Transport
Volume 165 Issue TR1

Flexible model structure for discrete-choice models
Ishaq, Bekhor and Shiftan

ice | proceedings

Proceedings of the Institution of Civil Engineers
Transport 165 February 2012 Issue TR1
Pages 39–47 http://dx.doi.org/10.1680/tran.10.00001
Paper 1000001
Received 17/12/2009 Accepted 19/10/2010
Published online 06/09/2011
Keywords: mathematical modelling/transport planning

ice
Institution of Civil Engineers
publishing

Flexible model structure for discrete-choice models

Robert Ishaq PhD
Doctoral student, Department of Transportation and Geo-Information Engineering, Civil & Environmental Engineering Faculty, Technion – Israel Institute of Technology, Haifa, Israel

Shlomo Bekhor PhD
Associate Professor, Department of Transportation and Geo-Information Engineering, Civil & Environmental Engineering Faculty, Technion – Israel Institute of Technology, Haifa, Israel

Yoram Shiftan PhD
Associate Professor, Department of Transportation and Geo-Information Engineering, Civil & Environmental Engineering Faculty, Technion – Israel Institute of Technology, Haifa, Israel

Discrete choice models are widely used for travel demand analyses, and several such models were developed based on random utility theory. The estimation process in such models generally uses the entire study population and the same model structure to estimate average parameter values, but this approach limits the possiblility that the model can fully explain the behaviour of different populations. A new approach, the flexible model structure (FMS), expands existing discrete choice models with the addition of two main components: the segmentation process and model structure search. The aim of the FMS process is to guide the progress of model estimation towards a more behaviourally realistic representation. The main contribution of this paper is to suggest and illustrate a framework that simultaneously searches for the best segmentation and the best model structure for each segment. This paper presents a numerical case study that illustrates the FMS concept. The results indicate that (*a*) the addition of the segmentation process acts to emphasise the heterogeneity that exists among segments and demonstrates the importance of segmentation by significantly improving the estimation results; and (*b*) the model structure can vary along segments.

1. Introduction and literature review

Discrete choice models are usually derived from the assumption of ‘utility-maximisation’, originally developed in the field of psychology by Thurstone (1927). Models developed by this method are called random utility maximisation (RUM) models. Among the many RUM models are the standard multinomial logit model (MNL), which was first proposed by Luce (1959) and is the most widely used model; the probit model (Daganzo, 1979); the generalised extreme value (GEV) class of models (McFadden, 1978), which encompasses several models, such as nested logit (NL) (see Ben-Akiva and Lerman, 1985) and cross-nested logit (CNL) (see Bekhor *et al.*, 2007).

Discrete choice models are widely used for travel demand analyses. Most of these models deal with a single-choice dimension. In daily life, however, people are often required to make multi-dimensional travel choices simultaneously, and modelling such choices can be a very complex process. Moreover, the large number of modifications to preplanned activities (Doherty and Papinski, 2004), such as start and end time and location, significantly complicates the modelling process.

The new generation of travel demand forecasting models, the activity based models (ABM), tries to deal with multiple choices in an integrated framework; however, these models simplify the simultaneity of travel choices by grouping them into a hierarchical, multi-dimensional decision structure (Shiftan and Ben-Akiva, 2010). The hierarchical model structure for the ABM is usually proposed by the researcher according to a logical order and experience, and it is usually fixed; that is, the researcher tests few structures but assumes a single-choice hierarchical order even though there are a number of possible options (for example, theoretically there are 125 possible hierarchy orders if there are five choice dimensions). Given the significant level of effort in estimating such models, it is not practical to test many different possible structures. There is also no evidence that the logical choice hierarchy order proposed best fits a given data set.

In model estimation, the entire study population is normally used to estimate the fixed model structure proposed by the researcher. The inherent assumption of these models, that the same causal structure governs the behaviour of the entire population, may be incorrect, as human behaviour is known to exhibit considerable variation within and between behavioural units (Pendyala, 1998). For example, mode choice for some population groups may be driven by destination choice; for others, by such determinants as car ownership; or there may be no mutual influence between mode and destination choices as manifested in their different error structures.

