Dat.mobility

ADVANCING ANALYTICS

VACANCY: 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 & van Essen, 2021, Van Essen & Voorhorst, 2022.). ODiN contains observations reported by its respondents themselves from a day of travel of about 45.000 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 a quarter of the number of persons as ODiN, but it gathers data year round

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 B01
CC. 2710 3813
IBAN NL59 INGB 0701 2168 08

(instead of just one day). Furthermore, it contains observed instead of reported data and the number of participants is still growing.

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.

RESULT / OBJECTIVE

The goal of this research is to find out what NVP time span (hence, number of observations) is required to estimate a logit model that matches the level of explanatory power and descriptive values (i.e. the confidence intervals of the parameter estimates) of models estimated on a given amount of ODIN data (say, one year).

ASSIGNMENT

To estimate logit models on both ODIN and NVP data with the same specifications (alignment of the variables in NVP data with those in ODiN data should take place, correction for autocorrelation using e.g. panel error component logit models is needed). You will need to estimate the model on the NVP dataset using different time spans (hence, different number of observations) for the comparison.

A first study on this topic compared panel mode choice modeling estimates with its cross-sectional counterparts (Mahdian, 2023). The current assignment continues on this study, with more emphasis on the definition and operationalization of a method to compare 'the level of explanatory power and descriptive values' of two models estimated on different datasets (What is done before in literature? What statistical methods exist and are commonly used?). Insights from the previous study and procedures to obtain cleaned datasets are available as a starting point.

Finally, (if time allows) a case study (in which both models on ODIN and NVP datasets are applied in a real-life application context) shows how differences in parameter estimates translate into differences in model outcomes and (possibly in) policy makers' decisions.

INFORMATION

When interested in this internship assignment please contact: Mariska van Essen (mvessen@goudappel.nl) and Luuk Brederode (lbrederode@dat.nl).

References

- Brederode, L., van Essen, M., 2021. Landelijke keuzemodelschattingen Octavius. Available from: https://www.slideshare.net/LuukBrederode/lunchlezing-landelijke-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.
- Van Essen, M., Voorhorst, J., 2022. Estimation and regionalisation of nationwide parameters for destination and mode choice. Presented at the European Transport Conference (ETC), 7-9 september 2022, Milan, Italy. Available from: https://aetransport.org/past-etc-papers/search-all-etc-conference-papers?abstractId=7630&state=b
- Mahdian, M., 2023. Panel Mode Choice Modeling and Comparing the Estimates with Cross-section Counterparts. (thesis) Tilburg University. Available from: https://arno.uvt.nl/show.cgi?fid=161463