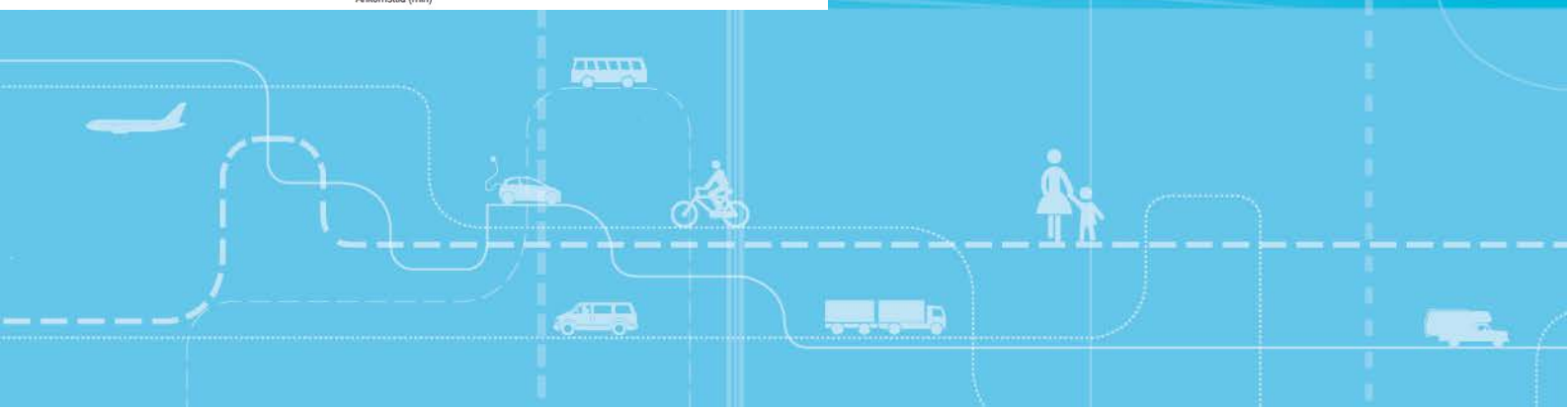
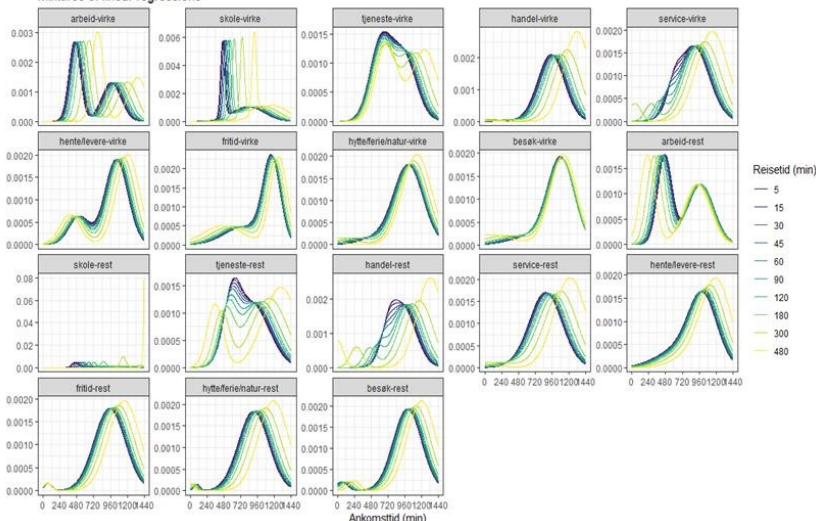


Videreutvikling av reisehensikts- og døgnfordelingsmodeller for togreiser

Mixtures of linear regressions



Videreutvikling av reisehensikts- og døgnfordelingsmodeller for togreiser

Stefan Flügel
Aino Ukkonen
Rikke Ingebrigtsen
Frants Gundersen
Nina Hulleberg

Transportøkonomisk institutt (TØI) har opphavsrett til hele rapporten og dens enkelte deler. Innholdet kan brukes som underlagsmateriale. Når rapporten siteres eller omtales, skal TØI oppgis som kilde med navn og rapportnummer. Rapporten kan ikke endres. Ved eventuell annen bruk må forhåndssamtykke fra TØI innhentes. For øvrig gjelder [åndsverklovens](#) bestemmelser.

Tittel: Videreutvikling av reisehensikts- og døgnfordelingsmodeller for togreiser

Forfattere: Stefan Flügel
Aino Ukkonen
Rikke Ingebrigtsen
Frants Gundersen
Nina Hulleberg

Dato: 02/2022

TØI-rapport: 1876/2022

Sider: 45

ISSN elektronisk: 2535-5104

ISBN elektronisk: 978-82-480-1920-6

Finansieringskilde(r): Jernbanedirektoratet

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

Authors: Stefan Flügel
Aino Ukkonen
Rikke Ingebrigtsen
Frants Gundersen
Nina Hulleberg

Date: 02/2022

TØI Report: 1876/2022

Pages: 45

ISSN: 2535-5104

ISBN Electronic: 978-82-480-1920-6

Financed by: Norwegian Railway Directorate

Prosjekt: 5072 Døgnmodell V2

Prosjektleder: Stefan Flügel

Kvalitetsansvarlig: Anne Madslie

Fagfelt: Transportmodeller

Emneord: Togreiser, reisehensikt, døgnmodeller, døgnstype, Trenklin

Project: 5072 Døgnmodell V2

Project Manager: Stefan Flügel

Quality Manager: Anne Madslie

Research Area: Transport models

Keywords: Train, trip purpose, desired arrival times, Trenklin

Sammendrag:

Fordelinger for reisehensikt og ønsket ankomsttid («døgnfordelinger») er sentrale inndata i transportmodellen for tog, Trenklin. I 2016/2017 etablerte TØI generiske modeller som kan predikere disse fordelinger for alle togstasjonspar i Norge. Denne rapporten dokumenterer en videreutvikling av disse modellene, samt etablering av en modell for predikering av døgnstype (virkedøgn og restdøgn) som er integrert i den nye reisehensiktsmodellen. Modellene er estimert på reisevanedata fra 2019 og representerer derfor en «pre-korona» situasjon. Sentrale forklaringsvariabler i modellen er reisetid og indeksverdier beregnet på sonedata (for reisehensiktsmodellen) og på pendlerstatistikk (for døgnmodellen). Modellen er segmentert etter 2 døgn typer og 9 reisehensikter. Resultatene fra modellen kan brukes som inndata i Trenklin-modellen.

Summary

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. The models are estimated on travel survey data from 2019 and therefore represent a "pre-corona" situation. Key explanatory variables in the model are travel time and index values calculated on zone data (for the travel purpose model) and on commuter statistics (for the 24-hour model). The model is segmented by 2 day-types and 9 travel purposes. The results from the model can be used as input in the Trenklin model.

Language of report: Norwegian

Transportøkonomisk Institutt
Gaustadalléen 21, 0349 Oslo
Telefon 22 57 38 00 - www.toi.no

Institute of Transport Economics
Gaustadalléen 21, N-0349 Oslo, Norway
Telephone +47 22 57 38 00 - www.toi.no

Forord

Fordelinger for reisehensikt og ønsket ankomsttid («døgnfordelinger») er sentrale inndata i transportmodellen Trenklin. Trenklin brukes av Jernbanedirektoratet for analyse av rutetilbud og nytte-kostnadsberegning av ulike infrastrukturprosjekter innenfor jernbanesektoren.

I 2016/2017 etablerte TØI generiske modeller som kan predikere disse fordelinger for alle togstasjonspar i Norge. Denne rapporten dokumenterer en videreutvikling av disse modellene, estimert på reisevanedata fra 2019.

Den nye modellen for hensiktsmodeller muliggjør det å predikere fordeling mellom et totalt antall reiser mellom to togstasjoner i restdøgn og virkedøgn. De største forbedringene i modellen for døgnfordelingen er forbedret metodikk for å beregne pendlerindekser og en mer finkornet segmentering av resultater.

Stefan Flügel har vært prosjektleder og har jobbet sammen med Aino Ukkonen med data-prosessering og videreutvikling av reisehensiktsmodellen. Rikke Ingebrigtsen har videreutviklet modeller for døgnfordelinger. Frants Gundersen har etablert en forbedret beregning av pendlerindeks. Nina Hulleberg har bidratt med innledende databearbeiding.

Modellen er utviklet på oppdrag fra Jernbanedirektoratet. Kontaktperson hos oppdragsgiver har vært Marius Sandvik. Vi takker for gode innspill og godt samarbeid gjennom prosjektet.

Forskningsleder Anne Madslie ved TØI har vært ansvarlig for kvalitetssikring av rapporten. Sekretær Trude Kvalsvik har hjulpet til med ferdigstilling av rapporten. Vi takker også Simen Sørbøe Solbakken som har bidratt med uttesting av å generere døgnfordelinger ved hjelp av maskinlæring.

