TNO PUBLIEK

Traffic & Transport
Anna van Buerenplein 1
2595 DA Den Haag
P.O. Box 96800
2509 JE The Hague
The Netherlands

www.tno.nl

T +31 88 866 00 00

TNO report

TNO 2022 R11882

Discrete Choice Models for Wellbeing oriented Policy Making

Date	August 2022
Author(s)	Dr. T. Bakri
No. of copies	-
Number of pages	20 (incl. appendices)
Number of appendices	1
Sponsor	TNO
Project name	Wise Policy Making
Project number	060.38077

Summary

The Early Research Program (ERP) “Wise Policy Making”, has the ambition to develop a methodology and tools that provide policymakers with support in prioritizing and steering towards societal wellbeing. As part of this ERP, a research team has dedicated their attention to an important aspect of wellbeing-oriented policy making; the choice behavior of humans.

In order to quantify the impact of policy measures in terms of how people’ s choice will be distributed among the available options, a range of methods are applied, each of which has its strengths and shortcomings. In this way, we hope to arrive at a balanced prediction that performs optimally under the different conditions and time scales, taking into account quantitative, qualitative and human bias aspects that matter in people’s choice. This is a necessary, but not sufficient, step towards quantifying the subsequent impact of policy measures on human well-being.

The approach consists at the front end of applying the MKBA (social Cost and Benefit Analysis) method followed by non-compensatory decision rules and a system dynamics approach to determine the number of relevant attributes and how they are qualitatively involved in people’s choice between the alternatives proposed by a policy measure. A discrete choice model based on random utility maximization (RUM) is then set up and its parameters are estimated from the input data. The estimated/trained model is used to estimate how people’s choice will be distributed among the available options based on the input data which in this case is a set of revealed preferences of an unbiased sample of the population. In other words the model is fitted onto the input data set to learn its parameters and use them to predict the distribution over the whole population. The number of variables/attributes taken into account in the utility function of each individual option will be limited to the most relevant and available for the use case under investigation. Elasticities and cross-elasticities are as well estimated giving the policy maker a powerful tool to estimate and predict the possible impact of measures on the modal split. After successfully estimating the modal on the input data, a comparison was made between the modal split of the whole Dutch population and the one of a subpopulation having predefined altruistic properties in order establish whether these altruistic properties contribute to less car use and more clean and sustainable mode choice like public transport or bicycle. A simulation of modal split was run as the population gradually shifted for the actual state to completely altruistic.

This approach is use case independent and can therefore be applied to different fields like the energy sector, sustainable and smart mobility or spatial planning. How the modelling in the mobility use case was carried out and implemented will be discussed in detail in the next introductory sections about Discrete Choice Models and data engineering.

Contents

1 Introduction

The Early Research Program (ERP) “Wise Policy Making”, has the ambition to develop a methodology and tools that provide policymakers with support in prioritizing and steering towards societal wellbeing. As part of this ERP, a research team has dedicated their attention to an important aspect of wellbeing-oriented policy making; the choice behavior of humans. How people react to various policy interventions, rules, regulations or incentives is a particularly important factor in understanding the eventual effect of the policy measure. There are plentiful examples of policy measures that did not attain the intended effects, because people reacted to the measure in an unforeseen manner. Such as that broadening the highways by adding extra lanes did not lead to less traffic jams, but instead it led to more people taking the car to work.

If a measure does not trigger the intended bahavior in people, a lot of time, energy, investments and sacrifices are made without reaching the end result. The ability to model prospective behavior as a reaction to a policy measure is of great value to policymakers, as it can guide them towards important insights in the expected effects of their measures.

2 Discrete choice modelling and utility functions

Discrete choice models (DCM), in the general setting, describe decision-makers' choices among alternatives. The decision-makers can be people, households, firms, or any other decision-making unit. The alternatives might represent competing products, courses of action, mobility modes for making a trip or any other options over which choices have to be made. In this introduction, we closely follow (Train, 2002)

DCM are applied in various fields ranging from econometrics, marketing to mobility policy to quantify impact of measures, estimate modal split and/or estimate market shares of certain product by predicting probability of adoption by the consumer. The idea/principle behind DCM is as follows:

- Suppose n individuals are faced with a choice between j alternatives.
- The benefit that person n experiences by choosing alternative i is called U_{ni}. This is a utility function belonging to alternative i. This is a 'monetization' of alternative i as experienced by person n and does not necessarily have to be expressed in Euro's. One of the properties of discrete choice models is that the value in itself of the utility function does not matter. It is more about the mutual order in utility between the options (which option has highest or lowest utility) that determine people's choices.
- A random utility maximization DCM in this case states that person n will choose alternative i if and only if: $U_{ni} > U_{nj}$ for all $j \neq i$..

Unfortunately, the utility functions U_{ni} as mentioned above are never fully known to the researcher/modeler because there is always some missing data on some features. Discrete choice models capture this lack of 'observables' by putting all unknowns in a stochastic variable ε_{ni}. The Utility function is given by:

$$U_{ni} = V(x_{ni}, s_n) + \varepsilon_{ni}$$

where,

$V(x_{ni}, s_n)$: The known part of the utility function as modeled by researcher. This can include continuous, linear functions, nonlinear concave and/or convex functions with the features and the (personal) characteristics as variables. Usually, a linear function of the features is used.

x_{ni}: A k dimensional vector containing the observed features of the alternative as experienced by decision maker/person n.

