10.1002/9781119993308.
[37] Koppelman FS, Wen CH. Alternative nested logit models: Structure, properties and estimation. *Transportation Research Part B: Methodological*. 1998;32(5):289-298. DOI: 10.1016/S0191-2615(98)00003-4.
[38] Tezcan HO. *Discrete choice modelling in transport*. Lecture notes (not published); 2019.
[39] Istanbul Metropolitan Municipality & Bimtaş. *Istanbul logistic master plan project*; 2018.
[40] De Jong GC, et al. Analysis of route and mode transport choice in Eastern South Asia following integration agreements. In: Dappe MH, Kunaka C. (eds) *Connecting to thrive: Challenges and opportunities of transport integration in Eastern South Asia*. World Bank, Washington, DC; 2021. DOI: 10.1596/978-1-4648-1635-2_ch2.

Berna Aksoy, Mustafa Gürsoy

Çeşitli Yük Türleri İçin Rastgele Fayda Teorisi ile Tutarlı Farklı Bir Nested Logit Model Yapısı: İstanbul İçin Bir Vaka Çalışması

Özet

Yük taşımacılığı şehir içi trafiğe önemli ölçüde katkıda bulunur ancak yolcu taşımacılığına kıyasla karar vericiler tarafından genellikle göz ardı edilir. Geleneksel ulaştırma modelleme çalışmaları, veri toplamanın zorluğu nedeniyle yük taşımacılığı için genellikle toplu yaklaşımlar kullanmaktadır. Ancak yük taşımacılığının doğası çok daha karmaşıktır. Bu nedenle yük taşıtı tercihlerinin belirleyicilerinin ayrık yaklaşımlarla incelenmesi, özellikle yerel çalışmalarda ortaya konulabilecek katkılar açısından büyük önem taşımaktadır. Yük taşımacılığına yönelik yerel ayrık yaklaşımların çalışılmasındaki bu belirgin boşluğu gidermek için, göndericiler ve taşıyıcılar için yük aracı tercihlerini belirleyen faktörleri araştırmak üzere ayrık modelleme tabanlı bir metodoloji sunuyoruz. Tahmin edilen nested logit model, rastgele fayda teorisinin ikinci kısmı olan RU2 yaklaşımı ile oluşturulmuştur, böylece jenerik katsayılar kullanıldığında ortaya çıkan teorik tutarsızlıklardan kaçınılmıştır. Sonuç olarak, model yapıları literatüre kıyasla tatmin edici sonuçlar vermiştir. Yük taşıtı seçimi tercihlerini etkileyen faktörlerin ambalaj tercihlerinden etkilendiği ve yük gruplarına göre farklılaştığı ortaya çıkmıştır. Bu yerel çalışma, İstanbul'da yük modellemesine yönelik ilk nested logit çalışması olup, ileride yapılacak ulusal çalışmalara ışık tutması hedeflenmektedir.

Anahtar Kelimeler

ulaştırma modellemesi; yük modellemesi; rastgele fayda teorisi; kesikli seçim; logit modeller; nested logit.