Oslo, februar 2022

Transportøkonomisk institutt

Bjørne Grimsrud
Administrerende direktør

Kjell W. Johansen
Andelingsleder

Innhold

Sammendrag

Summary

| | | |
|----------|--|-----------|
| 1 | Innledning..... | 1 |
| 1.1 | Bakgrunn..... | 1 |
| 1.2 | Bruk av modellene i praksis..... | 1 |
| 1.3 | Kort om metoden..... | 2 |
| 1.4 | Rapportstruktur..... | 3 |
| 2 | Oppdatert datagrunnlag..... | 4 |
| 2.1 | Reisevanedata (RVU)..... | 4 |
| 2.2 | Pendlerstatistikk..... | 5 |
| 2.3 | Sonedata..... | 6 |
| 2.4 | Dummyvariabler som karakteriserer togstasjoner..... | 7 |
| 2.5 | Reisetider mellom togstasjoner..... | 7 |
| 2.6 | Kundetilfredsundersøkelse..... | 8 |
| 3 | Definisjoner og deskriptive analyser..... | 9 |
| 3.1 | Døgntype..... | 9 |
| 3.2 | Reisehensikt..... | 10 |
| 3.3 | Døgnfordeling..... | 12 |
| 4 | Modell for døgntype- og reisehensikt..... | 19 |
| 4.1 | Estimeringsmodellen..... | 19 |
| 4.2 | Kalibrering og implementering..... | 23 |
| 4.3 | Utvalgte resultater..... | 25 |
| 5 | Modell for døgnfordeling..... | 29 |
| 5.1 | Estimeringsmodellen..... | 29 |
| 5.2 | Estimeringsresultater..... | 30 |
| 5.3 | Implementering og validering..... | 34 |
| 6 | Diskusjon..... | 39 |
| 6.1 | Usikkerhet..... | 39 |
| 6.2 | Kjente svakheter og forbedringsmuligheter..... | 39 |
| 6.3 | Anbefaling for implementering av døgnfordelinger i Trenklin..... | 40 |
| | Referanser..... | 41 |
| | Vedlegg..... | 43 |
| | Vedlegg 1.1 Togstasjoner ved flyplass og turistdestinasjon..... | 44 |
| | Vedlegg 1.2 Sammenligning av simulerte reisehensiktsandeler (før kalibreringen) og RVU-andeler..... | 45 |

Sammendrag

Videreutvikling av reisehensikt- og døgnfordelingsmodeller for togreiser

TØI rapport 1876/2022

Forfattere: Stefan Flügel, Aino Ukkonen, Rikke Ingebrigtsen, Frants Gundersen og Nina Hulleberg

Oslo 2022 45 sider

Fordelinger for reisehensikt og ønsket ankomsttid («døgnfordelinger») er sentrale inndata i transportmodellen for jernbane, Trenklin. I 2016/2017 etablerte TØI generiske modeller som kan predikere disse fordelinger for alle togstasjonspar i Norge. Denne rapporten dokumenterer en videreutvikling av disse modellene, estimert på reisevanedata fra 2019, samt etablering av en modell for predikering av døgnstype som er integrert i den nye reisehensiktsmodellen.

Bakgrunn

Arbeidet er bestilt og finansiert av Jernbanedirektoratet og modellene skal brukes i forbindelse med Trenklin-modellen. Reisehensikts- og døgnstype brukes i Trenklin som segmenteringsvariabel, mens døgnfordelingen (gitt reisehensikt) representerer en fordeling over ønsket ankomsttid og brukes for å beregne skjulte ventetider i Trenklin.

Kort fortalt predikerer modellene hvordan en samlet mengde persontogreiser kan oppdeles i a) døgnstype (restdøgn eller virkedøgn), b) reisehensikter, og c) ønsket ankomsttid (klokkeslett) på endestasjon. Oppdelingen skjer på nivå av togstasjonspar (startstasjon og endestasjon) innad i Norge. Fordelingene varierer altså på relasjonsnivå. Det finnes i underkant av 100 000 slike relasjoner (dvs. togstasjonspar) i Norge.

Metodiske forbedringer

Arbeidet bygger på metodikk etablert ved første versjon av modellene. Følgende forbedringer ble gjort:

- Bruk av oppdaterte data. Sentrale datakilder (RVU, pendlerdata) er fra 2019.
- Større datagrunnlag. Dette er oppnådd ved å inkludere bilturer i datagrunnlaget til døgnmodellen, og bilturer og øvrig kollektivtransport i modellen for reisehensikt og døgnstype. I implementert modell settes parameterne slik at resultater kan tolkes å kun gjelde for togobservasjoner.
- Forbedret metode for beregning av indeksverdier.
- Mer finkornet segmentering i modellene (18 segmenter).
- Uttesting av effekt av øvrige forklaringsvariabler med maskinlæring. Konklusjonen fra denne uttestingen var at inkludering av øvrige variabler (i tillegg til pendlerindeks og reisetid) hadde veldig begrenset effekt på predikeringsevnen til døgnfordelingsmodellen.
- Endogen predikering av døgnstype.
- Mer fleksibelt opplegg for (etter)kalibrering av hensiktsmodellen.
- Mer heterogenitet i døgnfordelinger for fritidsreiser (i mindre grad arbeidsreiser), ved å legge opp til at samlede døgnfordelinger beregnes som vektet gjennomsnitt av døgnfordelinger for underkategorier (handleturer, serviceturer, hente og levere, besøk, fritid og ferie/hytte/natur).

Utvalgte resultater

Predikert andel restdøgn (lørdag, søndag, helligdager og fellesferie) ligger for de fleste relasjoner mellom 10 og 50%.

Tabell S1 viser predikert andel restdøgn for noen utvalgte togstasjonsrelasjoner.

Tabell S1: Predikert restdøgnandel for utvalgte stasjoner.

| Til stasjon | Fra Oslo S | Fra Trondheim S | Fra Flå |
|---------------------|------------|-----------------|---------|
| Tønsberg stasjon | 19,0% | 43,0% | 31,9% |
| Torp stasjon | 17,0% | 34,4% | 30,8% |
| Lillestrøm stasjon | 16,3% | 40,3% | 31,5% |
| Lillehammer stasjon | 21,6% | 33,0% | 38,5% |
| Oslo lufthavn | 14,2% | 22,1% | 26,1% |
| Bergen stasjon | 38,8% | 39,3% | 40,0% |
| Geilo stasjon | 30,8% | 44,7% | 28,4% |
| Trondheim lufthavn | 33,4% | 13,2% | 43,0% |

Den mest sentrale forklaringsvariabelen for andel restdøgn er reisetid. For en gitt startstasjon predikerer modellen forholdsvis lite variasjon i andel restdøgn mellom endestasjoner med lignende reisetid. Predikerte andeler for restdøgn øker med reisetid innenfor vanlige reisetider (under 10 timer).

Gitt døgnstype predikerer reisehensiktsmodellen oppsplitting i 9 reisehensikter (arbeid, skole, tjeneste, handle, service, hente og levere, fritid og ferie/hytte/natur og besøk). Disse 9 reisehensikter aggregeres opp til tre reisehensikter (arbeid (inkl. skole), tjeneste og fritid) ved bruk av Trenklin.

Som et ledd i den implementerte reisehensiktsmodellen i Excel, har vi lagt opp til en kalibrering av fordelingen som fører til en større andel tjenestereiser til flyplasser og en svakere sammenheng mellom distanse og arbeids- og tjenestereiser. Denne kalibreringen var nødvendig for å utligne noen uønskete effekter som følge av at RVU-data definerer tilbringer- og frabringerreiser med tog til/fra flyplasser som (del av) flyreiser.

Tabell S2. viser noen modellresultater for aggregerte reisehensikter for utvalgte togstasjonsrelasjoner.