In travel demand modelling, the segmentation process has been used mostly for two main purposes: (*a*) to divide the data, as in the practice of the ‘four-step model’, according to trip purpose, and therefore to implement the model for each segment separately (Hensher and Button, 2000; Ourtuzar and Willumsen, 1990); (*b*) to understand travellers’ attitudes towards mode choice and to try to characterise groups that are more likely to switch from car to public transportation (Shiftan *et al.*, 2008).

In the 'four-step model', the segmentation process is fairly simple. The data are divided according to one criterion, usually 'trip purpose'. However, there is no evidence in the literature that segmenting the data according to trip purpose provides maximum heterogeneity among segments and maximum homogeneity within segments. Performing segmentation according to other variables (e.g. trip distances) may achieve better estimation and prediction results.

Some researchers have used segmentation by attitude as shown, for example, by Proussaloglou and Koppelman (1989), Golob and Hensher (1998), Golob (2001), Lieberman *et al.* (2001) and Anable (2005). Shiftan *et al.* (2008) used structural equation modelling (SEM) to simultaneously identify travellers' attitudes, travel behaviour, and the causal relationship between a traveller's socio-economic profile and this traveller's attitudes towards travel. Attitudinal factor scores derived from SEM were used by Shiftan *et al.* (2008) to group the data into distinct market segments.

Efforts to improve the segmentation process in travel behaviour analysis were first made by Badoe and Miller (1998), who developed an analytical procedure that automatically identified segments by dealing simultaneously with level of service and socio-economic and spatial factors to determine the relative role that each plays in determining travel behaviour; and by Bhat (1997), who used an endogenous segmentation approach to model mode choice that jointly determined the number of market segments in the travel population, assigning individuals probabilistically to each segment and developing a distinct modal choice model for each group.

The integrated development of discrete choice models and the segmentation process is provided by the latent class models (LCM). Over the past decade, researchers have recognised that the use of LCM can yield powerful improvements in model explanatory power over traditional MNL and NL.

LCM is used to identify segments of respondents who tend to have similar preferences as manifested in choice-based conjoint (CBC) data. LCM classifies respondents into different segments and estimates the utility for each segment. In other words, the theory underlying LCM posits that individual behaviour depends not only on observable attributes but also on latent heterogeneity, which varies with factors that are unobserved by the analyst.

LCM shows significant improvement in the estimation results, in comparison with traditional MNL and NL, and succeeds in capturing taste variations, thus indicating its efficacy and practicability (see, for example, Greene and Hensher (2003); Hess (2009)). Nevertheless, the same model structure is implemented for all segments, and no attempt is made to test different model structures for different segments.

The purpose of this study was to illustrate the concept of flexible model structure (FMS), based on two main hypotheses: (*a*) different groups of people have different elasticities of demand; and (*b*) different model structures may best describe the behaviour of different groups of people in multi-dimensional choice models.

These hypotheses are not new, but they were generally treated in specific contexts in the literature. The contribution of the FMS approach presented in this paper is that it offers a systematic way to tackle a general multi-dimensional choice problem by simultaneously segmentating the population and suggesting the best model structure for each segment to improve model estimation. To the best of the authors' knowledge, such an approach has not been illustrated in the literature. This approach should assist developers of multi-dimensional choice problem models and specific activity-based models currently being developed in many metropolitan areas to improve model design and estimation.

The paper is organised as follows: the next section presents the FMS approach; this is followed by a case study that illustrates the concept. The last section presents the conclusions of this study and proposes future research.

2. Methodology

The modelling approach for this study was based on the concept behind an FMS and illustrates its implementation to improve the development and estimation of multi-dimension discrete choice models towards more behaviourally realistic representations and explanatory power.

The FMS is composed of two main components.

(*a*) Segmentation process: the data are divided into two or more segments, according to one variable or a combination of variables.
(*b*) Model structure: the best model structure is then found for each segment.

In the case of destination and mode choice, for example, and assuming that the data are segmented into two segments, three possible model structures are possible: (*a*) a multi-nomial model; (*b*) nested logit in which the destination choice appears at the higher level; (*c*) nested logit in which the mode choice appears at the higher level. In an FMS, a different model structure could provide the best model estimation results for each segment of the data. For simplicity's sake and in order to focus on the general idea of the FMS approach, this study was limited to MNL and NL structures and avoided testing other flexible structures, such as cross-nested logit and mixed logit models.

