Transportmetrica, Vol. 4, No. 2 (2008), 83-96

USING CHOICE SETS FOR ESTIMATION AND PREDICTION IN ROUTE CHOICE

ROB VAN NES[1], SASCHA HOOGENDOORN-LANSER[2] AND FRANK S. KOPPELMAN[3]

Received 11 January 2007; received in revised form 20 August 2007; accepted 10 November 2007

This article investigates the relationships between choice set types and analysis purpose, especially estimation and prediction. The choice set types considered are generated objective choice sets and observed subjective choice sets. While for estimation purposes subjective choice sets might be preferred, although objective choice sets might have benefits as well, objective choice sets appear to be most suitable for prediction purposes. Empirical analysis of choice sets for multi-modal inter-urban train trips shows clear distinction between choice models estimated using objective choice sets and using subjective choice sets. Applying these choice models for prediction shows that choice models based on subjective choice sets have a poorer performance when applied to objective sets than vice versa. As a result it is recommended to use objective choice sets for both estimation and prediction purposes.

KEYWORDS: Choice sets, route choice, estimation, prediction, empirical analysis

1. INTRODUCTION

Choice sets are widely used in the analysis of travel behaviour, either for analysing choice behaviour, estimating parameters of choice models, or predicting choice probabilities. Defining choice sets, however, is not a trivial task. For mode choice analysis for instance it is common to define a choice set consisting only of public transport, car and slow modes, although each of these can be decomposed into more detailed alternatives. Further, when combinations of modes are considered for line haul, access and egress, the definition of the choice set becomes more complicated. This is even truer in the case of route choice analysis. Depending on trip type and network structure many possible alternatives might be distinguished, while travellers are likely to consider only a limited set of alternatives. Selection of a set of reasonable alternatives might be straightforward for an individual traveller, but it is difficult to formulate an algorithm that approximates the varying choice set decision rules among travellers. However, it is well known that the size and composition of choice sets that are used will influence the results of model estimation and prediction. Furthermore, as estimated choice models are used for prediction the purposes of estimation and prediction are inextricably linked. Therefore, defining proper choice sets requires serious consideration.

This article discusses the relationships between choice set types and analysis purposes in order to derive guidelines for choice set usage in estimation and prediction. First, a classification of choice set types is presented, followed by a discussion on choice set types that are likely to be available for estimation and prediction purposes. Next, an extensive database on multi-modal inter-urban train trips is used to estimate parameters for different types of choice sets. This is followed by an analysis of the performance of the estimated choice models in the case of prediction, with special attention to the situation in which choice models are applied to choice set types different from those with

[1] Transport & Planning, Delft University of Technology, The Netherlands. Corresponding author (E-mail: R.vanNes@tudelft.nl).
[2] Centre for Transport and Navigation, DG Public Works and Watermanagement, The Netherlands.
[3] Department of Civil and Environmental Engineering, Northwestern University, USA.

which they were estimated. The final section describes the conclusions that can be derived from this analysis.

2. CHOICE SET CLASSIFICATION

Choice sets might be classified from the perspective of the traveller or the researcher. Hoogendoorn-Lanser (2005) presents a scheme based on alternatives that are logical and are feasible on the one hand and that are known or are considered by the traveller on the other hand (see also Fiorenzo-Catalano et al., 2003; Hoogendoorn-Lanser and van Nes, 2004). The first group of characteristics might be defined in an objective way, while the second group clearly depends on the traveller himself. A third distinction can be made on the way the choice set has been derived: by the traveller (i.e. actual behaviour), observed by a researcher, or generated by a researcher.

In this article we take the perspective of the researcher who studies travel behaviour defining choice sets by explicit criteria as shown in Figure 1. For clarity we will limit the number of choice set types to be considered in this analysis to two:

1. Generated objective choice sets: choice sets generated by the researcher that consist of alternatives that are logical, i.e. do not contain loops, and are feasible for the traveller based only on objective criteria such as vehicle availability;
2. Observed subjective choice sets: alternatives that are reported by travellers, these alternatives are thus known and feasible.

