Determinants of the consideration mode choice set

Danique Ton[1,*], Dorine Duives[1], Oded Cats[1], Sascha Hoogendoorn-Lanser[2] & Serge Hoogendoorn[1]

[1] *Department of Transport & Planning, Delft University of Technology, The Netherlands*
[2] *KiM, Institute for Transport Policy Analysis, The Netherlands*
[*] *Corresponding Author, d.ton@tudelft.nl*

Keywords: consideration choce set, mode choice, discrete choice models, commute trips

1. Introduction

In discrete choice modelling, the specification of the choice set from which one alternative is chosen, is a complex task. This issue is mostly addressed in route choice as it is known for its large amount of possible alternatives (e.g. Prato & Bekhor, 2007). In mode choice, where generally only a limited number of alternatives is available (e.g. car, public transport, bicycle and walking), this issue has not yet received a lot of attention. However, also in case of mode choice, the size and composition of the individual's choice set influence the results of model estimation and consequently prediction and forecasting (Bovy, 2009). Ideally, the consideration choice set is used in modelling, which consists of all alternatives that are actively considered by the individual when making a choice (Hoogendoorn-Lanser & Van Nes, 2004).

Unveiling the consideration mode set of an individual is difficult, because asking about the considered modes in a survey or interview might result in misreporting of alternatives (Hoogendoorn-Lanser & Van Nes, 2004). Besides, the researcher cannot directly observe the consideration set when using revealed preference data. We argue however, that if data is available on the mode choices over a sufficiently long period of time (for example half a year), one will find all alternatives that are generally considered. Given that this aggregated data is available, it is possible to identify profiles of individuals (using their characteristics) that consider different sets of modes. Mansky (1977), introduced probabilistic availability of alternatives, where the choice of an alternative is conditional on the availability of alternatives in the choice set. Profiling individuals by identifying factors that influence the composition of the consideration choice set reflects this conditional availability, therefore using discrete choice models for identifying these factors seems an appropriate choice. These profiles would help in identifying this consideration choice set for other mode choice studies and in turn can improve model estimation and prediction.

This study, therefore aims at identifying which factors determine the composition of the individuals' consideration choice set for commute trips, so that profiles can be created. This is achieved by estimating discrete choice models using census data from the Netherlands, containing personal and household information, enriched with data on the reported modes during the last half year. The results of the study provide input for mode choice models by providing guidelines on how the consideration choice set depends on socio-demographics, ownership, built environment, household characteristics and the work environment.

2. Data collection and preparation

For this study, the census data of the Netherlands Mobility Panel (MPN) of the year 2016 is used, which contains data on the mobility patterns of individuals, and on personal and household characteristics (Hoogendoorn-Lanser et al., 2015). A companion survey on perceptions, attitudes and wayfinding styles towards active modes (coined PAW-AM) was distributed among the respondents of the MPN in 2017. This survey among other things addressed the use of modes in the last half year for commute trips, which is to reflect respondent's consideration choice set. A total of 2,775 respondents filled in both surveys, have a job and commute (resulting with 66% of the respondents).

In the companion survey, each respondent was asked to report which main modes (used for the largest part of the trip) were used. A total of five modes are considered for this study, namely the car, train, bus/tram/metro (BTM), bicycle, and walking. Walking and cycling are included separately, because they are very common in the Netherlands (about 46% of trips (CBS, 2016)). Furthermore, in the Netherlands the train is mostly used for inter-city travel, whereas BTM provides intra-city travel. Therefore, it is expected that these public transport modes are used differently and should be distinguished.

Several literature review papers have been published in the last decade that identify factors influencing mode choice (e.g. Munoz et al., 2016 and Buehler et al., 2016). Based on these reviews, we argue that the general consideration choice set is influenced by the following categories of factors; socio-demographics (e.g. age), ownership (e.g. owning a bicycle), built environment (e.g. urban density), household characteristics (e.g. number of people in the household) and work environment (e.g. number of working hours per week). Specific trip related aspects (e.g. distance or weather) will also influence the consideration choice set, however in this study only the general consideration choice set is identified (over the course of half a year). This list of factors is matched to the data of the MPN and the available factors are used in the research.

3. Methodology

As mentioned before discrete choice models can be used to evaluate which factors influence the composition of the consideration choice set. Therefore, allowing for identification of profiles of individuals with respect to different consideration choice sets. When combining five modes in the consideration mode set, a total of 31 alternatives can be identified (e.g. car-bicycle-walk). However, some of these are chosen by only few respondents, making it difficult for the model to determine the impact of factors on the consideration choice sets, which indicates that no valid (generalizable) argumentation can be provided. Consequently, alternatives were only included if more than 20 observations were recorded. This resulted in loss of 4.4% of the respondents (N = 2,652). Furthermore, a total of 12 alternatives remains.