In general, the variables that were used to segment the data, in addition to the number of segments, were determined according to the data analysis, the researcher's experience, and a process of trial and error. FMS allows a more behaviourally realistic representation having improved explanatory power. The concept and structure of FMS in comparison with the existing approach is presented in Figure 1. First, the best segmentation for the data is found and then all possible predefined model structures are empirically tested for each segment. The final model is composed of the best model structure for each segment. This idea will be demonstrated by a numerical case study in the next section. As will be shown in the following sections, FMS provides better estimation and prediction results than do existing models.

3. Numerical case study

To illustrate the implemention of FMS, a numerical case study was carried out using a sample from the 2006 Haifa Household Travel Behavior Survey (Yefe Nof, 2006). Haifa is the largest city in the north of Israel and the third largest metropolitan area in Israel. For this case study, it was decided to focus on the mutual decision of trip destination and mode. Different mode and destination choice model structures were tested, first for the entire data set and subsequently for different segments. This will be described in the subsections that follow.

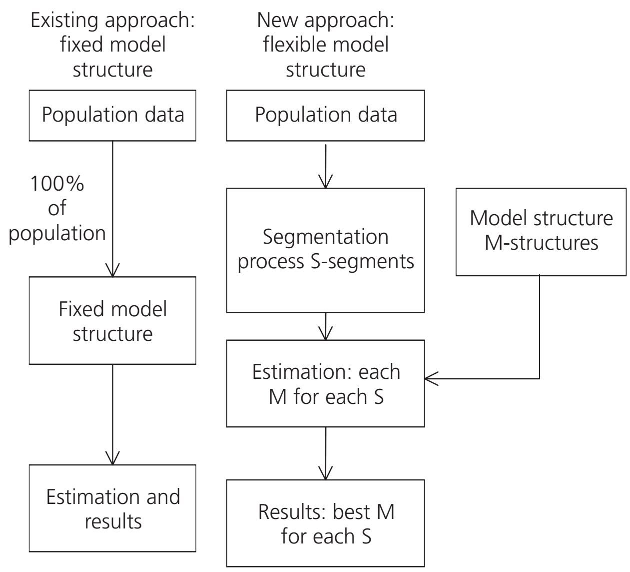

Figure 1. Comparison of the existing approach and the proposed new approach

3.1 Data description

The data used in the case study include 11 299 non-work trips, made by four modes: (*a*) car – as a driver; (*b*) car – as a passenger; (*c*) bus; and (*d*) walking. These trips took place in the inner-ring of the Haifa metropolitan area – which, for simplicity, was divided into 41 super zones. Each record of a trip includes personal characteristics (such as age, gender and household size), trip characteristics (such as mode choice, travel time, trip departure and arrival times), and zonal characteristics (population and employment).

Basic statistics of the data characteristics are presented in Table 1.

3.2 Model specification

The 41 destination choices and four mode choices yield a total of 164 alternative choices. The case study consists of two stages.

The first stage estimated three fixed-model structures using the entire data: MNL (Model 1); NL in which mode choice constitutes the higher level of the nest, and destination choice the lower (Model 2 M–D); and NL in which destination choice comes on the higher level, and mode choice on the lower (Model 3 D–M). As mentioned above, the focus of this study was on estimating different model structures within the MNL and NL framework. Obviously, other model structures, such as CNL or GNL, could also be tested.

The second stage divided the population into segments according to different variables, such as personal characterisations, income, and number of cars. In this case study, the segment variables and their values were chosen according to statistical method (using factor analysis), as in personal characterisations, and empirically according to the literature review and the researcher's experience.

Data characteristics – distribution of 11 299 trips					
Gender	Female	54%	Purpose	Errands	22%
	Male	46%		Leisure	20%
Licence	Yes	44%		Education	19%
	No	56%		Shopping	14%
Mode	Driver	37%		Other	25%
	Passenger	15%	Age	≤18	22%
	Bus	24%		19-59	54%
	Walk	24%		≥60	24%

Table 1. Data characteristics

Models for each segment were then estimated for the three different model structures: first, SMNL (segmentation multinomial logit) (Model 4); second, SNL_MD (segmentation nested logit) with mode choice at the higher and destination choice at the lower level (Model 5); third, SNL_DM with destination choice at the higher level and mode choice at the lower one (Model 6). For example, when the data were divided into three segments, nine models were estimated (three segments by three model structures).