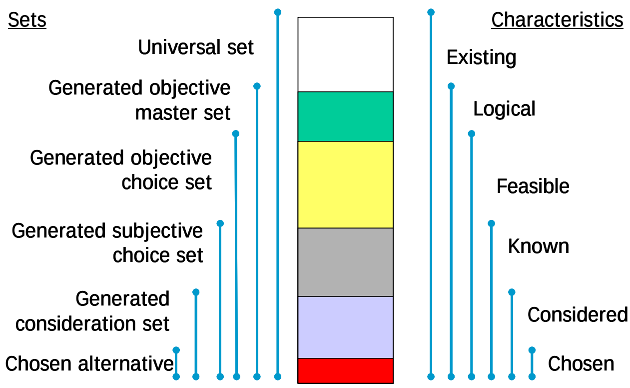

FIGURE 1: Relationships among existing, logical, available, feasible, known, preferred and chosen trip alternatives from the researcher's perspective

Alternatives that are not feasible, but are included in the generated objective master set, are considered not to be relevant. The distinction between consideration set and subjective choice set is difficult to establish. Therefore, all alternatives reported by the traveller are assumed to part of the subjective choice set. Generated objective choice sets are likely to be much larger than observed subjective choice sets. Please note that in this terminology a generated objective choice set should contain (nearly) all logical and feasible alternatives. Furthermore, it is assumed that the subjective choice set as well as the chosen alternative is part of the generated objective choice set.

3. CHOICE SET TYPE AND ANALYSIS PURPOSE

Choice sets may be used for three purposes:

1. Analysis of travel behaviour;
2. Estimation of parameters in choice models;
3. Prediction of probabilities of alternatives.

In the first case, the researcher has full knowledge of the characteristics of the choice set type he is using. Specific characteristics or even limitations thus might be accounted for in the analysis. The other two cases require more attention.

3.1. Estimating parameters

For estimation, the choice set type strongly influences the estimated parameters as well as the formulation of the choice model itself (e.g. Van der Waerden et al., 2004; Train, 2002; Hoogendoorn-Lanser, 2005). Researchers might generate objective choice sets themselves or might determine which alternatives are known by travellers (subjective choice sets). The differences between the two types might be manifold. To illustrate this difference the conceptual model for travel choice analysis presented by Bovy and Stern (1990) can be used. Due to information limitations, travellers are only aware of a subset of all alternatives that might be possible. Based on specific attributes they may eliminate alternatives, either based on preferences or on feasibility. The remaining alternatives might be ranked using a kind of trade-off between attributes. Finally a certain decision rule might be applied to determine the optimal alternative. This conceptual model differs somewhat from the notions presented in the previous section, however, it shows that many mechanisms and many criteria are possible to reduce the objective set to a smaller choice set. A subjective choice set could thus be a specific sample of the objective choice set.

Random Utility Models are commonly used to study travel behaviour. A typical characteristic of RUM-models is that they assume compensatory behaviour: alternatives which score badly for one attribute might still be attractive due to high scores for other attributes. In the framework of Bovy and Stern this concept is only applicable in the last step. The choice set type that best resembles this situation is the subjective choice set. This implies that for the application of RUM-models it would be best to use subjective choice sets. Using objective choice sets would mean that the choice model also would have to account other choice related mechanisms such as elimination based on specific preferences or feasibility. The choice model thus confounds the actual choice process with the determination whether the alternative is considered or not.

Consequence of this line of reasoning is that subjective choice sets should be derived, either by observation or using a generation method. Observing subjective choice sets is already quite complicated. Asking travellers for possible alternatives for a certain choice might easily invite them to come with more alternatives than they actually considered. Hoogendoorn-Lanser (2005) suggests making a distinction between alternatives that are directly mentioned and alternatives that come up after a moment of thought. Generating a subjective choice set is even more difficult to do. Beforehand, it is unclear which criteria should be applied to eliminate non-relevant alternatives from a set of possible alternatives. At the level of individuals the choice for any criterion would be arbitrary. This is important since the composition of the choice set influences the parameter estimates. Choices for generating subjective choice sets might thus directly affect parameter estimates. A possible option to deal with this problem might be to model a two-stage process, that is a model to determine whether an alternative is relevant or not, and a model to determine the actual choice from the relevant alternatives (see e.g. Manski, 1977; Cascetta and Papola, 2001; Swait, 2001a,b). However, it is questionable to which extent these methods are capable of dealing with large choice sets as might be expected for route choice problems.

Apart from this theoretical discussion the use of objective or subjective choice sets has another effect that should be considered. As mentioned earlier, objective choice sets are likely to be large and to have a large variety in attributes and attribute values, while subjective choice sets tend to be small, thus having less variation. Due to this larger variation, choice models based on objective choice sets can be more elaborate in number of attributes and in model structure, while the parameter estimates will be more significant. Higher quality models suggest a better description of travel behaviour. From this perspective, the use of objective choice sets should be preferred. A practical problem might be that objective choice sets are too large for estimation software. In that case it is possible to use a representative sample of alternatives (Ben-Akiva and Lerman, 1985; Train, 2002), e.g. a maximum of 50 alternatives per respondent.