S_n: A m dimensional vector containing decision maker/person-related characteristics. For example, age, gender, annual income married/not married, smoker not smoker etc.

ε_{ni} A stochastic vector that models the unknown characteristics that affect the alternative.

The known part of the utility function can be written as follows:

$$V(x_{ni}, s_n) = \alpha_i + \sum_{j=1}^{k} \beta_j \, x_{ni_j} + \sum_{j=1}^{m} \gamma_j s_{n_j}$$

Where β_j are parameters to be estimated for alternative feature j, $\gamma_j$are parameters to be estimated for decision-maker characteristic j and α_i a bias term usually referred to as a constant specific to alternative i.

Different assumptions about the underlying distribution of this stochastic variable ε_{ni} lead to different RUM discrete choice models. The best known of these are the Multinomial Logit, Nested Logit, Probit and Mixed Logit. Each of these models leads in turn to a probability distribution over the options given the utility and a model fit on the input data with persons characteristics and options features.

The most well-known and used RUM DCMs include:

- **Multinomial Logit (MNL)**: The strength of the MNL is its simple (closed) form that allows the parameters to be easily estimated without much computation and simulation. Guaranteeing convergence to a global minimum and a first order accuracy. This is one of the most widely used DCM as a first approach because of its simplicity and global accuracy. This is the model that has been implemented in this ERP.
- **Probit and Mixed-Logit** are somewhat more sophisticated models that do take into account the heterogeneity in populations, are more accurate and therefore perform clearly better than MNL in general. However, they require a lot of simulation and computational capacity and assume that the population distribution is normally distributed and/or known in advance. This is not always the case.

2.1.1 Logit model

The logit model is obtained by assuming that each that ε_{ni} is distributed independently, identically extreme value. The density for each unobserved component of utility is in this case given by:

$$f(\varepsilon_{ni}) = e^{-\varepsilon_{ni}} e^{-e^{-\varepsilon_{ni}}}$$

From this assumption follows, after some computations, the closed-form of the probability that decision-maker n chooses alternative i

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}}$$

2.1.2 Parameter estimation

Given a data set with revealed preferences of a representative subset of the population under study, one can estimate the parameters of the utility function using the Logit model by maximizing the log-likelihood function. The log-likelihood function of the Logit with linear utilities is globally concave in parameters β_j, and γ_j which guarantees convergence in the numerical maximization procedures. Different computer packages contain routines for best fit estimation of logit models with linear-in-parameters representative utility.

2.1.3 Modal split computation

When the Logit parameters have been estimated on a representative sample of the population, the modal split for the whole population is computed as follow:

1 Use the model to compute for each trip of in the database the most likely mode to be chose by the user.
2 Assign that mode to the trip and do this for all trips in the database
3 The modal split of the population is then the normalized frequency of each mode in the database.

2.1.4 Elasticities

Since choice probabilities in the Logit are a function of observed variables, it is often useful to know the extent to which these probabilities change in response to a change in some observed factor. The so called *elasticities*. For example, in a big city like Amsterdam a policy maker could be interested in the following question: To what extent are people willing to avoid the car when going to the city center if the parking costs are increased by 10%?
To address these questions, derivatives of the choice probabilities are calculated. The change in the probability that decision-maker n
chooses alternative i given a change in an observed factor, z_{ni}, entering the representative utility of that alternative (and holding the representative utility of other alternatives constant). One of the advantages of the Logit model is that Elasticities and cross-elasticities (see definition below) have a mathematical closed-form which means that they can be computed analytically.
For the elasticities we get the following formula:

$$\frac{\partial P_{ni}}{\partial z_{ni}} = \frac{\partial V_{ni}}{\partial z_{ni}} P_{ni}(1 - P_{ni})$$

One can also determine the extent to which the probability of choosing a particular alternative changes when an observed variable relating to another alternative changes. The so called *cross-elasticities.* Considering the example of parking costs, the cross-elasticity in this case between public transport and car parking cost would be how does an increase in parking cost for car affect the mode choice of public transport. In other words, which part of the car drivers will leave the car for public transport after the parking costs have increased by 10%.
For the cross-elasticities we have the following closed-form formula:

$$\frac{\partial P_{ni}}{\partial z_{nj}} = -\frac{\partial V_{nj}}{\partial z_{nj}} P_{ni} P_{nj}$$

Elasticities and cross-elasticities give a powerful means to the policy maker when assessing and predicting the possible impact of measures on the modal split. Note that one the parameters of the models have been estimated, the elasticities become trivial to compute using the elasticities formulas mentioned above.