The assumption, at this phase of the research, is that the choice for the mode set car and bicycle is independent from the choice for car alone or bicycle alone. Therefore, the multinomial model (MNL) structure, which assumes independence of irrelevant alternatives, is used. The utility function is defined such that for alternative i and observation n at time t specified in the following way (Ben-Akiva & Bierlaire, 1999):

$$U_{in} = V_{in} + \varepsilon_{in}, i \in C_n \tag{1}$$

Where V_{in} is the deterministic utility for alternative i (which is part of the choice set C_n) and observation n and ε_{in} represents the random error term, which captures uncertainty and is independent and identically (i.i.d.) Gumbel distributed. For the full paper, also more complex and behaviourally accurate model structures will be evaluated.

The model estimation proceeds iteratively, where all factors identified in the literature and available in the data are tested. The model is optimized in an iterative manner based on the significant influence of factors and model fit, in terms of adjusted rho-square and likelihood ratio. This way the factors that are most important for identifying which consideration choice set is chosen, are included.

Due to the interest in the predictive power of this model for identifying the consideration mode choice set, the model is estimated using 80% of the sample (randomly drawn). The remaining 20% of the sample is used for validation purposes. The models are estimated using the Python Biogeme package (Bierlaire, 2016).

4. First results

This section provides insight in the composition of the consideration choice sets to understand diversity and occurrence of different sets (4.1) and discusses the first results of the estimated models (4.2).

4.1.   The consideration mode choice set

The majority of the respondents have reported using only one main mode for the commute trips in the last half year (see Table 1.a). This, however, does not mean that these individuals use only one main mode in their complete mobility pattern, as for other trip purposes they could use more or different modes. For the commute trips, however, it seems that the respondents have created a rigid habit over time regarding mode use. This is also reflected in the top five most common consideration choice sets (see Table 1.b). Four out of five consist of a single-alternative choice set (number five contains two single-alternative choice sets that are chosen equally). However, the combination car and bicycle is also observed quite often.

TABLE 1: DISTRIBUTION OF RESPONDENTS OVER CONSIDERATION MODE CHOICE SET FOR COMMUTE TRIPS, WITH (A) TOTAL NUMBER OF MODES IN THE SET AND (B) MOST OBSERVED SETS.

# Modes	Percentage
1	79.8%
2	15.5%
3	3.6%
4	1.0%
5	0.2%

(A)

Top-5	Work (%)
1	Car (45.0)
2	Bicycle (25.4)
3	Car-Bicycle (9.4)
4	Train (5.1)
5	BTM and Walk (2.1)

(B)

4.2.   Model estimation results

The results of the final MNL model for commute trips are shown in Table 2. All categories of factors are represented in this model. Most of these factors, however, are dummy variables. Therefore, they were estimated as alternative specific parameters, so that influence per alternative could be determined. The car-only choice set is taken as the reference alternative.

TABLE 2: MNL RESULTS FOR COMMUTE TRIPS

Alternative	Significant parameters	# Obs. in estimation	% correctly predicted
Car	–	1021	88.5
Walk	-12 work hours (0.93), driver license (-1.57), own car (-1.89), 20-34 years (-0.74)	46	0
BTM	12-34 work hours (-1.46), 35+ work hours (-1.56), driver license (-1.72), own car (-2.00), PT subscription (0.89), PT reimbursement (3.94)	45	0
Train	12-34 work hours (-0.93), 35+ work hours (-1.08), driver license (-1.49), own car (-1.03), PT subscription (1.91), PT reimbursement (3.56), car reimbursement (-1.97), bicycle reimbursement (1.90), household size (-0.18)	115	55.6
Bicycle	-12 work hours (0.68), 35+ work hours (-0.66), driver license (-1.16), own car (-0.77), own bicycle (1.60), PT subscription (0.46), bicycle reimbursement (3.02), car reimbursement (-2.16), # children (-0.22), low urban density (-0.40), 12-19 years (1.25)	558	59.5
Bicycle – Walk	12-34 work hours (-1.08), 35+ work hours (-2.71), driver license (-1.64), own car (-1.64), bicycle reimbursement (3.60), very high urban density (1.05), 50-64 years (-1.36*)	24	0
Bicycle – BTM	driver license (-2.94), own car (-2.04), PT subscription (1.62), PT reimbursement (2.17), bicycle reimbursement (3.51), household size (-0.31), very high urban density (1.16), medium household income (-0.89*), medium education (-1.00)	28	16.7
Bicycle – Train	driver license (-2.29), own bicycle (-0.86), PT subscription (1.04), PT reimbursement (3.15), car reimbursement (-2.30), household size (-0.55)	17	0
Car – Train	own bicycle (-0.67), PT subscription (1.52), PT reimbursement (2.19), low urban density (-1.07*), 12-19 years (-2.47), 20-34 years (-2.80), 35-49 years (-3.40), 50-64 years (-3.68), medium education (-0.63*)	34	0
Car - Bicycle	35+ work hours (-0.57), bicycle reimbursement (2.32), car reimbursement (-0.88), 20-34 years (-0.96), 35-49 years (-1.09), 50-64 years (-1.08)	216	0
Car – Bicycle – Walk	12-34 work hours (0.77*), driver license (-1.73), car reimbursement (-1.66), household size (-0.49), # children (0.76*), 20-34 years (-1.96), 35-49 years (-1.83), 50-64 years (-1.49), medium household income (1.05)	17	0
Car – Bicycle – BTM	own car (-1.07), bicycle reimbursement (2.91), PT reimbursement (2.75), low urban density (-2.12), moderate urban density (-1.93*), very high urban density (-1.85*), 20-34 years (-2.11), 35-49 years (-2.51), 50-64 years (-4.19), medium household income (-0.97)	15	0
Model fit	Adj. rho-square: 0.528 Init. Log-likelihood: -5,307.761 Final Log-likelihood: -2,421.830 Likelihood ratio test: 5,771.860 # parameters: 86	Total: 2,136	Total: 59.1%