Obviously there is a dilemma. On the one hand it can be argued that given the variety of decision rules travellers may apply to limit their choice set, subjective choice sets are most appropriate for estimating choice models. On the other hand, objective choice sets are likely to yield higher quality choice models, with respect to model structure and parameter estimates. Key point is which approach is more likely to reflect actual travel behaviour.

3.2. Predicting probabilities

Estimated choice models are used for predicting probabilities of alternatives, for instance for forecasts. In that case no subjective choice sets are available. Generation of subjective choice sets suffer from the same problem as mentioned in the previous section: it is unclear which criteria should be used to determine the relevant alternatives. Using objective choice sets might offer computational problems, so usually approximations of objective choice sets are used. Such a choice set should have sufficient variety and include alternatives that are relevant for the purpose of the study (see Fiorenzo-Catalano (2007) for a discussion of criteria for choice sets).

A second reason for applying objective choice sets for prediction purposes is that this type of application usually considers trips between zones, while estimation generally focuses on individual trips. As a result, choice sets used for prediction should be much larger to account for the differences in spatial and individual characteristics. Objective choice sets are more suited for this.

If objective choice sets, or approximations of them, are used for prediction it would be best to use choice models that are based on objective choice sets as well. In that way the consistency between estimation and prediction is guaranteed.

3.3 Consequences

The discussion on the relationship between choice set type and modelling purpose suggests that from a theoretical perspective subjective choice sets should be preferred for estimation purposes, while objective choice sets are to be preferred for prediction purposes. On the other hand, however, it appears that objective choice sets have the benefit that they allow more advanced modelling and might yield higher quality parameters. This obvious dilemma on choice set usage for estimation and prediction purposes thus yields four possible scenarios as can be seen in Table 1.

TABLE 1: Possible scenarios of using choice sets for estimating choice models and their application for prediction

		Choice models following from estimation using	
		Subjective choice sets	Objective choice sets
Prediction using	Subjective choice sets	This case appears to be sensible. Key point is how to determine the subjective choice set, especially for prediction.	The first question is the availability of the subjective choice set for prediction. The second issue is whether the probability of the non chosen alternatives will be underestimated.
	Objective choice sets	In this case the point is whether the non-chosen alternatives will be overestimated.	This appears to be a sensible combination.

In the following sections these scenarios will be explored for an actual database on multi-modal inter-urban train trips. Choice models will be estimated using objective and subjective choice sets. The resulting choice models will then be analysed on their performance with both objective and subjective choice sets. The goal of the analysis is to determine whether it should be recommended to use objective choice sets for estimation as well as for prediction purposes. First, however, the dataset itself will be described.

4. DATASET

For the analyses multi-modal travel data are used from a survey conducted among train travellers in an urbanized corridor in The Netherlands, including the cities Dordrecht, Rotterdam, The Hague, and Leiden (Hoogendoorn-Lanser, 2005). The survey focused on the multi-modal trip itself (which modes were used and in which order, what were the transfer nodes, what were the boarding and alighting nodes) and on train-based trip alternatives known by the traveller, i.e. the observed subjective choice set. The trips are described at leg level (part of a trip in which a single mode is used) including differences within transport services, e.g. different bus lines. Modes that are distinguished are walking, cycling, car, bus, tram, metro, local train, express train, and intercity train services. Physical differences in for instance bicycle routes to or from railway stations are not included. The survey data was extended with detailed data on all trip components, such as in-vehicle times per mode and costs, as well as with similar data for all other reasonable non-reported alternatives for the same trip (generated objective choice set). The considered sample contains 708 home-based trips, most of them having 3 legs. All trips correspond to different respondents and different OD-pairs. Travellers’ home and activity addresses are located throughout the research area. 80 % of the trips have a trip length between 10 and 30 kilometres. The average travel time is 46 minutes. Commuting (44%) and education (29%) are the main trip purposes.

From these respondents only 189 respondents reported alternatives other than the chosen alternative. 50 % of these respondents have a choice set of 2 alternatives, while the average size is 2.8. The maximum choice set size is 7 alternatives. Analysis of the composition of the subjective choice sets showed that most alternatives only differ in a single leg from the chosen alternative, especially the home-end leg (between home-address and boarding railway station) and secondly for the activity-end leg (alighting railway station and activity-address) (see Hoogendoorn-Lanser and van Nes, 2004). The train part appears to be quite constant within each subjective choice set. Only 20 % of the alternatives differ for 2 legs or more from the chosen alternative.

