Summary

Further development of models for distributions of trip purpose and arrival time of train trips

TØI Report 1876/2022
Authors: Stefan Flügel, Aino Ukkonen, Rikke Ingebrigtsen, Frants Gundersen & Nina Hulleberg
Oslo 2022 45 pages Norwegian language

Distributions for travel purposes and desired arrival time are key inputs in the transport model Trenklin. In 2016/2017, TØI established generic models that can predict these distributions for all train station pairs in Norway. This report documents a further development of these models, as well as the establishment of a model for predicting the type of day (working day or weekend/holidays) that is integrated in the new travel purpose model.

Background

This report has been commissioned and financed by the Norwegian Railway Directorate and the models will be used in connection with the Trenklin model. Travel purpose and day type are used in Trenklin as a segmentation variable, while the day-time distribution (given travel purpose) represents a distribution over the desired arrival time and is used to calculate hidden waiting times in Trenklin.

In short, the models predict how a total number of passenger trips can be divided into a) type of day (working day og weekend/holiday), b) travel purposes, and c) desired arrival time at the final station. The level of analysis is train station pairs (start station and end station) within Norway. There are just under 100,000 such train station pairs in Norway.

Methodological improvements

The work is based on methodology established by the first version of the models. The following improvements were made:

- Use of updated data. Central data sources (RVU, commuter data) are from 2019.
- Larger data base. This has been achieved by including car trips in the data basis for the day-of-time model, and car trips and other public transport in the model for travel purpose and type of day. In the implemented model, the parameters are set so that results can be interpreted to apply only to train observations.
- Improved method for calculating index values.
- More fine-grained segmentation of the models (18 segments).
- Testing of the effect of other explanatory variables with machine learning. The conclusion from this test was that the inclusion of other variables (in addition to the commuter index and travel time) had a very limited effect on the ability to predict of the day-of-time model.
- Endogenous prediction of day type.
- More flexible functionality for calibration of the implemented trip purpose model.
- More heterogeneity in day-of-time distributions for leisure travel (to a lesser extent work-related trips), by allowing that overall day-of-time distributions are calculated as a weighted average of day-of-time distributions for subcategories (shopping trips, service trips, pick-up and drop-off, visits, leisure and cabin/holiday/nature).

Selected results

The predicted proportion of days in weekend and holidays is for most train station pairs between 10 and 50%.

Table S1 shows the predicted proportion of days in weekend and holidays for some selected train station pairs.

Table S1: Predicted residual daily share for selected stations.

End station	From Oslo S	From Trondheim S	Fra Flå
Tønsberg station	19.0%	43.0%	31.9%
Torp airport	17.0%	34.4%	30.8%
Lillestrøm station	16.3%	40.3%	31.5%
Lillehammer station	21.6%	33.0%	38.5%
Oslo airport	14.2%	22.1%	26.1%
Bergen station	38.8%	39.3%	40.0%
Geilo station	30.8%	44.7%	28.4%
Trondheim airport	33.4%	13.2%	43.0%

The most central explanatory variable for the prediction of the share of weekend/holidays is travel time. For a given start station, the model predicts relatively little variation in the proportion of remaining days between end stations with similar travel time. Predicted proportions for remaining days increase with travel time within usual travel times (less than 10 hours).

Given the type of day, the travel purpose model predicts distribution of 9 travel purposes (work, school, business, shopping, service, pick up and deliver, leisure and holiday / cabin / nature and visits). These 9 travel purposes are aggregated into three travel purposes (work (incl. school), business trips and leisure) when applied to Trenklin.

As part of the implemented travel purpose model in Excel, we have faciliated a calibration of the distribution that leads to a larger proportion of business trips to/from airports and a weaker correlation between distance and work- and business trips. This calibration was necessary to offset some undesirable effects due to the fact that RVU data defines access and egress trips by train to/from airports as (parts of) trips by air

Table S2. shows some model results for aggregate travel purposes for selected train station pairs.

Table S2: Predicted travel purpose distributions for selected train station pairs.

	Workdays after calibration			Weekend/holidays after calibration		
	% work	% business	% leisure	% work	% business	% leisure
From Oslo S to:						
Tønsberg station	64.8%	9.0%	26.2%	40.2%	4.4%	55.3%
Torp airport	39.7%	31.3%	29.1%	22.7%	9.6%	67.7%
Lillestrøm station	73.6%	6.3%	20.0%	53.8%	3.8%	42.4%
Lillehammer station	55.0%	13.1%	31.8%	27.8%	5.5%	66.7%
Oslo airport	12.2%	58.0%	29.8%	11.9%	9.8%	78.3%
Bergen station	15.8%	20.4%	63.9%	3.8%	3.7%	92.5%
Geilo station	34.3%	15.8%	49.9%	9.6%	4.1%	86.2%
Trondheim airport	4.8%	47.9%	47.3%	1.3%	6.7%	92.0%

The day-of-time distributions are estimated based on the reported start time in RVU and travel times by train as defined in the Level-of-Service matrices.