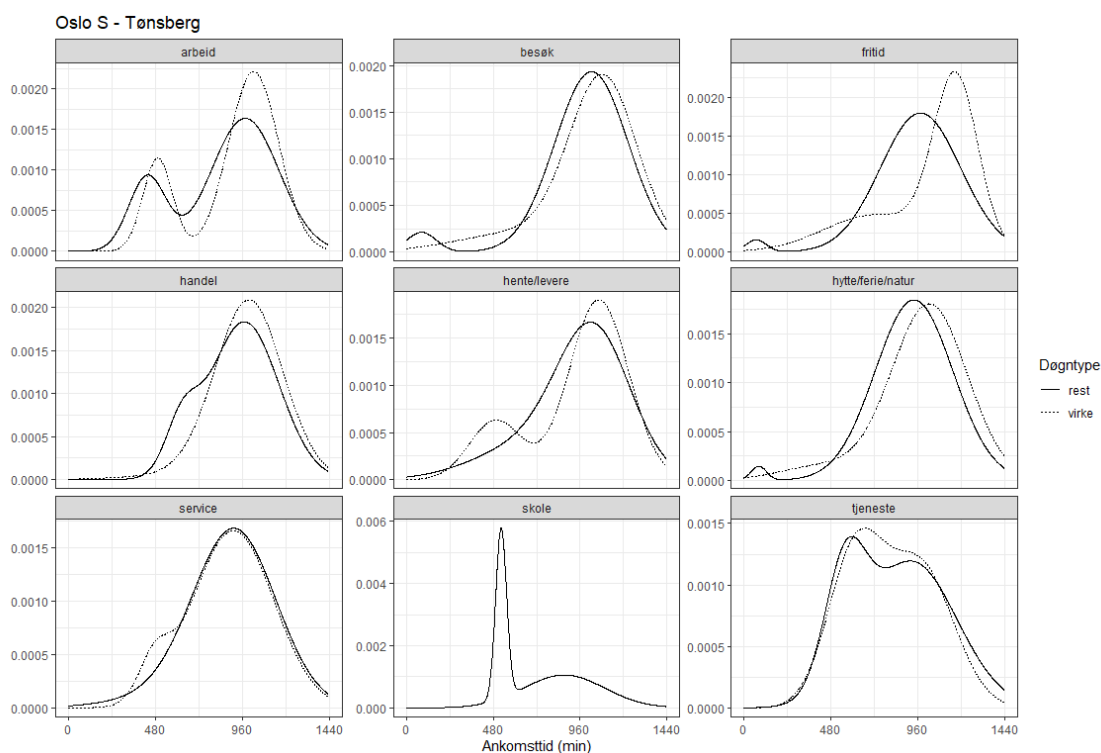
Tabell S2: Predikerte reisehensiktsfordelinger for utvalgte togstasjonsrelasjoner.

| | Virkedøgn etter kalibrering | | | Restdøgn etter kalibrering | | |
|------------------------|-----------------------------|------------|----------|----------------------------|------------|----------|
| | % arbeid | % tjeneste | % fritid | % arbeid | % tjeneste | % fritid |
| Fra Oslo S til: | | | | | | |
| Tønsberg stasjon | 64,8% | 9,0% | 26,2% | 40,2% | 4,4% | 55,3% |
| Torp stasjon | 39,7% | 31,3% | 29,1% | 22,7% | 9,6% | 67,7% |
| Lillestrøm stasjon | 73,6% | 6,3% | 20,0% | 53,8% | 3,8% | 42,4% |
| Lillehammer stasjon | 55,0% | 13,1% | 31,8% | 27,8% | 5,5% | 66,7% |
| Oslo lufthavn | 12,2% | 58,0% | 29,8% | 11,9% | 9,8% | 78,3% |
| Bergen stasjon | 15,8% | 20,4% | 63,9% | 3,8% | 3,7% | 92,5% |
| Geilo stasjon | 34,3% | 15,8% | 49,9% | 9,6% | 4,1% | 86,2% |
| Trondheim lufthavn | 4,8% | 47,9% | 47,3% | 1,3% | 6,7% | 92,0% |

Døgnfordelingene er estimert basert på rapportert starttid i RVU og reisetider med tog som definert i tilbudsmatriser.

Vi benytter samme parametriske modell for døgnfordelingene som ble benyttet i TØI rapport 1558: en «mixture of linear regressions» der vi lar forventningsverdien for hver komponent avhenge av forklaringsvariabelen reisetid. For arbeidsreiser estimerer vi også sammenhengen mellom den relative pendlerindeksen (som angir hvilken vei arbeidsreise-strømmen går) og mixture-vektene (som angir relativ størrelse av morgen- og ettermiddags-rushene).

Figur S1 viser modellerte døgnfordelinger for en utvalgt strekning (Oslo S–Tønsberg).



Figur S1: Døgnfordelinger for ulike reisehensikter på relasjonen Oslo S–Tønsberg, restdøgn og virkedøgn.

Usikkerhet

Som generelt usikkerhetsmoment må det sies at fordelingene er estimert på – og dermed gjenspeiler - «pre-korona» reiseadferd. Varige endringer etter pandemien, som økt hjemmekontor, kan ha betydelig påvirkning på framtidig hensiktsfordeling og døgnfordeling.

Modellen skal gjenspeile systematiske forskjeller mellom togstasjonsrelasjoner. Det forventes ikke at modellen klarer å fange opp effekter som er spesifikke for enkelte togstasjoner. Det hadde heller ikke vært mulig å estimere, med et relativt lite datagrunnlag og et stort antall relasjoner modellen skal anvendes på. Modellen vil derfor ikke treffe på alle togstasjonsrelasjoner.

Modellen for reisehensikt/døgnstype er estimert på et datagrunnlag for reiser inntil 20 timer (med veldig få observasjoner over 10 timer), mens modellen for døgnfordeling er estimert for reiser inntil 8 timer. For veldig lange reiser (over 8 timer) er det stor usikkerhet om modellen gir fornuftige svar.

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 facilitated 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).

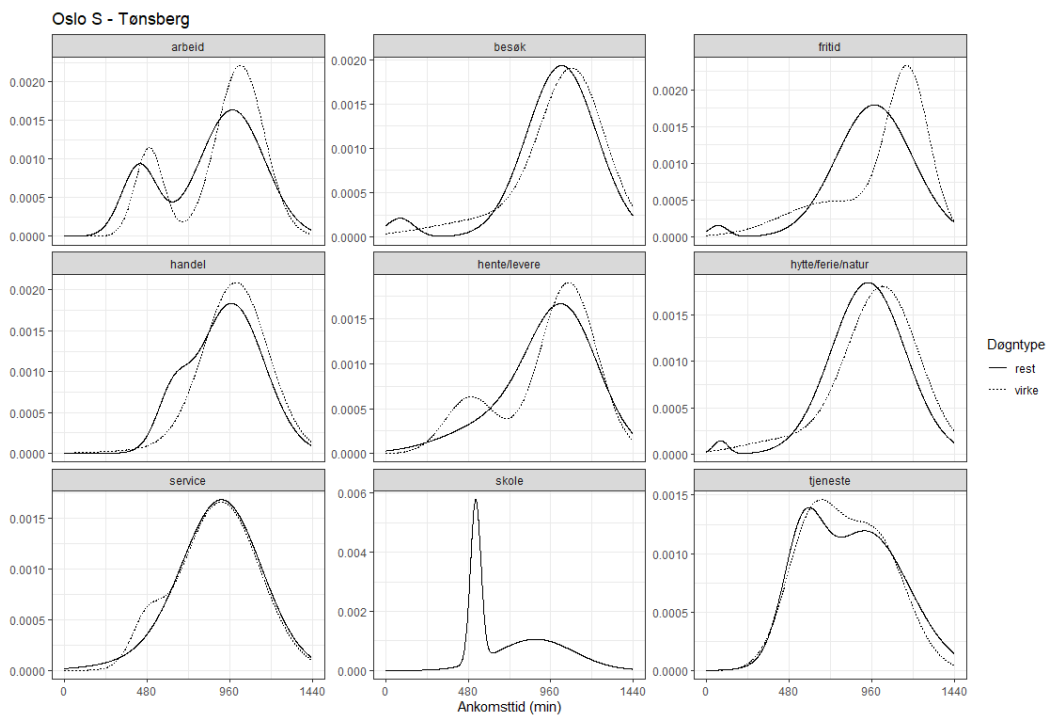


Figure S1: Day-of-time distributions for different trip purposes («arbeid» = work, «besøk» = visit, «fritid» = leisure, «handel» = shopping, «hente/levere» = pick-up and drop-off, «hytte/ferie/natur» = cabin/holiday/nature, «service» = service, «skole» = school, «tjeneste» = business) or unlike for Oslo 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.

1 Innledning

1.1 Bakgrunn

Denne rapporten dokumenterer arbeidet som ble gjort i forbindelse med et prosjekt som hadde som mål å lage oppdaterte modeller for hensikts- og døgnfordeling til bruk i Trenklin. Oppdraget inkluderte også etablering av en ny modell som kan predikere oppsplitting i virkedøgn og restdøgn («døgntype»). Denne modellen («døgntypemodell») er sammenflettet med hensiktsmodellen og bruker samme forklaringsvariabler og parametere. Kort fortalt predikerer modellene hvordan en samlet mengde persontogreiser kan oppdeles i a) døgntype, b) reisehensikter, og c) ønsket ankomsttid på endestasjon (klokkeslett). Oppdelingen skjer på nivå av togstasjonspar (startstasjon og endestasjon) innad i Norge. Fordelingene varierer altså på relasjonsnivå. Det finnes i underkant av 100 000 slike relasjoner (dvs. togstasjonspar) i Norge.