*Significant at 90% confidence interval, all other parameters are significant at 95% confidence interval.

Many factors significantly influence the composition of the consideration choice set, however some factors have a higher impact than others, indicating that these are more important and have a greater role in profiling of individuals. Reimbursements for bicycle and public transport are highly important with respect to the car-only alternative, as they benefit the alternatives that include cycling

and train or BTM. Consequently, if your employer provides reimbursement for using a mode you are more likely to choose it. Interestingly, these single-mode reimbursements also lead to using other modes (e.g. car-bicycle-BTM). Furthermore, age is an important element for the alternatives containing multiple modes, where a higher age is related to more disutility. Consequently, indicating that the older generation has a higher probability for habitual mode choice for the commute trip. Finally, gender was not found to be relevant for identifying the consideration set.

The model fit of the the model is very high, indicating that individuals are likely to be assigned to the correct choice. Furthermore, regarding the out-of-sample prediction, a total of 59.1% of the 516 observations were predicted correctly. These generally reflect the most observed alternatives. The model aims at estimating the correct consideration set for the largest part of the sample, which can be captured by the four most commonly chosen alternatives. Even though not all alternatives can be predicted correctly, these results provide a decent first insight into the factors that determine the consideration mode choice set.

5. Conclusions and future work

This abstract provides first insights into the factors that help identify the composition of the consideration mode choice set for commuter trips on an individual level. By applying discrete choice models, using census data from the Mobility Panel Netherlands, we were able to identify the significant influence factors that help in profiling of individuals.

The results show that many different categories of variables influence the consideration set, i.e. socio-demographics, ownership, urban environment, household, and work environment. Some of these factors are more important, like reimbursement for using a mode and age. Finally, the model is able to predict on out-of-sample data what their considered choice set is, but it gives priority to predicting correctly the alternatives that are chosen most, therefore not performing well on the less chosen alternatives.

In case the model predicts the wrong choice, which holds mostly for the multiple-mode alternatives, it often assigns it to one of the single mode alternatives that are part of it (e.g. in the car-bicycle alternative, the prediction is car or bicycle). This suggests that possible dependencies between alternatives may exist. Consequently, in the full paper we consider investigating (cross-) nested logit models, to test if these dependencies indeed arise and possibly lead to improvements in the model estimation and prediction.

Acknowledgements

References

Ben-Akiva, M., & Bierlaire, M. (1999). Discrete choice methods and their applications to short term travel decisions. In W. Hall, Handbook of Transportation Science (pp. 5-34). Dordrecht, The Netherlands: Kluwer.

Bierlaire, M. (2016). PythonBiogeme: a short introduction. Report TRANSP-OR 160706, Series on Biogeme. Transport and Mobility Labaratory, School of Architecture, Civil and Environmental Engineering. Ecole Polytechnique Fédérale de Lausanne, Switzerland.

Bovy, P. H. (2009). On modelling route choice sets in transportation networks. Transport Reviews, 43-68.

Buehler, R., Gotschi, T., & Winters, M. (2016). Moving toward active transportation: how policies can encourage walking and bicycling. Active Living Research.

CBS. (2016, July 29). Personenmobiliteit in Nederland. Retrieved July 14, 2017, from Statline: http://statline.cbs.nl [in Dutch]

Hoogendoorn-Lanser, S., & Van Nes, R. (2004). Multimodal choice set composition: Analysis of reported and generated choice sets. Transportation Research Record, 79-86.

Hoogendoorn-Lanser, S., Schaap, N., & Olde Kalter, M.-J. (2015). The Netherlands Mobility Panel: An innovative design approach for web-based longitudinal travel data collection. 10th International Conference on Transport Survey Methods (pp. 311-329). Transportation Research Procedia.

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

Muñoz, B., Monzon, A., & Daziano, R. A. (2016). The inreasing role of latent variables in modelling bicycle mode choice. Transport Reviews, 737-771.

Prato, C. G., & Bekhor, S. (2007). Modeling route choice behaviour: How relevant is the composition of choice set? Transportation Research Record, 64-73.