We use the same parametric model for the day-of-time distributions that was used in TØI report 1558: a "mixture of linear regressions" where we let the expected value for each component depend on the explanatory variable travel time. For business travel, we also estimate the relationship between the relative commuter index (which indicates which way the commuting travel flow goes) and the mixture weights (which indicate the relative magnitude of the morning and afternoon rush hours).

Figure S1 shows modeled day-of-time distributions for a selected train station (Oslo S – Tønsberg).

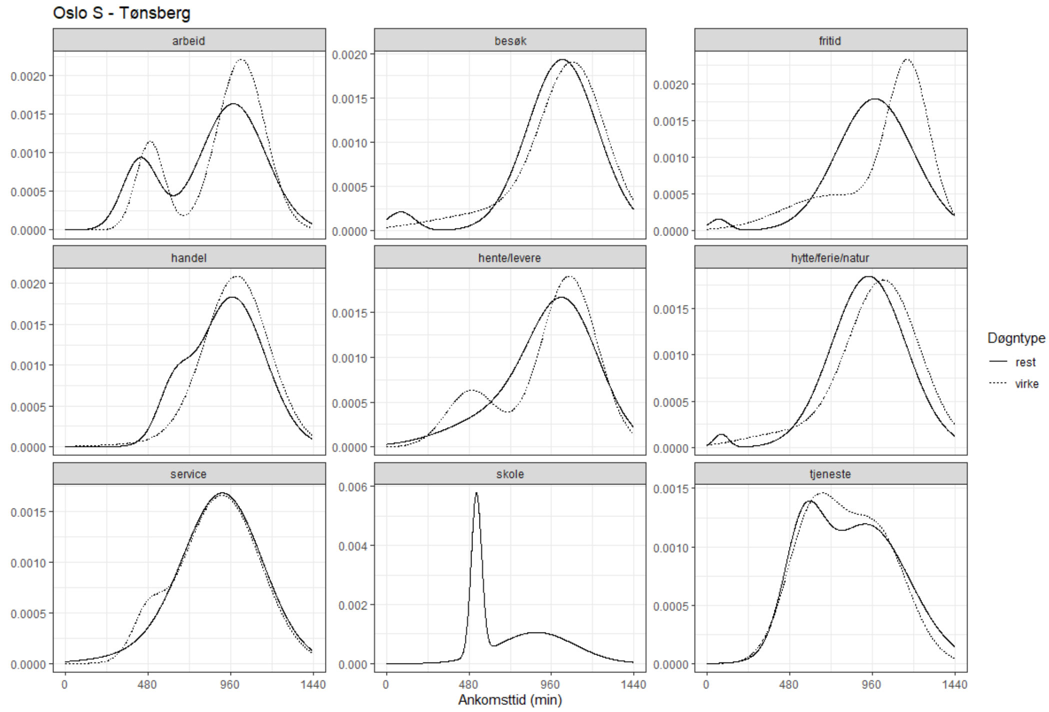

Figure S1: Day-of-time distributions for different trip purposes («abreid»=work, «besøk» = visit, «fritid» = leisure, «handel» = shopiping, «hente/levere» = pick-up and drop-off, «hytte/ferie/natur» =cabin/holiday/nature, «service» = service, «skole» = school, «tjeneste» = business) or ulike for Oslo S–Tønsberg given working days ("virke") and weekend/holiday ("rest").

Uncertainty

As a general element of uncertainty, it must be said that the distributions are estimated on - and thus reflect - "pre-corona" travel behavior. Permanent changes after the pandemic, such as an increased home office, can have a significant impact on future travel purpose and day-of-time distribution.

The model is supposed to reflect systematic differences between train station pairs. It is not expected that the model will be able to capture effects that are specific to particular train stations. It would also not have been possible to estimate with such precision given the relative small data set and a large number of train station pairs the model will be applied to.

The model for travel purpose and day type is estimated on a dataset for trips up to 20 hours (with very few observations over 10 hours), while the model for time-of-day distribution is estimated for trips up to 8 hours. For very long trips (over 8 hours), there is great uncertainty as to whether the model gives reasonable predictions.