Arbeidet er bestilt og finansiert av Jernbanedirektoratet og modellene skal brukes i forbindelse med Trenklin-modellen (TØI 2013, Ranheim 2017, Ukkonen og Flügel 2020).

Arbeidet bygger på metodikk etablert ved første versjon av modellene (Flügel et al 2017).¹

1.2 Bruk av modellene i praksis

Reisehensikts- og døgntype brukes i Trenklin som segmenteringsvariabel², mens døgnfordelingen (gitt reisehensikt) representerer en fordeling over ønsket ankomsttid og brukes for å beregne skjulte ventetider i Trenklin. Skjult ventetid er differensen mellom ønsket ankomsttid og ruteplanmessig ankomsttid. Det er bare sistnevnte som skal være avhengig av togtilbudet, mens ønsket ankomsttid skal gjenspeile generelle preferanser for tidspunkt å starte opp aktiviteter ved ankomststed.

Tabell 1.1 gir en kort oversikt over de tre modeller som skal etableres.

¹ Døgnmodellen fra andre versjon ble også forsøksvis implementert i RTM-systemet (Flügel og Hamre 2019)

² Reisehensikter håndteres innad i én Trenklin-kjøring, mens det kjøres separate Trenklin-modeller for virkedøgn og restdøgn.

Tabell 1.1: Oversikt over typer modeller.

| Modell | Nivå | Utdata | Inndata | Versjon 1 (2016/2017) | Versjon 2 (2021/2022) |
|----------------------|----------------------------------|--|---|-------------------------------|--|
| Døgntype | OD | Sannsynlighet for virke- og restdøgn (N=2) | Reiselengde, ulike indeksverdier | - | Mulitnomisk logit modell (nested logit modell ble testet) |
| Reisehensikt | OD gitt døgntype | Sannsynlighet for ulike reisehensikter (N=9) | beregnet fra sonedata, dummy variabler for spesielle togstasjoner | Mulitnomimisk logit modell | |
| Døgnfordeling | OD gitt reisehensikt og døgntype | Diskrete fordelinger (N=1440) | Reiselengde, pendlerindeks | Mixture of linear regressions | Mixture of linear regressions, (uttesting med dype nevrale nettverk) |

Trenklin-modellen opererer med tre reisehensikter: arbeidsreiser, tjenestereiser og øvrige reiser. Skolereiser er inkludert i arbeidsreiser, og fritid er en sammensatt gruppe av ulike fritidsreiser (handleturer, serviceturer, hente og levere, besøk, fritid og ferie/ hytte/natur). I dagens Trenklin-modell leses reisehensiktsfordelinger inn som tre matriser, for henholdsvis arbeidsreiser, tjenestereiser og øvrige reiser. Beregningen skjer utenfor selve Excel-filen som Trenklin kjøres i. Dette er i motsetning til døgnmodellen som anvendes direkte i Trenklin. Basert på parameterverdier, matriser for den relative pendlerindeksen og togtider, genereres døgnfordelinger som brukes implisitt³ videre i beregningene.

Se avsnitt 6.3 for en anbefaling om bruk av de nye reisehensikts- og døgnfordelinger i Trenklin.

1.3 Kort om metoden

I dette arbeidet bruker vi samme grunnleggende metode som ved første versjon (Flügel et al 2017).

Metoden kan sammenfattes slik:

- 1) Dataprosessering og etablering av avhengig og uavhengige variabler
- 2) Deskriptiv inspeksjon av fordelingene
- 3) Modellspesifisering og statistisk estimering av parameterne basert på RVU-data
- 4) Implementering av modellen for alle togstasjonspar i Norge (N=99856)⁴
- 5) Rimelighetsvurdering (og etterkalibrering av hensiktsmodellen)

³ Implisitt i den forstand at døgnfordelingen skrives ikke ut innad i Trenklin-modellen. De ble bare generert som del av VBA-koden i Trenklin.

⁴ Implementering av døgnmodellen i Trenklin er ikke en del av dette arbeidet. I avsnitt 6.3 anbefaler vi en framgangsmåte.

Følgende forbedringer er gjort sammenlignet med forrige versjon:

- Bruk av oppdaterte data. Sentrale datakilder (RVU, pendlerdata) er fra 2019.⁵
- Større datagrunnlag. Dette er oppnådd ved å inkludere bilturer i datagrunnlaget til døgnmodellen, og bilturer og øvrig kollektivtransport i modellen for reisehensikt og døgntype.
- Forbedret metode for beregning av indeksverdier.
- Mer finkornet segmentering i modellene (18 segmenter).
- Uttesting av effekt av øvrige forklaringsvariabler med maskinlæring (døgnmodell).⁶
- Endogen predikering av døgntype.
- Mer fleksibelt opplegg for (etter)kalibrering av hensiktsmodellen.
- Mer heterogenitet i døgnfordelinger for fritidsreiser (i mindre grad arbeidsreiser), ved å legge opp til at samlede døgnfordelinger beregnes som vektet gjennomsnitt av døgnfordelinger for underkategorier (handleturer, serviceturer, hente og levere, besøk, fritid og ferie/hytte/natur).

At vi denne gangen inkluderer bilobservasjoner i datagrunnlaget har flere grunner.

For det første øker vi antall observasjoner, noe som gir mer pålitelige parametere for forklaringsvariabler som har lignende effekt for tog og bil. Uten bilobservasjoner hadde vi ikke kunnet opprettholdt et så høyt antall segmenter.

I hensiktsmodellene, estimerer vi dummyparametere for togobservasjoner og kan utnytte visse egenskaper ved logitmodeller. Konkret bruker vi dummyparameterne for å rekalkibrere markedsandeler ved anvendelsen av modellen. Dette gjør at nivået i resultatene er representativt for togreiser selv om datagrunnlaget også inkluderte bilturer.

For døgnmodellene er et viktig poeng at bilreiser ikke er bundet til et rutetilbud.

Observerte ankomsttider for bilturer kan dermed ansees å være godt eget til å gjenspeile generelle planleggingspreferanser, også for togreiser. Med togturer har man en fundamental metodisk utfordring fordi observerte data (i dette tilfelle ankomsttider) om togreisende alltid være betinget på underliggende rutetabeller. Merk også at Trenklin antar samme døgnfordelinger for eksisterende togbrukere og nye togbrukere. Å inkludere bilturer virker dermed også forsvarlig med tanke på senere anvendelse.

1.4 Rapportstruktur

Kapittel 2 dokumenterer oppdatert datagrunnlag.

Kapittel 3 etablerer definisjoner for døgntype, reisehensikt og døgnfordeling og viser deskriptiv statistikk basert på reisevanedata.

Kapittel 4 dokumenterer etablering av døgntype- og reisehensiktsmodellen, og kapittel 5 gjør det samme for døgnfordelingsmodellen.

Kapittel 6 diskuterer begrensninger og kjente svakheter, samt forbedringsmuligheter i modellene.

⁵ Første versjon brukte data fra 2013/2014.

⁶ Konklusjonen fra denne uttestingen var at inkludering av øvrige variabler (i tillegg til pendlerindeks og reisetid) hadde veldig begrenset effekt på predikeringsevnen til modellen.

2 Oppdatert datagrunnlag

2.1 Reisevanedata (RVU)

RVU er den eneste landsdekkende datakilden som inneholder detaljert informasjon om start og sluttsted, reisehensikt og reisetidspunkt. Som ved etablering av versjon 1, bruker vi derfor RVU-data som estimeringsgrunnlag.

I dette prosjektet har vi brukt den nasjonale RVU for år 2019. Reiser i RVU er stedfestet basert på grunnkretser der reisen starter og slutter. Vi har fått en tilpasset RVU-fil der alle personidentifiserende variabler er tatt ut. Dette var en betingelse for å få tilgang til grunnkretser på åttesifret format uansett størrelse på grunnkretsen.

RVU-data inneholder en stor andel respondenter som er trukket fra det såkalte regionale tilleggsutvalget. Det ligger utvalgsvekter i RVU som korrigerer for at dette. Som i første versjon av modellen, brukes ikke utvalgsvektene ved etablering av modellene.⁷

Vi har ekskludert følgende observasjoner fra RVU-datasettet.