The fourth model, which is the FMS (Model 7), is a combination of the best estimated model for each segment. That is to say, if there are two segments and the SNL mode-destination produced the best estimation results for the first segment and the SNL_DM produced the best results for the second segment, then the FMS would be a combination of SNL_MD and SNL_DM. Therefore, the first six models are all special cases of Model 7. The estimation process was performed using the BIOGEME software (Bierlaire, 2008). Figure 2 illustrates the stages of the experiment and the different model structures tested.

3.3 Estimation results

The estimation results for the first stage are presented in Table 2. In addition to the 22 personal and level-of-service variables in the MNL model, NL model 2 contains four nesting coefficients. NL model 3 contains the same 22 variables + 41 nesting coefficients. Full estimation results will be presented later only for the best model (see Table 6 later).

According to the 'adjust-rho-square' and 'Akaike information criterion (AIC)' measures, Model 2 – NL (mode-destination) and Model 3 – NL (destination-mode) showed very close performance, with a very small advantage to Model 3. These two models cannot be compared by the chi-squared test, as neither is a linear constraint of the other; however, the various goodness-of-fit measures could still be investigated.

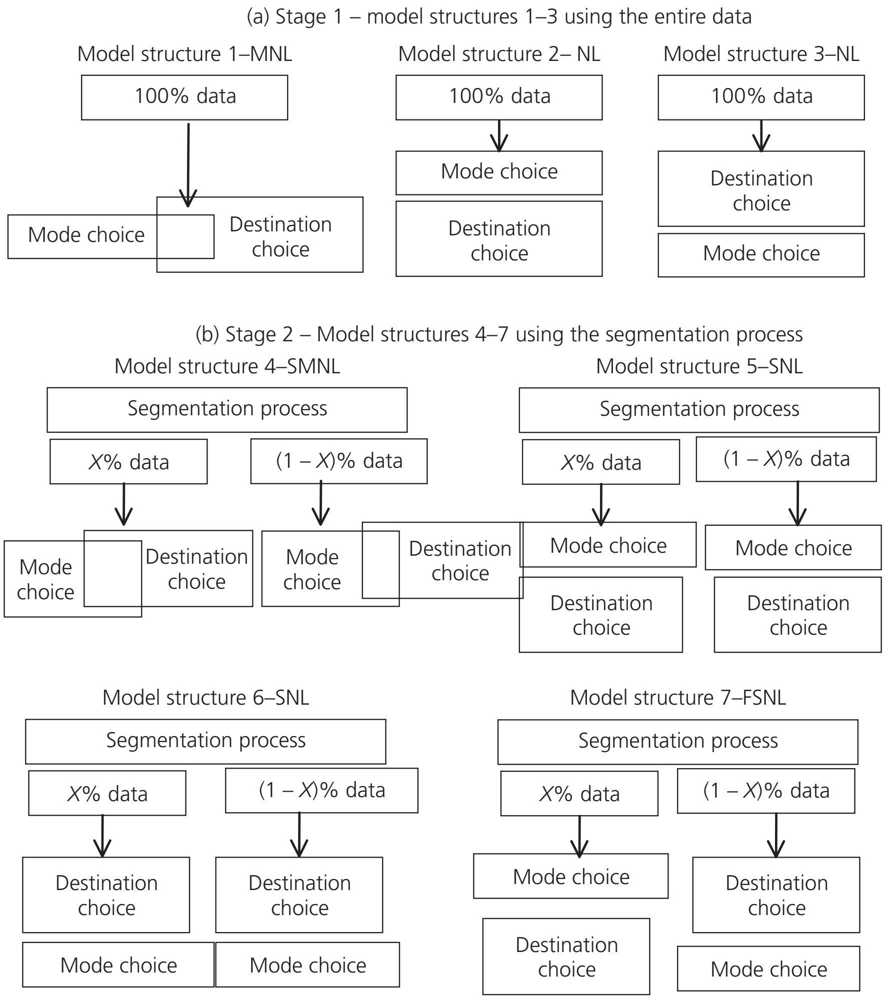

Figure 2. Experiment stage

No clustering	Estimated parameters	Initial likelihood	Final likelihood	Ad.-rho-sq.	AIC
Model 1–MNL	22	−56 192	−38 499	0·314	77 042
Model 2–NL (M–D)	26	−56 192	−38 029	0·323	76 110
Model 3–NL (D–M)	63	−56 192	−37 987	0·323	76 100

Table 2. Results of the first stage

The second stage was repeated for three different segmentation schemes: two segments for 'Personal characterisations' (age, number of children, gender, number of workers in household, driver's licence, household size), three segments for 'Income level', and three segments for 'Number of cars'. The segmentation process for 'Personal characterisations' uses the factor analysis technique. The representative characteristics for each segment are presented at the bottom of Table 3.