For all respondents objective choice sets were generated using a new deterministic, run-based, selective enumeration method using branch and bound techniques, which is applicable for generating choice sets for individual travellers taking into account traveller and trip characteristics (Hoogendoorn-Lanser et al., 2007). Key characteristic of the algorithm is a set of constraints that reflects observed travel behaviour. All feasible route alternatives between an origin and destination that satisfy such a set of constraints are generated. A subset of these constraints is applied to limit the generation process within feasibility bounds of space and time, while another subset is applied to restrict the generation to plausible routes. An advantage of this approach is its comprehensiveness: within the given bounds it generates all feasible paths. Another advantage is its deterministic property; the generation is perfectly reproducible in contrast to so-called stochastic generation approaches, where the outcomes such as size and composition of choice sets are random variables (see Fiorenzo-Catalano, 2007). Since the search process is exhaustive, the full choice set is known. It proved to be possible to generate objective choice sets that include the chosen alternative as well as the reported alternatives. Only in a limited number of cases it was necessary to relax some constraints, e.g. with respect to the maximum distance to the nearest railway station for cycling.

The generated objective choice sets are significantly larger than the reported subjective choice sets. The average choice set size is 48 alternatives (median is 39). For one respondent a choice set size of 278 alternatives was found. The alternatives in the choice set differ for all components of the multi-modal train trip. As a result the overlap between alternatives is large. More than 90 % of the alternatives differ for 2 legs or more from the chosen alternative. Furthermore, the choice set also includes alternatives having more legs than the chosen route. For computational reasons the choice sets used in the analysis were limited to a random sample of 50 alternatives from the generated alternatives at most.

5. ESTIMATION

For both choice set types, generated objective choice set and observed subjective choice set, random utility models are estimated. The following cases are considered:

- Objective choice sets (maximum of 50 alternatives): all respondents (N=708, Model L-O) and the subset of respondents reporting subjective choice sets (N=189, Model S-O)
- Subjective choice sets: only respondents reporting subjective choice sets (N=189, Model S-S)

Since both cases differ with respect to number of respondents and number of alternatives, additional analyses are performed to determine how these differences affect the estimation results.

Route overlap is an important aspect that should be accounted for when estimating route choice models. Estimation results for objective choice sets showed that accounting for overlap by inclusion of a path size-variable based on the number of legs yields better models than the MNL-model (Hoogendoorn-Lanser, 2005; Hoogendoorn-Lanser et al., 2005). Generalized nested logit models lead to even better results. However, subjective choice sets are too small for this type of advanced choice models. Comparing the results of an advanced model for objective choice sets with a MNL-model for subjective choice sets would put the subjective choice sets at a disadvantage from the start. The simpler alternative of including a path-size variable based on the number of legs might also confuse the analysis. The path-size parameter appeared to be only significant in the case of 708 respondents considering objective choice sets (Model L-O) and not in the case of subjective choice sets (Model S-S). Obviously, in the case of subjective choice sets, many choice sets consist of two alternatives, in which case the path size variable based on the number of legs would be identical for both alternatives thus making the inclusion of a path-size-variable meaningless. Thus, for an honest comparison only MNL-results are presented.

5.1 Using objective choice sets

The best MNL-model established for the case of 708 respondents using objective choice sets of 50 alternatives at maximum makes a distinction between home-end part, train part, and activity-end part of trips. In this way the difference in availability of modes and in knowledge of the transport system between home-ends and activity-ends can explicitly be accounted for (see Hoogendoorn-Lanser et al., 2006). The best MNL-model consists of 25 parameters describing mode-specific constants for home-end and activity-end, and parameters for in-vehicle times by mode, train types, railway station types, costs for UPT and parking, UPT headway, walking time to UPT-stops, and number of high- and low-frequency legs. All parameter estimates appeared to be highly significant (see Appendix)

Since subjective choice sets are only available for 189 respondents, a second choice model using objective choice sets is estimated for this smaller dataset (Model S-O). The choice model is slightly simplified and consists of 20 parameters (see Appendix). In both choice models, the parameters have been scaled using the train in-vehicle time, which allows for a direct comparison of the results. The question is which way these parameter estimates differ from each other. Figure 2 shows that the parameters of the second choice model are in line with those of the first: the R^2 is 0.93 (parameters that are unique for Model L-O are excluded from the comparison). In general the parameter estimates for the smaller dataset are 9% higher. The higher parameters estimates might suggest that the 189 respondents might have a larger variety in their choice sets than average. No clear outliers can be distinguished. The number of respondents thus appears to have a limited impact on estimation results.

5.2 Using subjective choice sets

The choice model specification derived with objective choice sets could not be estimated using subjective choice sets. Due to the smaller number of respondents and the differences in choice sets (size and composition), the best model that could be estimated contained only 12 parameters, while not all parameters are significant at 95% confidence level (see Appendix) (Model S-S). No parameter estimates for train stations could be established, which is due to the limited variation in boarding and alighting stations in the subjective choice sets. Furthermore, there appears to be less difference between home-end and activity-end parts of trips.