3 Data preparation

For the mobility use case, we focused on mode choice behavior. Human mode choice behavior is one of the most complex and intriguing phenomena policy makers are faced with. Especially when considering all the new modes that are emerging nowadays and innovative initiatives like Mobility as a Service (MaaS) where mobility is decoupled from owning a car but becomes a service one can buy to get from an origin to a destination. These initiatives in mobility can have a huge impact on how policy makers should organize the infrastructure, livability of their city and ultimately in the well-being of the population. Understanding which modalities people will choose and at what cost ratio they are willing to switch to cleaner modes is crucial. Having a good DCM to predict mode choice is important here but not enough. A sound, adequate and nonbiased set of input data is at least as important as the model itself. This is usually an underexposed item in peer reviewed articles. One often starts by assuming the input data is fit for purpose. In this section however, we intend to take the interested reader into the world behind data preprocessing and engineering.

3.1.1 Description of the input data

- *Onderweg in Nederland (ODiN)*
  The ODiN is a Dutch annual revealed preferences survey conducted by CBS during the year among roughly 50 thousand people. Respondents are asked to record, for a specific day of the year, which trips they make, for what purpose, with which means of transport and how long the trip takes. This is then enriched and expanded by CBS to create a complete file which can be used to link mobility behaviour to personal and household characteristics. For this project, the ODiN data from 2017 was used as input. Thew data consists of:
  - 160+ variables in total
  - 20+ on household: composition, income, transport mode possession, …
  - 60+ on trip: origin, destination, purpose, distance, duration, departure time, mode, …
  - 20+ on legs: mode, duration, used train station, …

  Selected variables:
  - Household composition
  - Children in household
  - Car ownership
  - Education
  - PT discount card
  - Household income
  - Driving license
  - Gender
  - Age
  - Trip purpose
  - Urbanization of O/D
  - Trip distance
  - Mode choice
  - Trip travel time
- *General Transit Feed Specification (GTFS) data*
  GTFS is a standard format in which information about the public transport network (lines, stops, timetables) is represented. For the Netherlands, roughly

every week a new GTFS file is published containing the timetables of the various public transport providers.

- OpenStreetMap (OSM)
  OSM provides freely available maps at a high level of detail. These maps are partly based on donated commercial data, partly on GPS data collected and maintained by thousands of volunteers worldwide. Snapshots of the data can be downloaded and include the road network (car, pedestrian and bicycle) at a given date.
- OpenTripPlanner *(OTP) tool*
  OTP is a multi-modal route planner that uses public OSM and GTFS data (mentioned above). The route planner can plan trips between two points for a given time and day, using (a combination of) the modalities car, bicycle, public transport or walking.

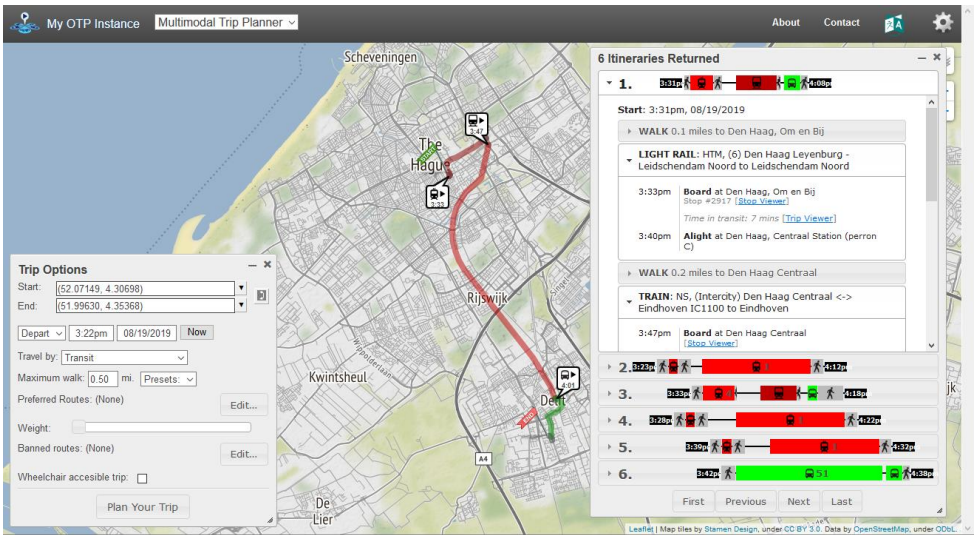

Figure 1 Screenshot of the OTP Graphical User Interface.

- *Zip code 4 areas (PC4)*
  CBS provides annual demographic and socioeconomic key figures for the numerical portion of zip code zones (PC 4, e.g., 2595). For the current project, the zip code zones were mapped onto their resulting geographic center of gravity for simplicity. For this project, we chose to use zip code zones of 2017 (Central Bureau of Statistics (CBS), 2017), because they correspond to the ODiN year used in this study.
- *Parking charges per PC4 area*
  The RDW offers, through the National Parking Register, open data on parking facilities and corresponding charges of all public parking facilities within the Dutch municipalities. The data contains 6446 active parking facilities, 64% (4133) have location information that can be linked to a zip code zone. This does not include private garages such as those of a specific store or a residential tower, but does include permit areas (38%), paid street parking (33%), garages (8%), carpool (5%) and other zones (9%: waivers, blue zones, etc...) 26% (1097) of all PC4 zones have at least 1 linked parking facility. The coverage of this data at the PC4 level is visualized in Figure 2, all colored zones have parking facility information. From this data, an average static parking rate per PC4 area based on the rate of each parking facility on Tuesday afternoon at 15:00 was derived and used in the utility function for the car to account for the

parking charges.