- Observasjoner uten gyldig grunnkrets
- Reiser som starter og slutter i samme grunnkrets
- Reiser som fått tilkoblet samme start- og endestasjon (dette inkluderer reiser der det ikke finnes togtilbud, som Nord-Norge)
- Observasjoner med følgende reisehensikt:
 - Luftet hund
 - Gikk/jogget/skitur
 - Sykkeltur
 - Ridetur
 - Biltur
 - MC-tur
 - Båttur
 - Tur med fritidsbåt
 - Kombinasjon av arbeid/tjenestereise og private formål (på samme sted)
- Reiser med følgende transportmidler:
 - Båt og ferje
 - Gang og sykkel
- Togreiser med reisetider ifølge tilbudsdata (se avsnitt 2.5) som er lengre enn 1200 minutter (20 timer)

⁷ Deskriptive analyser viste forholdsvis små forskjeller (Se avsnitt 3.1 og 3.2.) og test-estimeringer viste relativt liten effekt på estimerte parametere. Vi har derfor besluttet å ikke bruke utvalgsvekter i våre analyser. Det er heller ikke sikkert at generelle utvalgsvekter vil passe for togreisende. Det finnes også en pågående faglig diskusjon om hvor godt disse utvalgsvektene egentlig fungerer. Siden vi skal predikere på togstasjonsrelasjonsnivå ville det heller ikke vært mulig å bruke vektene i selve predikeringen (kun ved estimering av de underliggende parameterne).

Etter ekskluderingsene består datasettet av 37345 reiser. I estimeringsmodellen for reisehensikt og døgnstype har vi videre ekskludert observasjoner med manglende info eller verdi 0 for sentrale forklaringsvariabler. Modellen er derfor estimert basert på 37218 observasjoner.

Ved estimering av døgnmodellen ekskluderer vi øvrige transportmidler (bruker kun tog og bilturer) og setter en strengere grense for reiselengde (8 timer). Med dette blir antall observasjoner 29423.

2.2 Pendlerstatistikk

TØI har gjennom et annet prosjekt⁸ tilgang til bosted og arbeidssted for alle registrert med et ansettelsesforhold i Norge. Bostedet og arbeidsstedet er geografisk bestemt med grunnkretsen til boligen og arbeidsstedet til den ansatte. Pendlerstatistikk knyttet til togreiser kan dermed produseres som oversikter over antall bosatte i områder rundt togstasjoner med arbeidsplass i områder rundt andre togstasjoner. Formålet med pendlerstatistikken er å beskrive asymmetrien mellom stasjonspar mht. morgen- og ettermiddagsrush.

Tog på arbeidsreisen vil være mindre relevant desto lenger fra stasjonen en person bor eller jobber. Ved summering av antall pendlere mellom to stasjoner har vi derfor vektet hver person etter den samlede avstanden som boligen og arbeidsplassen ligger fra nærmeste togstasjon. Vektene er justert noe fra forrige modellering (Flügel m.fl. 2017) Inndelingen er mer finkornet og lengre distanser har fått lavere vekter. Dette virker å være mer realistisk, spesielt i byområder (Hansen et al 2019). Vektene er vist i tabell 2.1.

Tabell 2.1: Vekt etter samlet avstand fra bolig og arbeidsplass til nærmeste stasjon.

| Lengde fra bolig til nærmeste stasjon pluss lengde fra arbeidsplass til nærmeste stasjon | Vekt |
|--|------|
| 0-2 km | 1,0 |
| Over 2 til og med 5 km | 0,8 |
| Over 5 til og med 10 km | 0,6 |
| Over 10 til og med 15 km | 0,4 |
| Over 15 til og med 20 km | 0,3 |
| Over 20 til og med 30 km | 0,2 |
| Over 30 til og med 40 km | 0,1 |
| Over 40 til og med 60 km | 0,05 |
| Over 60 til og med 100 km | 0,01 |
| Over 100 km | 0 |

Avstand mellom stasjon og henholdsvis bolig og arbeidssted er beregnet ut fra grunnkrets (som vi har pendlerdata for). For hver grunnkrets i Norge er det ved hjelp av Matrikkelen beregnet et geografisk tyngdepunkt ut fra bygningene som finnes i grunnkretsen. Dette tyngdepunktet er deretter, ved hjelp av et GIS-verktøy, «snappet» til nærmeste kjørbare vei hentet fra Elveg-databasen. Ut fra dette nye punktet er dermed avstand mellom alle grunnkretser beregnet langs kjørbare vei fra Elveg. Elveg inneholder en del feil (f.eks. ved at noen

⁸ NFR-prosjektet *Digitalisation and mobility: Smart and sustainable transport in urban agglomerations (DIGMOB)*. Aggregerte tall er gjort tilgjengelig som pendlingstall for dette modellprosjektet.

vei-lenker ikke henger sammen o.l.), slik at noe manuell retting er foretatt og noen maskinelle rutiner er benyttet for å minimalisere feil. Blant annet er det kjørt reisevei mellom alle grunnkretser begge veier. Der de to reiseveiene gir ulikt resultat velges den korteste.

I forhold til forrige versjon av modelleringen er det innført to justeringer som øker validiteten. For det første er det justert for avtalt arbeidstid for alle ansettelsesforhold. En må anta at de med lav arbeidsbrøk foretar tilsvarende færre arbeidsreiser. De med under 20 timer avtalt per uke får vekt 0,4. De med 20 til 30 timer avtalt arbeidstid per uke får vekt 0,7 og de med mer enn 30 timer per uke vektes 1.

For det andre er det tatt hensyn til om en må bruke ferje fra bostedet til nærmeste togstasjon eller fra togstasjon til arbeidssted. Innenfor rammene av prosjektet har det ikke vært mulig å innarbeide faktisk reisetid med ferjene eller reell ventetid, men siden bruk av ferje representerer en betydelig reduksjon av tilgjengelighet til togstasjonene er det lagt til 20 km i reisevei når ferje benyttes. Det kan nevnes at de absolutt aller fleste pendlere er heltidsansatt og uten behov for ferje, så justeringene som er gjort er ikke radikale.

Antall pendlere fra en stasjon til en annen er dermed personer som bor innenfor omlandet til en stasjon og arbeider i omlandet til en annen stasjon, vektet med avstandsvekt (der avstanden er justert for eventuell ferjebruk) og arbeidstidsvekt. Som tidligere er det gjort noen manuelle justeringer av hvilke grunnkretser som danner omlandet til hvilke togstasjoner. Dette omfatter 193 grunnkretser og gjelder spesielt der en finner to eller flere stasjoner i samme grunnkrets. I tillegg er det noen få steder der avstandsregninger mangler (Elveg har ingen vei i grunnkretsen eller veilenkene er brutt før en kommer til grunnkretsen).

2.3 Sonedata

Sonedata er hentet fra RTM-modellsystemet og inkluderer befolkningsdata, arbeidsplassdata, arealdata, samt data om hoteller, hytter, skoler osv. Sonedata er på grunnkretsnivå og er oversatt til togstasjonsnivå på tilsvarende måte som pendlerdata er vektet sammen fra grunnkretsnivå til togstasjonsnivå (se tabell 2.1 ovenfor). Dette gir indeksverdier per togstasjon for ulike sonedata.

I modelleringen ble det gjort følgende transformasjoner av indeksverdier:

- 1) For togstasjonsrelasjoner beregnes det sum av indeksverdier for start- og sluttstasjon. Dette tilsvarende fremgangsmåten i versjon 1 og bidrar til at hensiktsfordelingen er symmetrisk.⁹
- 2) Vi beregner forholdstallene mellom ulike type sonedata:
 - a. Indeks for ulike arbeidsplasser er delt med indeks for totalt antall arbeidsplasser
 - b. Indeks for skoleplasser, hoteller, hytter er delt med indeks for befolkning
 - c. Indeks for befolkning er delt med indeks for totalt antall arbeidsplasser
- 3) Forholdstallet er videre log-transformert. Siden noen få stasjoner har indeks med verdi 0 for befolkning og/eller arbeidsplasser, innfører vi en (veldig) liten tilleggsterm for å unngå numeriske problemer i beregningen.

⁹ Hvis reisetiden mellom stasjon A og B ikke er identisk med reisetiden mellom B og A vil hensiktsfordelingen ikke være helt symmetrisk.

2.4 Dummyvariabler som karakteriserer togstasjoner

Befolkningsdata er videre brukt til en inndeling av områdetype for start- og sluttstasjon. Hver stasjon får en inndeling i områdetypene by, tettsted eller landlig på basis av befolkningsstørrelse, vist i tabell 2.2.

Tabell 2.2: Områdetype for togstasjon, definert på basis av befolkningsindeks rundt stasjoner.

| Områdetype | Befolkning |
|------------|--------------|
| By | 25000 - |
| Tettsted | 5000 - 25000 |
| Landlig | - 5000 |

For hvert par av stasjoner er det definert fire dummyer som viser hvilken kombinasjon av områdetyper stasjonsparet tilhører. Disse dummyene er definert på basis av stasjonsparets områdetyper, vist i tabell 2.3.