Figure 3 presents a graphical comparison, which shows that parameter estimates using subjective choice sets are generally 50% smaller than found for the objective choice set. The R^2 is still reasonable: 0.81. If the two largest outliers are excluded (parameter estimates for the number of legs) the correlation between the parameter estimates is more than 0.95. This implies that that there is a systematic difference in the size of the parameter estimates between both choice models. It is thus expected that the resulting choice models is less distinctive with respect to route choice probabilities in the case of prediction. Comparison with the results of Model S-O yields similar results.

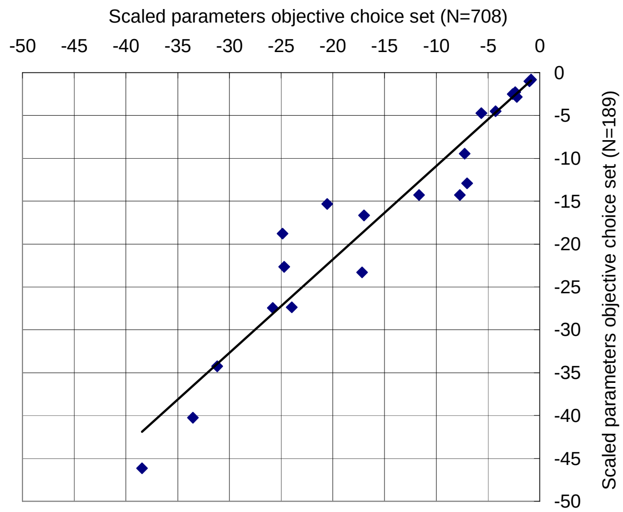

FIGURE 2: Relationships between scaled parameters for choice models estimated on objective choice sets for all respondents (N=708) and for respondents reporting subjective choice sets (N=189) (Model L-O and Model S-O respectively)

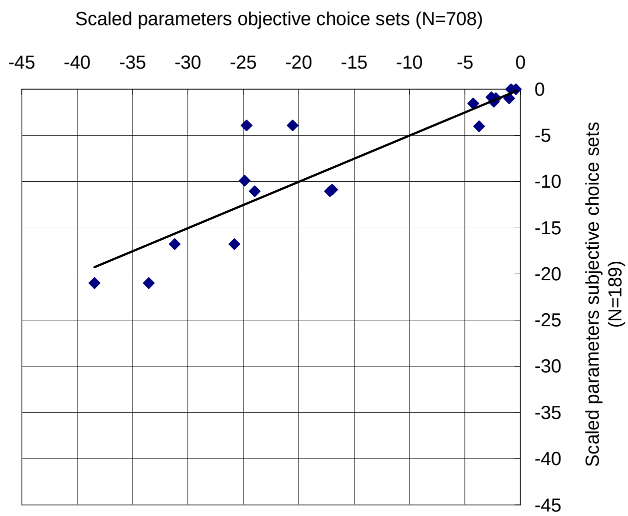

FIGURE 3: Relationships between scaled parameters for choice models estimated on objective choice sets (N=708, Model L-O) and on subjective choice sets (N=189, Model S-S)

5.3 Impact of number of respondents and choice set size

Given the differences in number of respondents and in choice set size, two additional analyses have been performed:

- Impact of a smaller number of respondents;
- Impact of choice set characteristics (size and composition).

The first analysis focuses on the number of respondents. The question then is whether the respondents reporting subjective alternatives are a representative sample of the full set of 708 respondents. To this end 50 random samples of 189 respondents have been drawn from the population of 708 respondents. For each sample the parameter estimates of choice model S-O are estimated. Subsequently average values and standard deviations are derived. Comparison shows that all parameter estimates of Model S-O (189 respondents considering objective choice sets) are within the 95% confidence interval. The correlation between the parameter estimates is 0.96 (see Figure 4). The parameter estimates of Model S-O, however, are on average 18% smaller. Since the estimates of Model S-O were generally higher than those of Model L-O, the 18% difference suggests that the differences between Model L-O and Model S-O are not due to specific characteristics of the choice sets of respondents reporting subjective alternatives. To conclude, these findings indicate that the respondents reporting subjective alternatives are a representative sample of all respondents.