Figure 2 PC4 zones in The Netherlands with parking facility information from the NPR.

- *KNMI precipitation data*
  KNMI provides precipitation data on a 1x1 km grid with a time resolution of 5 minutes. The data comes from two weather radars in Den Helder and Herwijnen and is validated and corrected with data from roughly 300 KNMI ground stations. For this project, precipitation data from 2017 is used (KNMI & Overeem, 2017) to match ODiN's baseline year.

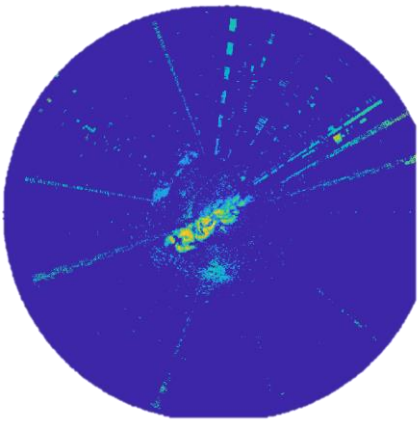

Figure 3 KNMI data showing precipitation intensity derived from radar reflections.

3.1.2 Data preprocessing and fusion

Figure 4 shows roughly the process that was followed to arrive at the final dataset. The trips from ODiN were enriched with information about the amount of rain at departure/arrival and with information (travel time, distance, number of transfers) about alternative ways to make the same trip. A price was then determined for each trip and alternative based on the distance and parking rates at the destination. This results in a dataset that shows both the choice made from ODiN, and the corresponding features of the alternatives and environmental factors.

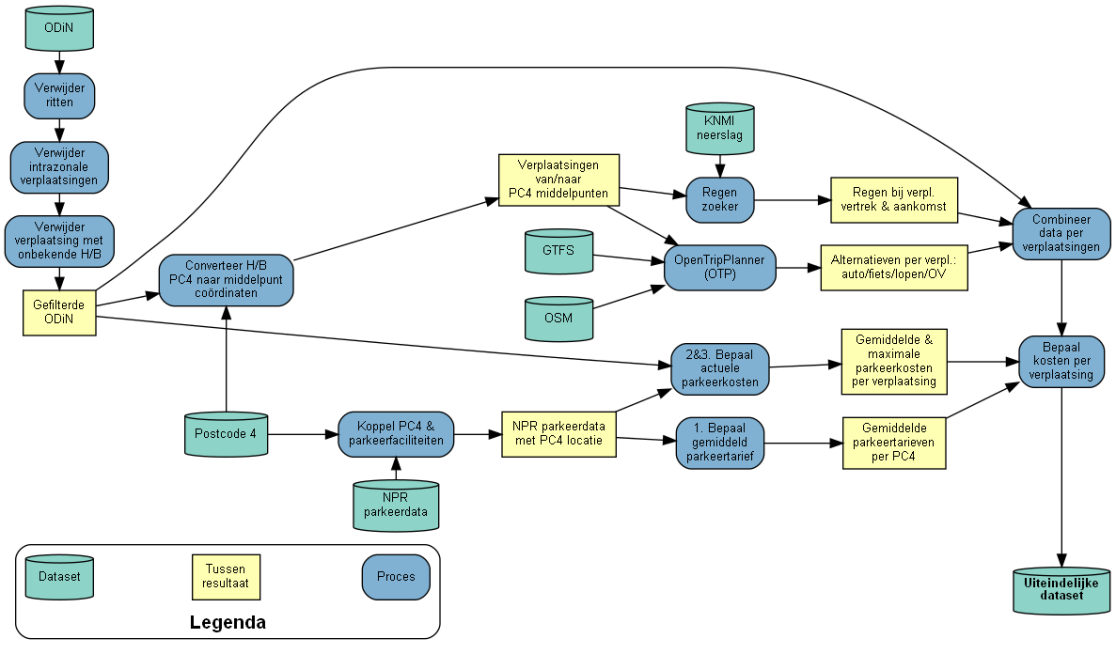

Figure 4 Data preprocessing and fusion process.

The full process is described below.

- **Filtering**: the ODiN data was first filtered (top left of Figure 4):
  - Only data at the displacement level was included, not at the before/after transportation level. After all, the project focuses on the modal shift over the entire door-to-door movement, not just the change in the before/after transport. In addition, this simplifies the processing considerably. In this step, a relevant subset of the 203 columns was chosen.
  - Intra-zonal trips were omitted because no alternative trip can be found for this using OpenTripPlanner (OTP) tool, as the origin and destination would be the same.
  - Movements with an unknown origin or destination have been omitted because no alternatives can be found for these, nor can the amount of rain at the origin/destination be determined.
- **Conversion PC4**: for the trips, the origin/destination is only known at the zip code 4 level (e.g. 2595). To link the data to the other sources, exact coordinates are needed (e.g. 52.082, 4.325). For this purpose, the 2017 zip code file was used to determine the coordinate of the center of gravity for each zip code zone.
  - **Assumption**: the coordinates of the center of gravity are the exact origin/destination. As a result, trips only occur from the center of one PC4 zone to the center of another PC4 zone. This is a bit of a limitation in the model that is inherent to the PC4 granularity of the revealed preferences data of the CBS (ODiN). As a consequence, the center of a PC4 zone can sometimes end up in, remote industrial sites, or sometimes even in the middle of a lake or river.
- **Rain**: based on historical data of the KNMI, each trip was assigned the amount of precipitation there was around the departure and arrival time by mapping KNMI radar data onto the PC4 geolocations.