Table 3 presents the estimation results of the segmentation process according to the three segmentation schemes. Generally, segmentation improved the estimated results compared to the results shown in Table 2, despite the increase in the number of parameters. Among the three schemes, segmentation according to the 'Number of cars' variable yielded the best estimation results.

According to the chi-squared test (significance level of 5%), the nested logit structures in all segmentation schemes achieved better results than did the multinomial logit structure ($\chi^2_{5\%,(26-22)=4} = 9{\cdot}488$, $\chi^2_{5\%,(63-22)=41} = 56{\cdot}94$).

Table 4 compares the goodness-of-fit measures of Model 5 (SNL_MD) and Model 6 (SNL_DM) in order to search for the best model. The one with the highest adjust-rho-square value and the lowest AIC value was considered the best model fit.

For segmentation by 'Personal characteristics', Model 5–SNL_MD has a small advantage for segment 1 whereas Model 6–SNL_DM better fits segment 2.

For segmentation by 'Income level' and 'Number of cars',

Segmentation scheme	Segment	Description	No. of obs.	Init. likelihood	FL Model 4–SMNL (estimated parameters)	FL Model 5–SNL_M–D (estimated parameters)	FL Model 6–SNL_D–M (estimated parameters)
No clustering			11 299	−56 192	−38 499 (22)	−38 029 (26)	−37 987 (63)
Personal characterisations	Segment 1	See note*	5 472	−27 364	−18 428 (22)	−18 245 (26)	−18 205 (63)
	Segment 2	See note†	5 827	−28 828	−19 837 (22)	−19 565 (26)	−19 489 (63)
	Total		11 299	−56 192	−38 265 (44)	−37 810 (52)	−37 694 (126)
Income level	Segment 1	Below average	3 643	−18 257	−12 153 (22)	−12 048 (26)	−11 977 (63)
	Segment 2	Average	3 044	−15 166	−10 151 (22)	−10 066 (26)	−10 098 (63)
	Segment 3	Above average	4 612	−22 769	−15 518 (22)	−15 268 (26)	−15 290 (63)
	Total		11 299	−56 192	−37 892 (66)	−37 382 (78)	−37 365 (189)
Number of cars	Segment 1	0 cars	2 909	−14 196	−8 943 (22)	−8 745 (26)	−8 869 (63)
	Segment 2	1 car	5 867	−29 344	−19 744 (22)	−19 593 (26)	−19 453 (63)
	Segment 3	2+ cars	2 523	−12 652	−7 776 (22)	−7 666 (26)	−7 673 (63)
	Total		11 299	−56 192	−36 463 (66)	−36 004 (78)	−35 995 (189)

Init likelihood, ±56 192.
* Average age 60, no children, mixed gender, mixed driving licence, average household size 3.
† Average age 20, 1 child, mixed gender, mixed driving licence, average household size 4.
$\chi^2_{5\%,(26-22)=4} = 9{\cdot}488$, $\chi^2_{5\%,(63-22)=41} = 56{\cdot}94$