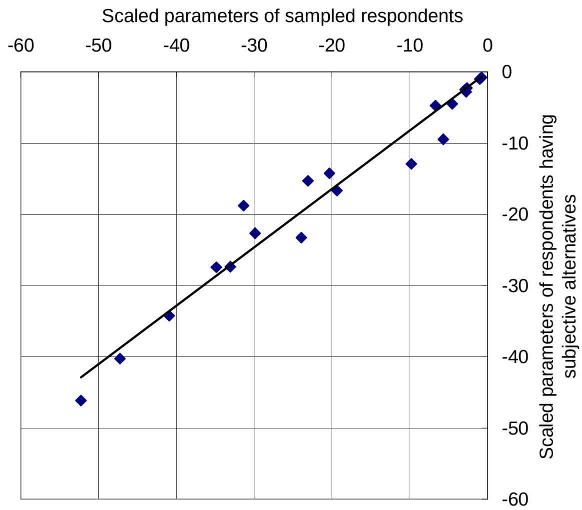

FIGURE 4: Relationships between scaled parameters for utility function estimated on objective choice sets for samples of 189 respondents and for respondents reporting subjective choice sets (N=189) (Model S-O)

The second analysis focuses on the number of alternatives in the choice set. For the 189 respondents reporting subjective alternatives 50 samples of small choice sets have been drawn from the objective choice sets. Again, for each set of samples the parameters of Model S-S (subjective choice set) have been estimated, yielding the mean and standard deviation of each parameter. In this case the differences are larger. For the number of legs the parameter estimate in Model S-S is significantly different, while for four other parameters the t-values are larger than 1.28. The differences in parameter estimates can also be seen in Figure 5. The R^2 is 0.77, while the parameter estimates of Model S-S are 9% higher. However, it should be noted that if the number of legs is omitted, the results improve significantly: R^2 becomes 0.93. This analysis shows that the subjective choice set appears to be a specific sample of all possible alternatives. Apparently, the subjective choice set is biased with respect to the number of legs within a multi-modal trip. As a result, the corresponding choice model is less sensitive for the number of legs. Subjective choice sets appear to have a majority of alternatives having the same number of legs as the chosen alternative, while alternatives objective choice sets show a larger variation in the number of legs.

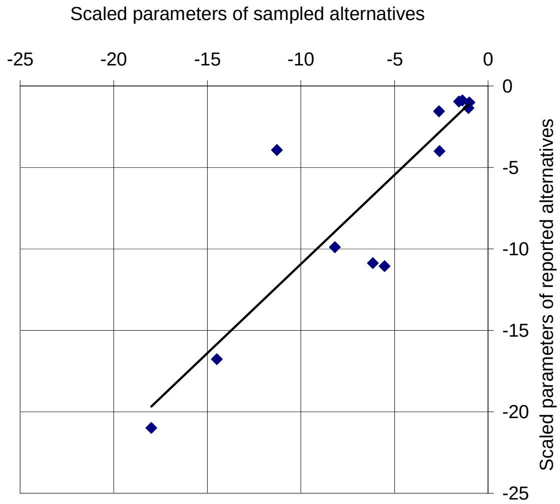

FIGURE 5: Relationships between scaled parameters for utility function for respondents reporting subjective choice sets estimated on small samples of objective choice sets and on subjective choice sets (N=189)

6. PERFORMANCE

The performance of the three estimated choice models for prediction is analysed by applying these to the three following datasets:

- 708 respondents, generated objective choice sets;
- 189 respondents, generated objective choice sets;
- 189 respondents, observed subjective choice sets.

Results are compared with respect to the log-likelihood and the prediction rate, that is the percentage of cases (respondents) for which the chosen alternative has the highest probability. Given the size of the parameter estimates of the three choice models, it can be expected that Model S-O performs best, while Model S-S is likely to yield the poorest results: larger parameters lead to a clearer distinction between alternatives.

The results for the log-likelihood are shown in Table 2. Obviously the score is best if the model is applied to the dataset it was estimated with. These results are indicated in italic. The differences between the performances of Model L-O and Model S-O are small. Apparently, the number of respondents does affect the level of detail of the choice model, but the main characteristics are already captured when the small dataset of 189 respondents is used. Furthermore, it is clear that Model S-S, which is based on subjective datasets, has the poorest performance: the log-likelihood increases with 13%. This confirms the notion that the model based on the subjective choice sets is less distinctive: the probability of the chosen alternative is lower, while those of unattractive alternatives are relatively higher. Applying Models L-O and S-O on the subjective choice sets leads to an increase of the log-likelihood of 8%.