Tabell 2.3: Dummyparametere for stasjonspar definert på basis av områdetype i Tabell 2.2.

| Stasjonspar | Dummy |
|---|----------------|
| by – by | d_by_by = 1 |
| by – tettsted eller tettsted – by | d_by = 1 |
| by – landlig eller landlig – by | d_by = 1 |
| tettsted – tettsted | d_tettsted = 1 |
| tettsted – landlig eller landlig – tettsted | d_tettsted = 1 |
| landlig – landlig | d_landlig = 1 |

Videre har vi definerte dummyvariabler for flyplass og «turisttogstasjoner» basert på en liste over togstasjoner som også ble brukt ved etablering av framskrivingsmodell for Trenklin (Flügel et al 2021). Listen er gjengitt i vedlegg 7.1.¹⁰

Stasjoner som er definert som flyplass og turiststasjoner for noen dedikerte parameterne i hensiktsmodellen (se avsnitt 4.1) og det er også lagt opp egne kalibreringsmulighet for disse stasjoner i implementert modell (se avsnitt 4.2).

2.5 Reisetider mellom togstasjoner

Reisetid mellom togstasjoner inngår som forklaringsvariabel i modellene og informasjonen brukes også ved beregning av ankomsttider i døgnfordelingsmodellen.

Reisetid mellom togstasjoner er et datasett med nasjonale tilbudsdata fra 2015, korrigert i 2019 og oppdatert med nye stasjoner i 2021, levert fra Jernbanedirektoratet. Dette datasettet inneholder fire matriser; ombordtid, ventetid, antall bytter og ventetid ved bytte for hver togstasjonsrelasjon. Fra dette datasettet er ombord og ventetid ved bytte summert for hver stasjonsrelasjon og dette er togtiden som er brukt i modellen. Reisetiden er koblet via stasjons-ID til RVU.

I likhet med versjon 1, inkluderer vi i modelleringen både reisetid og kvadratene av reisetid. Med denne spesifiseringen er sammenhengen mellom reisetid og andel av en gitt

¹⁰ Stasjonen er satt skjønnsmessig og representerer ikke en offisiell list over turiststasjoner definert av Jernbanedirektoratet.

reisehensikt (og i sum reisehensiktsfordeling) noe mer fleksibelt. Kvadratrotten av reisetid er skalert ned med faktor 1000 for å gjøre konvergeringen av estimeringsalgoritmen enklere.

2.6 Kundetilfredsundersøkelse

Vi har fått tilgang til data fra en omfattende spørreundersøkelse som TNS Kantar utfører på bestilling fra Jernbanedirektoratet. Datasettet inkluderer 140244 (gyldige) svar på spørsmål om reisehensikt, derav 13247 fra 2019. I dette prosjektet bruker vi kundeundersøkelsen som et sammenligningsgrunnlag for resultater fra hensiktsmodellen. Svarene i kundeundersøkelsen er segmentert på linjenivå (ikke relasjonsnivå) og inndeling i reisehensikter avviker noe fra definisjonen i RVU. Derfor er det ikke mulig å direkte sammenligne tallene. Undersøkelsen kan likevel gi en god pekepinn på fordelingen av reisehensikter for togreiser.

Tabell 2.4: Andeler for ulike markedsandeler i Kundeundersøkelsen (tall for 2019), totalt og for utvalgte linjer.

| | N | % arbeid | % tjeneste | % fritid |
|------------------------------------|-------|----------|------------|----------|
| Totalt | 13247 | 48,9% | 14,8% | 36,2% |
| L12 Kongsberg - Eidsvoll | 1106 | 50,2% | 21,1% | 28,6% |
| R11 Skien - Eidsvoll | 1223 | 34,5% | 24,8% | 40,7% |
| R10 Drammen - Lillehammer | 1010 | 30,6% | 30,7% | 38,8% |
| L13 Drammen - Dal | 1332 | 53,8% | 9,9% | 36,3% |
| L2 Stabekk - Ski | 1265 | 68,7% | 8,7% | 22,6% |
| L1 Spikkestad - Asker - Lillestrøm | 1909 | 62,0% | 7,3% | 30,8% |
| 41 Oslo - Bergen | 208 | 6,7% | 26,0% | 67,3% |
| 21 Oslo - Trondheim | 149 | 7,7% | 22,6% | 69,7% |
| 50 Kristiansand - Stavanger | 117 | 18,3% | 28,4% | 53,4% |
| 71 Trondheim - Bodø | 83 | 6,5% | 19,1% | 74,4% |

En tendens i dataene er at andel tjenestereiser er høyere for strekninger med flyplasser.¹¹ F.eks. har linje L12 (som stopper ved Oslo Lufthavn) en andel tjenestereiser av 21,1% mens linje L1 har kun en andel av 7,3%.

Langdistansetogene virker naturlignok å ha en lav andel arbeidsreiser.

Merk at alle reiser i kundeundersøkelsen er togreiser og implisitt oppfattet som «helreiser». I RVU derimot er mange togreiser representert som tilbringerreiser (delreiser). Det gjelder spesielt reiser til flyplasser. Se avsnitt 3.2. for en videre diskusjon.

¹¹ Mens langdistansetogene har enda høyere andel tjenestereiser

3 Definisjoner og deskriptive analyser

3.1 Døgntype

Døgntype defineres basert på reisedato i RVU 2019. Det deles inn i restdøgn og virkedøgn. Restdøgn inkluderer:

104 helgedager
 10 helligdager
 3 dager påskeferie
 2 dager juleferie
 5 dager vinterferie (dato for Oslo er brukt)
 5 dager høstferie (dato for Oslo er brukt)
 1 inneklemt fredag
 20 arbeidsdager i juli

Dette er i sum 150 dager (av totalt 365, altså 41,1%) i år 2019.

Tabell 3.1 viser antall observasjoner og andel restdøgn for ulike reisehensikter (se definisjon i neste avsnitt). Andel restdøgn er oppgitt som uvektet og vektet¹² og for alle transportmidler og for tog.

Tabell 3.1: Deskriptive analyser av virkedøgn og restdøgn i datautvalg fra RVU 2019.

| | Antall virkedøgn (uvektet) | Antall restdøgn (uvektet) | Andel restdøgn - alle transportmidler - uvektet | Andel restdøgn - alle transportmidler - vektet | Andel restdøgn tog - uvektet | Andel restdøgn tog - vektet |
|-------------------|----------------------------|---------------------------|---|--|------------------------------|-----------------------------|
| Arbeid | 10053 | 1889 | 15,8% | 14,6% | 13,8% | 13,1% |
| Skole | 1077 | 148 | 12,1% | 10,8% | 12,1% | 11,7% |
| Tjeneste | 1357 | 263 | 16,2% | 14,1% | 10,4% | 10,4% |
| Handle | 3684 | 1905 | 34,1% | 36,1% | 23,3% | 21,6% |
| Service | 2288 | 882 | 27,8% | 24,5% | 19,4% | 11,8% |
| Hente/Leverer | 2269 | 919 | 28,8% | 29,4% | 22,2% | 16,8% |
| Fritid | 2837 | 1914 | 40,3% | 40,1% | 35,4% | 33,4% |
| Ferie/hytte/natur | 764 | 719 | 48,5% | 48,5% | 44,0% | 47,7% |
| Besøk | 2224 | 2153 | 49,2% | 49,1% | 39,1% | 39,8% |
| Alle hensikter | 26553 | 10792 | 28,9% | 28,4% | 19,5% | 19,0% |

Som forventet er restdøgnandelen høyest for Ferie/hytte/natur og Besøk. Restdøgn inkluderer ikke fredager, noe som trolig bidrar til at det er en nokså stor andel virkedøgn for disse reisehensikter.

¹² Ved bruk av utvalgsvektene i RVU («nasjonal_vekt»)

Merk at vi ikke er direkte interessert i restdøgn gitt reisehensikt i denne rapporten. Modellen prøver å predikere restdøgn uansett reisehensikt og reisehensikt gitt restdøgn (og gitt virkedøgn).

3.2 Reisehensikt

Reisehensikt defineres ut fra svar i RVU-data (variabel «q38rc_detaljert_gjoremal» og «q45_START»). Tabell 3.2 gir en oversikt over den anvendte inndelingen.