Table 3. Results of the second stage – final likelihood

Segmentation scheme	Segment	Description	Model 5–SNL (mode-destination)			Model 6–SNL (destination-mode)		
			Final likelihood	Adj.-rho-square	AIC	Final likelihood	Adj.-rho-square	AIC
Personal charac.	Segment 1	See note*	−18 245	0·332	36 538	−18 205	0·332	36 536
	Segment 2	See note†	−19 565	0·320	39 178	−19 489	0·322	39 104
	Total		−37 810	0·326	75 716	−37 694	0·327	75 640
Income level	Segment 1	Below average	−12 048	0·339	24 144	−11 977	0·341	24 080
	Segment 2	Average	−10 066	0·335	20 180	−10 098	0·330	20 322
	Segment 3	Above average	−15 268	0·328	30 584	−15 290	0·326	30 706
	Total		−37 382	0·333	74 908	−37 365	0·331	75 108
Model 7−FMS: final likelihood = −37 311 Adj.-rho-square = 0·334 AIC = 74 852								
Number of cars	Segment 1	0 cars	−8745	0·382	17 538	−8869	0·371	17 864
	Segment 2	1 car	−19 593	0·331	39 234	−19 453	0·335	39 032
	Segment 3	2+ cars	−7666	0·392	15 380	−7673	0·389	15 472
	Total		−36 004	0·356	72 152	−35 995	0·356	72 368
Model 7 − FMS: final likelihood = −35 864 Adj.-rho-square = 0·360 AIC = 71 958								

* Average age 60, no children, mixed gender, mixed driving licence, average household size 3.
† Average age 20, 1 child, mixed gender, mixed driving licence, average household size 4.

Table 4. Comparison between Model 5 and Model 6 for each segment

different model structures best fit different segments. For example, for the 'Income level' segmentation scheme, Model 6−SNL_DM best fits segment 1 (below average), whereas Model 5−SNL_MD best fits segments 2 and 3 (average and above average). The results show interesting travel behaviour implications where different nesting structures according to income level show that individuals in households that are identified as low income have more variety in choosing their mode, as there might be competition for the car within the household; therefore, destination choice appears higher in the hierarchy. On the other hand, individuals in households identified as average and high income (segment 2 and 3) are more likely to use their cars, and therefore mode choice appears higher in the hierarchy.

The same phenomenon occurs for the 'Number of cars' segmentation scheme: Model 5−SNL_MD best fits segments 1 and 3 (no cars and 2+ cars in a household), whereas Model 6−SNL_DM best fits segment 2 (1 car in a household). The 'Number of cars' segmentation scheme achieves the best estimation results among the three schemes. Here, as well, these results show interesting behavioural implications where individuals in households that do not own cars (segment 1) are less flexible to choose their mode, and therefore mode choice appears higher in the hierarchy; in contrast, individuals in households with only one car (segment 2) have more variety in choosing their mode, as there may be competition for the car within the household, with the result that destination choice appears higher in the hierarchy. Individuals in households that own two or more cars (segment 3) are 'captive' to their cars, and therefore mode choice appears, again, higher in the hierarchy.

As mentioned above, FMS (Model 7) is a combination of the best estimated models for each segment. Therefore, the FMS for the 'Income level' segmentation scheme consists of Model 6−SNL_DM for segment 1 and Model 5−SNL_MD for segments 2 and 3. Similarly the FMS for the segemanation by 'Number of cars' consists of Model 5−SNL_MD for segments 1 and 3 and Model 6−SNL_DM for segment 2. As shown in Table 4, the FMS model yields the best performance in terms of the adjust-rho-square and AIC measures. The full FMS model specification for segmentation by 'Number of cars', which was the best model obtained here, is presented later in Table 6.

In order to illustrate the heterogeneity of segments, Table 5 presents the marginal rate of subtitution as reflected in the coefficient ratios of the three 'Number of cars' segments of Model 6−SNL_DM. It is remarkable that the magnitude and sign of the coefficient ratio differ significantly (at the 5% level of significance) among the three segments of the same model structure. This fact emphasises the heterogeneity of the segments and demonstrates the importance of segmentation.

Time coefficient	Segment 1	Segment 2	Segment 3
cbd_car/time_car	2·24	−18·90	−21·22
crd_car/time_car	−5·73	−15·88	−17·07
male_car/time_car	−3·48	−5·27	0·76
time_passenger/time_car	0·15	0·37	0·38

cbd, central business district; crd, central residence district.