TABLE 2: Comparison of log-likelihood values for combinations of datasets and choice models

Dataset	Model L-O	Model S-O	Model S-S
708 respondents, objective choice set	-1616.4	-1665.1	-1833.3
189 respondents, objective choice set	-462.7	-456.0	-515.9
189 respondents, subjective choice set	-158.6	-158.4	-146.5
Relative difference per dataset			
708 respondents, objective choice set	-	103%	113%
189 respondents, objective choice set	101%	-	113%
189 respondents, subjective choice set	108%	108%	-

In the case of prediction rates it is relevant to account for the choice set sizes. The generated objective choice sets are much larger than the observed subjective choice sets. Therefore Table 3 also includes the theoretical prediction rate in the equal probability case. The first finding to be noticed is that the differences in prediction rates are small. Apparently all models capture the main components of choice behaviour. In the case of the large dataset Model S-S, estimated on subjective choice sets, again has the poorest performance: the prediction rate drops 2.1%. The results for the second dataset, however, shows that this result is not generally true for all datasets considering generated objective choice sets. Reversely, applying Model L-O, based on generated objective choice sets, on subjective choice sets leads to a reduction of the prediction rate of 1.1%. The prediction rates further show that Model S-O has a more robust performance than Model L-O. It even yields better prediction rates for the two other datasets!

TABLE 3: Comparison of prediction rates for combinations of datasets and choice models (%)

Dataset	Model L-O	Model S-O	Model S-S	Equal probability
708 respondents, objective choice set	29.0	29.5	26.8	2.6
189 respondents, objective choice set	22.8	24.9	24.9	2.6
189 respondents, subjective choice set	57.7	59.3	58.7	36.0
Differences per dataset				
708 respondents, objective choice set	0.0	0.6	-2.1	
189 respondents, objective choice set	-2.1	0.0	0.0	
189 respondents, subjective choice set	-1.1	0.5	0.0	

Both comparisons suggest that Model S-S estimated using subjective choice sets, has a poorer performance in the case of prediction, although the differences are limited.

7. DISCUSSION AND CONCLUSIONS

This article presented an analysis of the relationships between choice set type and analysis purpose, in particular estimation and prediction. A distinction is made between (generated) objective choice sets which include all relevant and feasible alternatives or a representative subset, and (observed) subjective choice sets encompassing of alternatives considered by the traveller. Objective choice sets tend to be large, while subjective choice sets are generally small.

For estimation purposes it might be argued that subjective choice sets should be used to study choice behaviour because they are individual specific. Using objective choice sets confounds the choice model with the decision whether an alternative is actually considered or not. Objective choice sets on the other hand, might have the benefit that they have a larger variety in attributes and attribute values, thus allowing more detailed and advanced choice modelling. For prediction purposes an aggregate level of detail is used, while clear rules for generating subjective choice sets are lacking, which implies that only objective choice sets can be used.

Empirical analysis of choice sets for multi-modal inter-urban train trips in The Netherlands, confirms the theoretical analysis with respect to choice set characteristics. It was especially found that subjective choice sets have a limited variety with respect to the chosen alternative.

Using objective choice sets leads to more detailed choice functions and higher parameter estimates. This is true, even in the case that a small set of respondents is analysed. Subjective choice set usage leads to simpler models and smaller parameter estimates. Main difference is the valuation of the number of legs. Apparently, choice models based on objective choice sets incorporate the decision whether an alternative having many legs (i.e. many transfers) is relevant for the traveller by high parameter estimates for the number of legs. In the case of subjective choice sets, such alternatives are already eliminated by the traveller, allowing for smaller parameter estimates.

Analysis of the performance of choice models based on objective or subjective choice sets showed that the latter yield poor results when applied to objective choice sets. Since this is usually the case where prediction is concerned, it is recommended that choice models should be estimated on objective choice sets, that is, if it is intended to use the choice models for prediction purposes as well. Since subjective choice sets are difficult to establish, this finding also yields the additional benefit, that it makes it easier for researchers to construct appropriate choice sets for travel behaviour analysis. Furthermore, it allows for advanced modelling approaches leading to an even better performance for estimation and prediction.

ACKNOWLEDGEMENT

This research has been supported by the Transport Research Centre Delft. Furthermore, the authors would like to thank Michel Bierlaire (EPFL) for providing Biogeme for the logit estimations.

REFERENCES

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

Bovy, P.H.L. and Stern, E. (1990) Route Choice: Wayfinding in Transport Networks. Kluwer Academic Publishers, Dordrecht.

Cascetta, E. and Papola, A. (2001) Random utility models with implicit availability/ perception of choice alternatives for the simulation of travel demand. Transportation Research Part C, 9, 249-263.

Fiorenzo-Catalano, S. (2007) Choice set generation in multi-modal transport networks. TRAIL Thesis Series T2007/6, TRAIL, Delft.

Fiorenzo-Catalano, S., Hoogendoorn-Lanser, S. and van Nes, R. (2003) Choice set composition modelling in multi-modal traveling. IATBR, 10th International Conference on Travel Behaviour Research, Moving through Nets: The Physical and Social Dimension of Travel, Lucerne, August 2003.