  - **Assumption**: rain in a 3x3 km quadrant around the PC4 center is averaged. This is an approximation, the geographical layout of the KNMI data does not always match exactly with the PC4 centers.
  - Assumption: rain 15 minutes around the departure/arrival time is included in the determination of the amount of rain. Whereby the 5 minutes interval containing the departure/arrival time should be centered around the 15 minutes interval. For example, if a trip starts at 13:11, the sum of the KNMI rain values from 13:05-13:20 is taken into account at that location.
- **Alternatives**: for each realized trip in the ODiN, OTP was used to find out what the travel time, distance and number of transfers would have been if the modes car, walking, bicycle or public transport had been used for the same trip. The time of day and the day of the week are taken into account in the application to the route planner
  - **Assumption 1**: only the first result of the route planner is considered. This is often, but not always, the best/fastest travel option for the given alternative.
  - **Assumption 2**: The route planner uses a maximum walking distance when determining if a route is possible for a given mode. If a route requires walking more than this threshold distance, the route is excluded and the route planner does not find an alternative using given mode. The walking distance threshold differs per main mode chosen:
    - For car this distance is 800 m.
    - By bicycle the distance is 200 m.
    - For public transport the distance is 2000 m.
    - For walking the distance is unlimited, but the maximum travel time is set to 2 hours otherwise OTP will not consider walking as an option.
- **Data fusion**: the trips from the filtered ODiN data were combined with the precipitation data at departure/arrival and merged with information about alternative trips for each trip. This involves determining whether an alternative was actually available to a traveler based on the assumptions below.
  - **Assumption 1**: the car is only available if the traveler has a driver's license according to ODiN or has used the car as a mode in ODiN.
  - **Assumption 2**: public transport is only available if OTP found an OV route or if the traveler took public transport as mode in ODiN.
  - **Assumption 3**: cycling is only available if the travel distance is less than or equal to 10 km (95th percentile cycling distance in ODiN 2017) or if the traveler has taken the bicycle as travel mode according to ODiN.
  - **Assumption 3**: walking is only available if the travel distance is less than or equal to 2 km or if the traveler made the trip walking according to ODiN.
  - **Assumption** 4: travel as a car passenger is always possible, there are no restrictions here.

The final dataset that serves as input to the DCM follows from the above steps. For more detail on the used variables in terms of modes, corresponding features and decision maker’s characteristics we refer to Appendix A

4 Model fit results (Full dataset)

Estimation of the Logit model on the complete input data set yields the following modal split for the Netherlands:

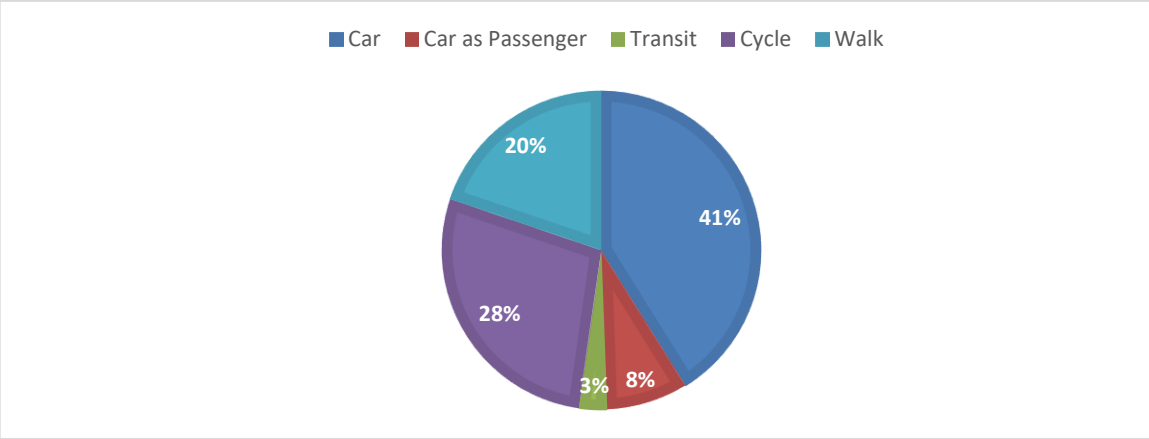

Figure 5 Predicted modal split (Full dataset).