Table 5. Comparison of coefficients ratios in three segments for 'Number of cars' segmentation scheme (Model 6−SNL)

	Segment 1 NL_MD	Segment 2 NL_DM	Segment 3 NL_MD
Constant_Driver	−3·80 (*−6·0*)	0·16 (*<u>1·7</u>*)	1·64 (*8·3*)
Constant_Passenger	−3·43 (*−15·63*)	−0·61 (*−6·1*)	0·338 (*<u>1·5</u>*)
Constant_Bus	−1·20 (*−11·8*)	−0·48 (*−4·0*)	−0·0598 (*<u>−0·2</u>*)
Constant_Walk	0 fixed	0 fixed	0 fixed
Under age 18_Driver	−3·79 (*<u>−0·5</u>*)	−3·60 (*−7·5*)	−2·56 (*−6·9*)
Under age 18_ Passenger	−0·71 (*−3·0*)	−1·43 (*−13·2*)	−0·01 (*−0·05*)
Under age 18_ Bus	−1·20 (*−10·3*)	−1·49 (*−11·5*)	−1·04 (*−4·4*)
CBD_Driver	−0·87 (*<u>−0·8</u>*)	3·44 (*24·3*)	3·99 (*14·2*)
CBD_ Passenger	1·55 (*6·3*)	3·05 (*18·2*)	2·77 (*8·7*)
CBD_ Bus	1·46 (*17·2*)	3·24 (*21·0*)	2·77 (*8·1*)
CRD_Driver	2·24 (*4·2*)	2·89 (*30·1*)	3·21 (*15·4*)
CRD_ Passenger	1·69 (*7·3*)	2·23 (*19·5*)	1·39 (*6·8*)
CRD_ Bus	1·73 (*23·3*)	2·45 (*21·9*)	1·65 (*6·4*)
Male_Driver	1·36 (*2·6*)	0·96 (*10·0*)	−0·14 (*<u>−0·7</u>*)
Male_Passenger	−0·47 (*−2·4*)	−0·16 (*<u>−1·6</u>*)	−0·83 (*−5·1*)
Male_Bus	−0·41 (*−4·5*)	−0·08 (*<u>−0·7</u>*)	−0·08 (*<u>−0·4</u>*)
Employee_Driver	0·17 (*<u>0·37</u>*)	0·80 (*7·2*)	1·99 (*5·9*)
Employee_Passenger	−0·71 (*−2·5*)	0·31 (*2·5*)	1·11 (*2·9*)
Employee_Bus	0·77 (*7·0*)	0·66 (*5·2*)	1·23 (*3·2*)
Travel time _Driver	−0·39 (*−9·8*)	−0·179 (*−53·4*)	−0·188 (*−42·1*)
Travel time _Passenger	−0·26 (*−18·5*)	−0·200 (*−40·8*)	−0·247 (*−29·7*)
Travel time _Bus	−0·061 (*−40·0*)	−0·067 (*−35·2*)	−0·071 (*−21·3*)
Travel time _Walk	−0·052 (*−43·9*)	−0·029 (*−52·8*)	−0·034 (*−25·5*)
Nest_Driver	1·00 (*78·8*)	41 nest coefficients	0·91 (*53·4*)
Nest_Passenger	0·95 (35·0)		0·81 (*37·4*)
Nest_Bus	0·81 (63·6)		0·86 (*33·6*)
Nest_Walk	0·54 (*29·2*)		0·38 (*12·0*)

VALUE (<u>*t*-test</u>) – statistically insignificant at level of 5%.

Table 6. FMS – models specification, 'Number of cars' segmentation scheme

4. Conclusions and future research

The following conclusions highlight the importance of the FSM approach.

(*a*) The segmentation process improves the estimation results significantly. This can be seen by comparing the results of the Final likelihood in Table 2 (no segmentation) and in Table 3 (with segmentation).

(*b*) The segmentation process emphasises the heterogeneity of segments and demonstrates the importance of classification. It is remarkable that the coefficient values differ significantly (at the 5% level of significance) in magnitude among segments.

(*c*) The model structure can vary among segments. Tables 4 and 6 show that a different model structure is suitable for some segments than for others with the same data. This explains the travel behaviour of the different segments. For individuals in households with only one car, the destination choice appears to be higher in the nest hirearchy than is mode choice for individuals who do not own cars; for a household that owns two or more cars, the nesting order appears to be reversed.

(*d*) In the present case, the improvement in model fit was achieved mainly by the segmentation process and less by different model structures. It should be noted that the focus here was on testing different MNL and NL structures for different segements of the population. More flexible model structurces, such as mixed logit, were not tested.

In summary, the numerical case study illustrates the importance of integrating both components, the segmentation process and the model structure fit, in the estimation of complex discrete choice models. Although each of these components is used to some extent in discrete choice model estimation, this study showed that an integrated approach that simultaneously searches for the best segmentation scheme and for the best model structure for each segement for multi-dimensional choice models can significanly improved such models and their ability to explain behaviour of differetnt segments of the population.