Tabell 3.2: Gruppering av reisehensikter.

| Reisehensikt i Trenklin | Reisehensikt i estimerte modeller | Detaljert reisehensikt i RVU |
|-------------------------|-----------------------------------|---|
| Arbeid | Arbeid | * Reise til/fra arbeid (inkl. hjemreiser som startet ved et fast oppmøtested på arbeid) |
| Arbeid | Skolereise | * Reise til/fra skole eller studiested/utdanning |
| Tjenester | Tjenestereiser | * Møte utenfor egen arbeidsplass, forhandlinger, salg, innkjøp o.l. * Serviceoppdrag (...) * Kurs, konferanse, kongress * Forhandlinger, salg, innkjøp, messe * Annen forretnings- og tjenestereise * Militærreise (Reiser til/fra tjenestested for vernepliktige) (inkl. hjemreiser som startet «Skole/studiested / Tjenestested arbeid (ikke fast møtested)») |
| Fritid | Handleturer | * Innkjøp av dagligvarer * Innkjøp av større ting/artikler/varer * Andre innkjøp (alle andre innkjøp og shoppingaktiviteter) |
| Fritid | Serviceturer | * Lege, tannlege, fysioterapeut, legevakt, sykehus eller andre medisinske tjenester * Henting/levering av utstyr og/eller ting * Andre ærend/tjenester |
| Fritid | Hente og levere | * Til/fra barnehage/park/dagmamma/skole - barn * Til/fra sports- og idrettsaktiviteter - barn * Til/fra andre fritidsaktiviteter - barn * Til/fra lege, tannlege, legevakt, sykehus eller andre medisinske tjenester - barn * Andre hente-/bringe-/følgereiser – barn * Til/fra fritidsaktiviteter – voksne * Til/fra lege, tannlege, legevakt, sykehus eller andre medisinske tjenester – voksne * Til/fra andre aktiviteter - voksne |
| Fritid | Besøk | * Private besøk (familie, venner eller andre, også sykebesøk) |
| Fritid | Fritid | * Kino, teater, konsert, utstilling eller lignende * Fotballkamp, sportsarrangement mv som tilskuer * Annen aktivitet som publikum * Sport- eller idrettsaktivitet * Musikkaktivitet og annen kulturaktivitet * Organisasjonsarbeid * Annen organisert aktivitet * Gym / treningssenter * Kafé, restaurant, pub eller lignende (inkl hjemreiser som startet «Annet sted i Norge») |
| Fritid | Ferie/ hytte/ natur | * Annen fysisk uteaktivitet * Andre ikke organiserte aktiviteter * Reise til hytte/fritidsbolig * Reise til fritidsbåt/marina * Andre ferie-/fritidsreiser (inkl hjemreiser som startet «bosted 2» eller «Annet land») |

Tabell 3.3 viser reisehensiktsfordeling gitt døgnstype i RVU, både uvektet og vektet.

Tabell 3.3: Hensiktsfordelingen gitt døgnstype i datautvalg fra RVU 2019.

| | Virkedøgn- ikke vektet | Virkedøgn- vektet | Restdøgn - ikke vektet | Restdøgn - vektet | Totalt - ikke vektet | Totalt - vektet |
|-----------------------|---------------------------|----------------------|---------------------------|----------------------|-------------------------|--------------------|
| Arbeid | 37,9% | 38,1% | 17,5% | 16,5% | 32,0% | 32,0% |
| Skole | 4,1% | 4,0% | 1,4% | 1,2% | 3,3% | 3,2% |
| Tjeneste | 5,1% | 5,1% | 2,4% | 2,1% | 4,3% | 4,2% |
| Handle | 13,9% | 13,9% | 17,7% | 19,8% | 15,0% | 15,6% |
| Service | 8,6% | 9,1% | 8,2% | 7,5% | 8,5% | 8,6% |
| Hente/Leverer | 8,5% | 8,5% | 8,5% | 9,0% | 8,5% | 8,6% |
| Fritid | 10,7% | 10,5% | 17,7% | 17,8% | 12,7% | 12,6% |
| Ferie/hytte/natur | 2,9% | 2,6% | 6,7% | 6,2% | 4,0% | 3,6% |
| Besøk | 8,4% | 8,2% | 19,9% | 19,9% | 11,7% | 11,5% |
| Alle hensikter | 100,0% | 100,0% | 100,0% | 100,0% | 100,0% | 100,0% |

Den største forskjellen i hensiktsfordeling gitt døgnstype er at andel arbeidsreiser er høyere i virkedøgn. Som forventet er andelen for fritid, ferie og besøk høyere for restdøgn enn for virkedøgn.

Tabell 3.4 splitter resultater opp etter transportmidler.

Tabell 3.4: Hensiktsfordelingen (uvektet) gitt døgnstype og transportmiddel i datautvalg fra RVU 2019.

| | Virkedøgn- tog (N=1418) | Restdøgn- tog (N=344) | Virkedøgn- bil (N=19808) | Restdøgn- bil (N= 8913) | Virkedøgn- annen kollektiv- transport (N=5327) | Restdøgn- annen kollektiv- transport (N=1535) |
|-----------------------|-------------------------------|-----------------------------|--------------------------------|-------------------------------|--|---|
| Arbeid | 56,2% | 37,2% | 35,1% | 15,5% | 43,4% | 24,8% |
| Skole | 8,7% | 4,9% | 1,3% | 0,4% | 13,2% | 6,0% |
| Tjeneste | 8,5% | 4,1% | 5,1% | 2,4% | 4,4% | 2,3% |
| Handle | 3,9% | 4,9% | 15,9% | 19,5% | 9,0% | 9,8% |
| Service | 3,8% | 3,8% | 9,7% | 8,7% | 5,8% | 6,0% |
| Hente/Leverer | 1,5% | 1,7% | 10,9% | 10,1% | 1,7% | 1,1% |
| Fritid | 9,7% | 21,8% | 9,9% | 16,0% | 14,0% | 26,8% |
| Ferie/hytte/natur | 2,0% | 6,4% | 3,1% | 6,6% | 2,3% | 7,0% |
| Besøk | 5,7% | 15,1% | 9,1% | 20,8% | 6,3% | 16,2% |
| Alle hensikter | 100,0% | 100,0% | 100,0% | 100,0% | 100,0% | 100,0% |

Togreiser har generelt en høyere andel arbeidsreiser og skolereiser enn de andre transportmidlene. Sammenlignet med bil har togreiser betydelig færre handle og hente/leverer reiser. For togreiser har arbeidsreiser høyest andel, også for restdøgn (37,2%).

Tabell 3.5. og tabell 3.6 viser aggregerte reisehensikter med inndelingen som brukes i Trenklin, for henholdsvis virkedøgn og restdøgn.

Tabell 3.5: Aggregert hensiktsfordeling for virkedøgn (uvektet) for togreiser gitt reisetid med tog, samt for relasjoner med flyplass i datantvalget fra RVU 2019.

| Aggregert reisehensikt | alle | 0-30min (N=631) | 30-60 min (N=491) | 60-120 min (N=224) | 2-6 timer (N=56) | 6+ timer (N=27) | Til eller fra flyplasser (N=45) |
|------------------------|-------|--------------------|----------------------|-----------------------|---------------------|--------------------|------------------------------------|
| Arbeid | 64,9% | 64,2% | 71,0% | 66,8% | 26,8% | 33,3% | 57,8% |
| Tjeneste | 8,5% | 6,1% | 6,1% | 14,1% | 30,4% | 18,5% | 17,8% |
| Fritid | 26,6% | 29,7% | 22,9% | 19,1% | 42,9% | 48,1% | 24,4% |

Tabell 3.6: Aggregert hensiktsfordeling for restdøgn (uvektet) for togreiser gitt reisetid med tog samt for relasjoner med flyplass i RVU 2019.

| | alle | 0-30min (N=154) | 30-60 min (N=96) | 60-120 min (N=63) | 2-6 timer (N=30) | 6+ timer (N=14) | flyplass (N=21) |
|----------|-------|--------------------|---------------------|----------------------|---------------------|--------------------|--------------------|
| Arbeid | 42,2% | 50,7% | 40,4% | 44,3% | 17,9% | 0,0% | 28,6% |
| Tjeneste | 4,1% | 2,7% | 4,3% | 4,9% | 7,1% | 7,7% | 0,0% |
| Fritid | 53,8% | 46,6% | 55,3% | 50,8% | 75,0% | 92,3% | 71,4% |

Andel arbeidsreiser for lange distanser kan virke høy og en mulig forklaring er at respondenter i RVU ikke klarer å skille mellom arbeidsreiser og tjenestereiser.

For alle togreiser (uansett reiselengde) ligger andel tjenestereiser på 8,5% (virkedøgn) og 4,1% (restdøgn). Begge tallene er betydelig lavere enn andelen 14,8% som er funnet i kundeundersøkelsen (se avsnitt 2.6).

En grunn til dette er at mange togturer med formål tjenestereise er tilbringerreiser til flyplasser. Disse registreres i RVU som flyturer (med tilbringertransportmiddel tog).

Fra tabell 3.4 ser vi også at andel tjenestereiser i virkedøgn i RVU for togreiser til og fra flyplasser ligger på kun 17,8%. Det er en god del lavere enn andre undersøkelser finner. Ellis og Nordheim (2018) finner at andel tjenestereiser til Gardermoen ligger på 58% med flytoget og 29% med NSB.