To have an idea how well the modal split prediction is we show the predicted versus real modal split. In Figure 6

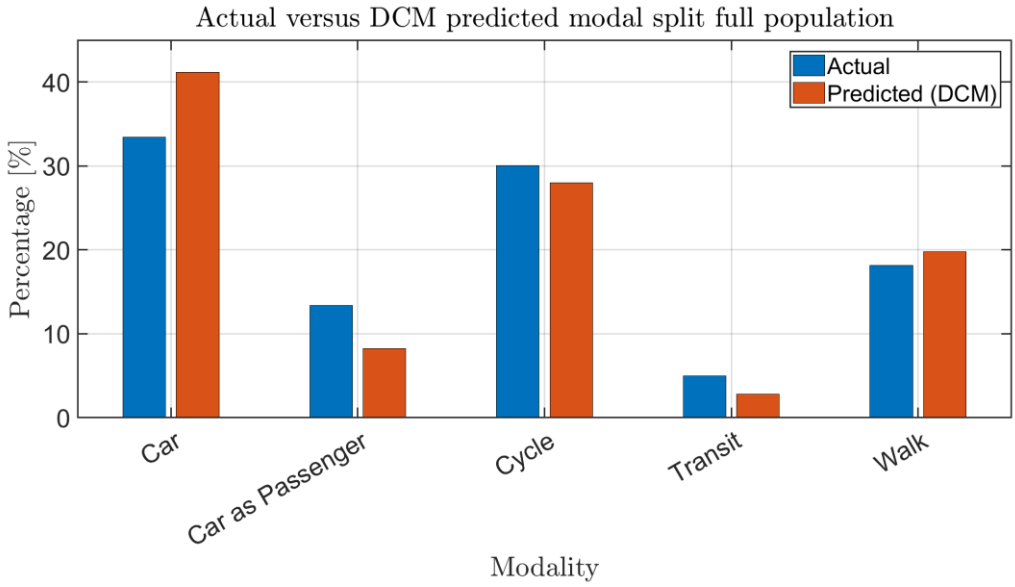

Figure 6 Plot of the actual versus predicted modal split (Full dataset).

The overall performance of the modal is not bad. The modal tends to overestimate a bit car use and underestimate the car as passenger mode and public transport. For a quantitative assessment of the goodness of fit, see the estimation report in Figure 7. The goodness of fit parameters Rho-square and Rho-square-bar are quite good! The other parameters are used to compare different models with each other.

Number of estimated parameters:	66
Sample size:	75043
Excluded observations:	0
Init log likelihood:	-96025.95
Final log likelihood:	-37466.46
Likelihood ratio test for the init. model:	117119
Rho-square for the init. model:	0.61
Rho-square-bar for the init. model:	0.609
Akaike Information Criterion:	75064.93
Bayesian Information Criterion:	75673.83
Final gradient norm:	1.32E+02
Diagnostic:	b'CONVERGENCE: REL_REDUCTION_OF_F_<=_FACTR*EPSMCH'
Database readings:	5257
Iterations:	4808
Optimization time:	01:18:48.3
Nbr of threads:	64

Figure 7 Estimation report (Full dataset).

5 Model fit results (altruistic datasubset)

A subset of the population that we arbitrarily called ‘altruistic’ based on the criteria defined below was also used as input for the model and the modal was once again estimated. The model split from both populations were compared and a simulation was run as function of the penetration rate α of the altruistic population. With $\alpha = 0$ meaning the full dataset and $\alpha = 1$ corresponding to the ‘altruistic’ dataset. The used criteria based on the decision makers characteristics were as followed:

- Number of persons in household: <= 4
- Age: 30+
- High income 50+ kEuro/year
- Civicly engaged
- Highly educated
- Has an electric car or running on lpg fuel

The estimation results are summarized below:

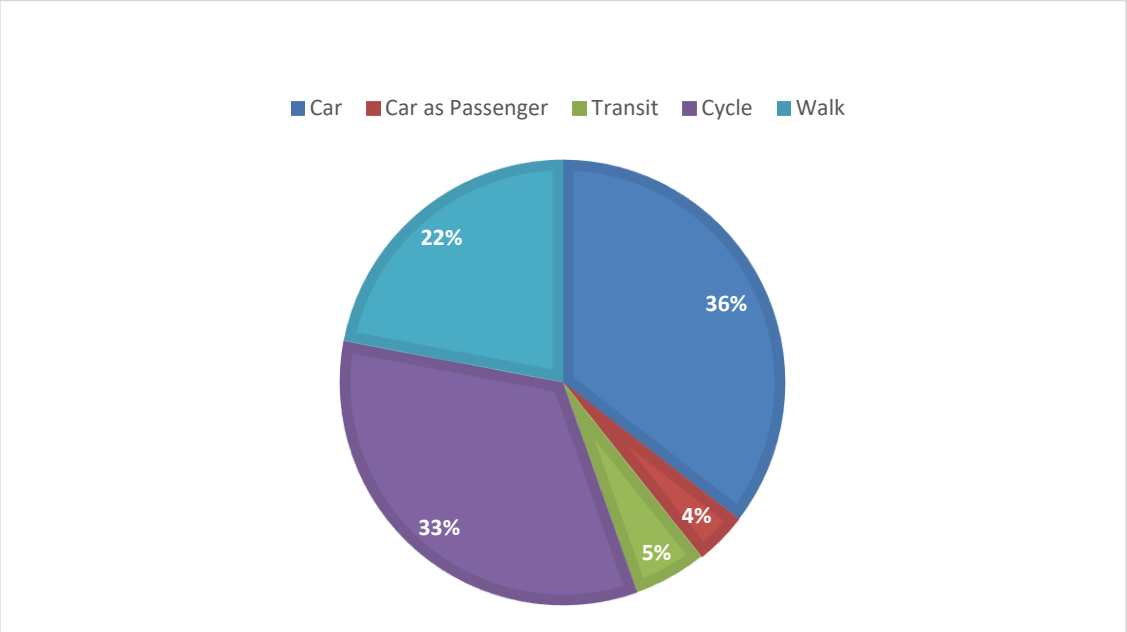

Figure 8 Predicted modal split (Sub dataset).