Future work will include the development of a more general integrated methodological framework for discrete choice models. Such a framework will include an optimisation algorithm to search for the best segmentation scheme and model structure. This means that the framework should recommend the number of segments and point to the best combination of variables that segments the data to obtain homogeneous segments. Moreover, it should adjust the best model structure for each segment.

Acknowledgement
The authors thank the 'Yefe Nof Company' for providing their data for the work presented in this paper.

REFERENCES

Anable J (2005) Complacent car addicts or aspiring environmentalists? Identifying travel behaviour segments using attitude theory. *Transport Policy* **12(1)**: 65–78.

Badoe DA and Miller EJ (1998) An automated segmentation procedure for studying variations in mode choice behavior. *Journal of Advanced Transportation* **32(2)**: 190–215.

Bekhor S, Reznikova L and Toledo T (2007) Application of cross-nested logit route choice model in stochastic user equilibrium traffic assignment. *Transportation Research Record* **2003**: 41–49.

Ben-Akiva M and Lerman S (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press, Cambridge, MA, USA.

Bhat CR (1997) An endogenous segmentation mode choice model with an application to intercity travel. *Transportation Science* **31(1)**: 34–48.

Bierlaire M (2008) *An introduction to BIOGEME Version 1·6*. Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. See http://biogeme.epfl.ch (accessed 12/08/2011).

Daganzo C (1979) *Multinomial Probit: The Theory and Its Applications to Demand Forcasting*. Academic Press, New York, NY, USA.

Doherty S and Papinski D (2004) Is it possible to automatically trace activity scheduling decisions? *Conference on Progress in Activity-Based Analysis, Vaeshartelt Castle, Maastricht, The Netherlands*.

Golob TF (2001) Joint models of attitudes and behavior in evaluation of the San Diego I-15 congestion pricing project. *Transportation Research, Part A: Policy and Practice* **35(6)**: 495–514.

Golob T and Hensher D (1998) Greenhouse gas emissions and Australian commuters' attitudes and behavior concerning abatement policies and personal involvement. *Transportation Research, Part D* **3(1)**: 1–18.

Greene W and Hensher D (2003) A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodology* **37(8)**: 681–698.

Hensher D and Button K (2000) *Handbook of Transport Modelling*. Elsevier Science, Oxford, UK.

Hess S (2009) Advantages of latent class choice models over continuous mixed logit models. *Proceedings of the 12th International Conference on Travel Behaviour Research, Jaipur, India*. International Association for Travel Behaviour Research, Washington, DC, USA.

Lieberman W, Schumacher D, Hoffman A and Wornum C (2001) Creating a new century of transit opportunity: Strategic planning for transit. *Transportation Research Record* **1747**: 60–67.

Luce D (1959) *Choice Behavior: A Theoretical Analysis*. Wiley, New York, NY, USA.

McFadden D (1978) Modeling the choice of residential location. In *The Economics of Housing*. Edward Elgar, London, UK, vol. I, pp. 531–552.

Ourtuzar J de D and Willumsen LG (1990) *Modelling Transport*. Wiley, New York, USA.

Pendyala R (1998) Causal analysis in travel behaviour research: a cautionary note. In *Travel Behaviour Research: Updating the State of Play* (Ortuzar JD, Hensher D and Jara-Diaz S (eds)). Elsevier Science, New York, NY, USA, pp. 35–48.

Proussaloglou KK and Koppelman FS (1989) Use of travelers' attitudes in rail service design. *Transportation Research Record* **1221**: 42–50.

Shiftan Y and Ben-Akiva M (2010) A practical policy-sensitive, activity-based, travel-demand mode. *The Annals of Regional*

Science, http://dx.doi.org/10.1007/s00168-010-0393-5.

Shiftan Y, Outwater ML and Zhou Y (2008) Transit market research using structural equation modeling and attitudinal market segmentation. *Transport Policy* **15(3)**: 186–195.

Thurstone L (1927) A law of comparative judgement. *Psychological Review* **34(4)**: 273–286.

Yefe Nof (2006) *Travel Behaviour Survey for Haifa Metropolitan Area, Report, 2006*. Yefe Nof, Haifa, Israel.