Hoogendoorn-Lanser, S. (2005) Modelling travel behaviour for multi-modal transport networks. TRAIL Thesis Series T2005/4, TRAIL, Delft.

Hoogendoorn-Lanser, S. and van Nes, R. (2004) Multi-modal choice set composition: analysis of reported and generated choice sets. Transportation Research Record, 1898, 79-86.

Hoogendoorn-Lanser, S., van Nes, R. and Bovy, P.H.L. (2005) Path-size modelling in multi-modal route choice analysis. Transportation Research Record, 1921, 27-34.

Hoogendoorn-Lanser, S., van Nes, R., Hoogendoorn, S.P. and Bovy, P.H.L. (2006) Home-activity approach to multi-modal travel choice modeling. Transportation Research Record, 1985, 180-187.

Hoogendoorn-Lanser, S., van Nes, R. and Bovy, P.H.L. (2007) Application of constrained enumeration approach to multi-modal choice set generation. Transportation Research Record, in press.

Manski, C.F. (1977) The structure of random utility models. Theory and Decision, 8, 229-254.

Swait, J. (2001a) Choice set generation within the generalized extreme value family of discrete choice models. Transportation Research Part B, 35, 643-666.

Swait, J. (2001b) A non-compensatory choice model incorporating attribute cut-offs. Transportation Research Part B, 35, 903-928.

Train, K.E. (2002) Discrete Choice Models with Simulation. Cambridge University Press, Cambridge.

Van der Waerden, P., Borgers, A. and Timmermans, H.J.P. (2004) Choice set composition in the context of pedestrians' route choice modeling. Transportation Research Board, 83rd Annual Meeting of the Transportation Research Board, January 2004, Washington D.C.

APPENDIX. PARAMETER ESTIMATES

	Model L-O		Model S-O		Model S-S	
Number of respondents	708		189		189	
Choice set type	Objective (max 50)		Objective (max 50)		Subjective	
	p (scaled)	t-value	p (scaled)	t-value	p (scaled)	t-value
Car home-end constant	-24.87	-4.67	-18.79	-2.22	-9.90	-1.42[2]
Bus home-end constant	-33.53	-12.73	-40.27	-7.77		
Tram home-end constant	-25.82	-9.91	-27.45	-5.17		
Metro home-end constant	-17.17	-5.32	-23.29	-3.39		
Bus activity-end constant	-38.44	-14.74	-46.14	-8.68		
Tram activity-end constant	-31.18	-12.39	-34.26	-6.91		
Metro activity-end constant	-23.97	-8.61	-27.37	-5.08		
Bike constant (h-act)	-16.99	-7.08	-16.66	-3.77	-10.88	-2.51
Bus constant (h-act)					-20.99	-4.31
Tram constant (h-act)					-16.76	-3.43
Metro constant (h-act)					-11.05	-2.06
Express train constant	-5.64	-4.69	-4.73	-2.07		
Local train constant	-7.02	-3.96	-12.91	-3.45		
Local home-end station constant	-7.71	-3.95				
Express activity-end station constant	-7.26	-2.46	-9.47	-1.57[2]		
Local activity-end station constant	-11.66	-5.66				
Local station constant			-14.26	-2.37		
Access/egress walk time to station	-2.38	-17.00	-2.29	-8.36	-1.35	-4.55
Access/egress bike time to station	-2.61	-11.77	-2.50	-6.62	-0.88	-3.13
Access/egress car time to station	-4.26	-6.14	-4.49	-4.07	-1.55	-2.05
Bus in-vehicle time	-0.40	-2.51				
Tram in-vehicle time	-0.84	-4.61	-0.80	-2.52		
Train in-vehicle time	-1.00	-6.74	-1.00	-3.25	-1.00	-3.61
Number of low-frequency legs (<8)	-24.70	-15.77	-22.65	-7.74		
Number of high-frequency legs (>=8) incl. walk	-20.55	-11.57	-15.31	-4.95		
Number of legs					-3.93	-1.31[2]
Total headway	-0.13	-1.79[1]				
Walk time to home-end/ activity-end UPT stops	-2.21	-8.27	-2.82	-5.08	-0.95	-1.80[1]
Parking costs	-2.55	-1.95				
UPT costs	-3.72	-4.72			-4.00	-2.57
Null loglikehood	-2382.25		-651.78		-177.32	
Final loglikehood	-1616.39		-455.99		-146.50	
Likelihood ratio test	1531.71		391.58		61.64	
Adjusted rho-squared	0.31		0.27		0.11	
Free parameters	25		20		12	

Parameters scaled to units of train in-vehicle time

[1] $1.65 \leq$ t-value ≤ 1.96, [2] t-value ≤ 1.65