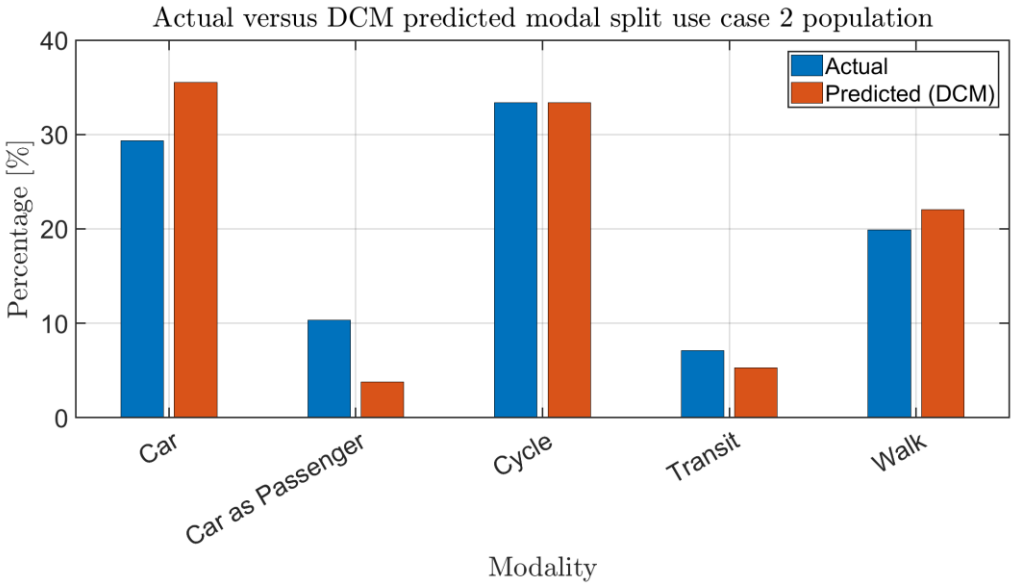

Figure 9 Plot of the actual versus predicted modal split (Sub dataset).

Number of estimated parameters:	57
Sample size:	26090
Excluded observations:	0
Init log likelihood:	-33806.36
Final log likelihood:	-13640.23
Likelihood ratio test for the init. model:	40332.26
Rho-square for the init. model:	0.597
Rho-square-bar for the init. model:	0.595
Akaike Information Criterion:	27394.47
Bayesian Information Criterion:	27860.12
Final gradient norm:	4,95E+01
Diagnostic:	b'CONVERGENCE: REL_REDUCTION_OF_F <= FACTR*EPSMCH'
Database readings:	13357
Iterations:	12120
Optimization time:	0:41:33.604054
Nbr of threads:	64

Figure 10 Estimation report (Sub dataset).

Comparing the results we see that in the altruistic population a clear shift from car as modality towards cleaner modes like the bicycle and to a lesser extent Public Transport.

6 Conclusion and discussion

Within this ERP an elaborate DCM tool was developed and estimated on preprocessed and enriched revealed preferences dataset. Elasticities were computed and impact on the different modes can be predicted as function of the elasticities and the change in a given feature or observed variable. Estimating the model on a subset of the population and computing the corresponding elasticities and cross-elasticities is straight forward. This gives the policy maker a powerful means to experiment and quantify the potential impact on modal split of a range of measures.

6.1 Random Regret models

Besides the RUM models there is another category of DCM models, inspired by Regret Theory, that can be more suitable in certain case studies, the so called Random Regret model. Those models are based on minimizing the regret as function of the performance of the chosen modality relative to that of the alternatives. This type of models was also carried out during this project as part of a master thesis jointly supervised by this ERP and the University of Utrecht, The Netherlands. The master thesis with the results (van der Pol, 2020) will be added to this document as an appendix. Random Regret models are more elaborate, highly nonlinear (meaning no guarantee to convergence to a global minimum) and capture a different aspect of the human nature. Namely the loss aversion encrypted in the human reptile brain when making choices. An interesting set of models to combine with RUM (random utility maximization) models in order to get a better modelling of human choice. To be considered in future work.

7 Bibliography

API (OV). (2021). GTFS Nederland. Retrieved from http://gtfs.ovapi.nl/nl/archive/NL-20210716.gtfs.zip.

Centraal Bureau voor de statistiek (CBS), & Rijkswaterstaat (RWS-WvVL). (2020). Onderzoek Onderweg in Nederland—ODiN 2019. [Application/pdf,.sps,.dat,.dta,.sav,.csv]. Centraal Bureau voor de Statistiek (CBS). Retrieved from https://doi.org/10.17026/DANS-XPV-MWPG.

