Dat.mobility advancing analytics

Comparing logit models estimated on cross-sectional with longitudinal data

INTRODUCTION

In strategic transport models, traffic is a result of the process of travellers choosing a destination, mode of travel, route and departure time. Most strategic traffic models contain behavioural choice models for each individual choice tied together partly sequentially and partly simultaneously.

Most choices are modelled using discrete choice models employing the concept of utility maximization. Estimation of these models comes down to employing log-likelihood maximization to find parameters that best fit to a dataset containing observed choices (usually data from a survey). The utility functions employed may contain attributes of the alternative (e.g. the travel time for mode car or the access time, travel time and number of interchanges for public transport), but also attributes of the decision maker (e.g. his/her age, income or household situation).

Goudappel and Dat.mobility have estimated nationwide-representative mode- and destination choice logit models using ODiN data as the dataset with observed choices (Brederode and van Essen, 2021). ODiN contains observations reported by its respondents themselves from a week of travel of about 30000 persons). To obtain sufficient observations, the dataset was comprised of several years of 'stacked' ODiN data.

Since a few years, Dat.mobility together with Kantar and Mobidot have started its own mobility panel (Nederlands VerplaatsingsPanel – NVP) which continuously monitors travel behaviour using a smart phone app component that has access to the geolocation of the phone. From this data origins, destinations, used mode of transport, departure time and routes can be derived. The NVP currently contains about half the number of persons as ODiN, but it gathers data year round (instead of just a week). Furthermore, it contains observed instead of reported data and the number of participants is still growing.

Dat.mobility BV is part of Goudappel Groep.

Snipperlingsdijk 4 7417 BJ Deventer The Netherlands P.O. Box 161 7400 AD Deventer The Netherlands

+31(0) 570 666 222 info@datmobility.nl www.datmobility.nl VAT NL 0062 45 079 B01CC. 2710 3813IBAN NL59 INGB 0701 2168 08

Dat.mobility

Ontwikkeling bestemmingen (gemiddelde per weekdag)



Figure 1: illustrative example of NVP data usage: relative changes of number of trips per purposes during COVID-19 pandemic (from:

<u>https://www.goudappel.nl/nl/expertises/data-en-it-oplossingen/nederlands-</u> verplaatsingspanel)

PROBLEM DESCRIPTION

Mobility behavior is changing rapidly due to new concepts (e.g.: shared vehicles, Mobility-as-a-Service) and societal developments (e.g. the COVID-19 pandemic and the rising fuel cost). Because NVP data is gathered continuously, it is expected that it allows to update the choice models more frequently, without having to wait until several years of ODiN data is available. However, it is unknown to what extent this expectation is true, as no logit models have yet been estimated on NVP data.

RESULT / OBJECTIVE

The goal of this research is to re-estimate (a subset) of the logit models that where estimated on ODiN data, but now using NVP data to find out how many NVP observations are required to match the descriptive value (i.e. the confidence intervals of the parameter estimates) of models estimated on a given amount of ODiN data.

ASSIGNMENT

To re-estimate the models, alignment of the variables in NVP data with those in ODiN data should take place. Furthermore, the non-chosen alternatives should be

Dat.mobility

ADVANCING ANALYTICS

extracted from the Mobility Spectrum (Possel et al., 2020) and added to the NVP dataset in exactly the same way as was done for the ODiN dataset.

The re-estimations should take the longitudinal nature of NVP data into account (i.e. compensate for autocorrelation using e.g. panel error component logit models) and a method has to be found (or developed) to be able to compare the parameter estimates and their confidence intervals apple-to-apple.

The re-estimations should be run multiple times, each time increasing the amount of NVP data. Finally, for different choice models, a relationship between amount of NVP data and descriptive value (relative to the baseline formed by the ODiN models) should be derived and synthesized into a rule of thumb that approximates the relationship over all models.

INFORMATION

When interested in this internship assignment please contact: Mariska van Essen (<u>mvessen@goudappel.nl</u>) or Luuk Brederode (<u>lbrederode@dat.nl</u>).

References

- Brederode, L., van Essen, M., 2021. Landelijke keuzemodelschattingen Octavius. Available from: https://www.slideshare.net/LuukBrederode/lunchlezinglandelijke-keuzemodellen-voor-octavius
- Possel, B., Graaf, S.D., Brederode, L., 2020. The Mobility Spectrum: A data driven strategic transport model for the whole of The Netherlands. Presented at the European Transport Conference 2020, online due to COVID19, p. 12.