KNMI, & Overeem, A. (2019). Precipitation—5 minute precipitation accumulations from climatological gauge-adjusted radar dataset for The Netherlands (1 km, extended mask) in KNMI HDF5 format. *KNMI Data Services. https://dataplatform.knmi.nl/dataset/rad-nl2*.

OpenStreetMap (contributors). (2021). OpenStreetMap data for The Netherlands. Geofabrik. Retrieved from http://download.geofabrik.de/europe/netherlands-latest.osm.pbf.

OpentTripPlanner contributors. (2020). OpenTripPlanner (1.5.0) [Java]. Software Freedom Conservancy. Retrieved from http://www.opentripplanner.org.

RDW. (2019). Open Data Parkeren: TARIEFDEEL | Open Data | RDW. Retrieved from https://opendata.rdw.nl/Parkeren/Open-Data-Parkeren-TARIEFDEEL/534e-5vdg.

Statistiek(CBS), C. B. (2019). Kerncijfers per postcode. Centraal Bureau voor de Statistiek. Retrieved from https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische-data/gegevens-per-postcode.

Train, K. (2002). *Discrete Choice Methods with Simulation.* Cambridge University Press.

van der Pol, W. (2020). The accumulated regret of trip chaining. *Master thesis University of Utrecht-TNO*.

A Enriched and fused input data

The enriched and fused input data set consists of among, others, the following columns:

- trip_id: unique travel ID from ODiN
- sted_{o, d}: degree of urbanization origin (o) or destination (d) postal code;
  - 1: Very urban (surrounding address density of 2500 or more)
  - 2: Highly urban (ambient address density of 1500 to 2500)
  - 3: Moderately urban (ambient address density from 1000 to 1500)
  - 4: Low-urban (ambient address density from 500 to 1000)
  - 5: Non-urban (ambient address density of less than 500)
- ovstkaart: indicates whether a traveler has a week (1), weekend (2) or no OV student card (0).
- weekday: indicates the day of the week;
  - 1: Sunday
  - 2: Monday
  - etc...
  - 7: Saturday
- d_hhchildren: do (1) or not (0) have children in the household.
- d_high_educ: the traveler has (1) or not (0) completed a college education or higher.
- gender: gender of the traveler;
  - 0: woman
  - 1: man
- age: age group of the traveler;
- 1: 6 through 17 years
- 2: 18 up to 54 years
- 3: 55+
- driving_license: the traveler has (1) or no (0) car license.
- car_ownership: the household has (1) or no (0) car.
- main_car_user: the traveler has either (1) or no (0) car in his/her name.
- hh_highinc10: the household has either (1) or no (0) top 10% income.
- hh_lowinc10: the household has either (1) or no (0) bottom 10% income.
- hh_highinc20: the household has either (1) or no (0) top 20% income.
- hh_lowinc20: the household has either (1) or no (0) bottom 20% income.
- pur_{home, work, busn, other}: the purpose of the trip is to go home (1) or not (0), work, business or other.
- {departure, arrival}_rain: precipitation in mm/15min at the origin (departure) or destination (arrival).
- choice: the chosen modality according to ODiN;
  - 1: Car driver
  - 2. Car passenger
  - 3: Public transport (bus, streetcar, metro or train)
  - 4: Bicycle
  - 5: Walk
- dist_{car, carp, transit, cycle, walk}: distance of the displacement or alternative by car (car), as passenger (carp), by public transport (transit), bicycle (cycle) or on foot (walk) in meters.

- t_{car, carp, transit, cycle, walk}: travel time of the journey or the alternative by car, as passenger (carp), public transport (transit), bicycle (cycle) or on foot (walk) in seconds.
- c_{car, carp, transit, cycle, walk}: cost of the journey or the alternative by car (car), as passenger (carp), by public transport (transit), bicycle (cycle) or on foot (walk) in Euros.
    - c_car is NaN for all rows because a definitive parking cost method was not chosen at the time of data preparation (see columns below).
- {vc, pc}_car: distance-dependent (vc) and parking (pc) costs for the car driver.
    - pc_car is NaN for all rows because a final parking cost method was not chosen at the time of data preparation (see columns below).
- pc_car_tue: parking cost based on average rate on Tuesday afternoon.
    - Also available with _nan suffix, herein the NaN values are not replaced with 0.
- av_{car, carp, transit, cycle, walk}: availability (1) or unavailability (0) of the modality for a trip or alternative.
- actduur: activity duration at the destination in minutes from ODiN.
- traveltime_sec: actual traveltime in seconds from ODiN.
- afstv_m: actual travel distance in meters from ODiN.
- aankpc: numeric part of the destination zip code from ODiN.

Distribution list (TNO 2022 R11882)

TNO

Referent Adelbert Bronkhorst	email-alert
Projectleider J. Sassen – Van Meer	email-alert
Research manager (projectleider) A. Woering	email-alert
Research manager auteur J.P. Dezaire	email-alert
Auteur Dr. T. Bakri	email-alert email-alert
TNO Bibliotheek locatie Soesterberg	hard copy & cd