En annen tendens som vi observerer er at andel tjenestereiser øker med reiselengden. For korte reiser (under 30 min) ligger andelen i tabell 3.4 på 6,1%, mens den ligger på hele 30,4% for reiser mellom 2 og 6 timer.¹³ Fra kundeundersøkelsen finner vi samme tendens, men effekten virker mye lavere. Den sterke effekten i RVU kan trolig også forklares med at relativt korte tilbringerreiser til flyplasser ikke er registrert som togturer i RVU.

I etterkalibreringen av modellen har vi prøvd å utligne disse effektene (se avsnitt 4.3).

3.3 Døgnfordeling

For senere anvendelse i Trenklin skal døgnfordelinger representere fordelinger av ønsket ankomsttid ved endestasjon. Siden ønsket ankomsttid ikke er observerbar i RVU-data (eller i andre datakilder), bruker vi et anslag for ankomsttid basert på rapportert starttid i RVU («q41_STARTTID») og togtid ifølge tilbudsdata (se avsnitt 2.5).¹⁴

¹³ Antall observasjoner for 6 + timer er noe lavt (N=27) slik at tallene der er mer usikre.

¹⁴ Ved første versjon av døgnmodellen ble det testet ulike alternative beregningsmåter for ankomsttid. Disse hadde relativt liten effekt på resultatet. Vi har derfor valgt den enkleste mulige tilnærming her.

Dette vil ikke være et presis anslag på faktisk ankomsttid, hverken for togreiser eller bilreiser, men det er heller ikke formålet med denne analysen. Formålet er å modellere systematiske forskjeller i (et anslag om) ønsket ankomsttider på tvers av togstasjonsrelasjoner og reisehensikter/døgntyper.

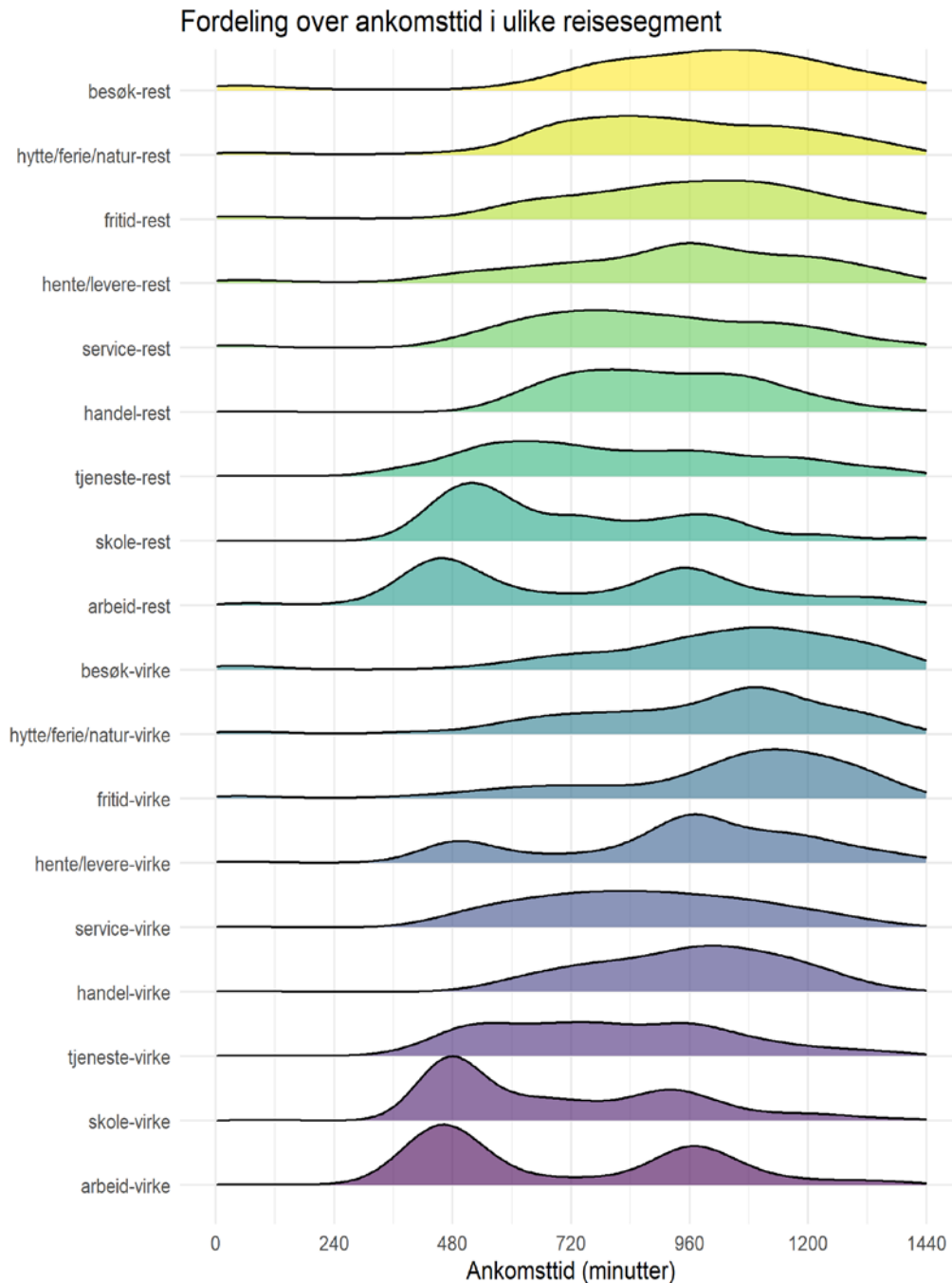
Etter dataprosesseringen som beskrevet i slutten av avsnitt 2.1 har vi informasjon om 29423 reiser fra RVU. Reisene fordeler seg på 18 segmenter som vist i tabellen under (9 reisehensikter for henholdsvis restdøgn («rest») og virkedøgn («virke»)).

Tabell 3.7: Segmenter og antall observasjoner i RVU, estimeringsgrunnlag for døgnfordelingene.

| Segment | Antall |
|---------------------------|--------|
| arbeid - virke | 7609 |
| skole - virke | 358 |
| tjeneste - virke | 1084 |
| handel - virke | 3034 |
| service - virke | 1921 |
| hente/levere - virke | 2115 |
| fritid - virke | 1998 |
| hytte/ferie/natur - virke | 604 |
| besøk - virke | 1801 |
| arbeid - rest | 1477 |
| skole - rest | 51 |
| tjeneste - rest | 220 |
| handel - rest | 1666 |
| service - rest | 751 |
| hente/levere - rest | 872 |
| fritid - rest | 1443 |
| hytte/ferie/natur - restd | 582 |
| besøk - rest | 1837 |

Legg merke til at estimeringsgrunnlaget varierer en god del mellom de ulike segmentene. For eksempel er det (ikke overraskende) få skolereiser på restdøgn.

Vi bruker *kernel density estimation* for å estimere ikke-parametriske fordelinger over ankomsttid i de ulike reisesegmentene. Figur 3.1 viser en empirisk fordeling for hvert segment.



Figur 3.1: Empiriske døgnfordelinger i estimeringsgrunnlaget.

Arbeidsreisene følger helt tydelig en bimodal fordeling med en topp om morgenen og en om ettermiddagen. Skolereiser i virkedøgn og hente/levering i virkedøgn ser også ut å ha to toppe. De andre fordelingene er i all hovedsak unimodale. For å sjekke dette med noe annet enn visuell inspeksjon, bruker vi *Hartigans' Dip Test for Unimodality* implementert i R-pakken **diptest**¹⁵. Nullhypotesen for denne testen er at observasjonene følger en unimodal fordeling. Ved $p < 0.05$ forkaster vi nullhypotesen.

¹⁵ Martin Maechler (2021). diptest: Hartigan's Dip Test Statistic for Unimodality - Corrected. R package version 0.76-0. <https://CRAN.R-project.org/package=diptest>

Tabell 3.8 viser p-verdi med Dip Test for ankomsttidene i hvert segment.

Tabell 3.8: Statistisk test for unimodal døgnfordeling.

| Segment | P-verdi | Unimodal |
|-------------------------|-----------|----------|
| arbeid-virke | 0,0000000 | Nei |
| skole-virke | 0,0110786 | Nei |
| tjeneste-virke | 0,1123057 | Ja |
| handel-virke | 0,4811879 | Ja |
| service-virke | 0,9756358 | Ja |
| hente/levere-virke | 0,0000000 | Nei |
| fritid-virke | 0,8460277 | Ja |
| hytte/ferie/natur-virke | 0,9983048 | Ja |
| besøk-virke | 0,3643726 | Ja |
| arbeid-rest | 0,0000000 | Nei |
| skole-rest | 0,6437710 | Ja |
| tjeneste-rest | 0,9912137 | Ja |
| handel-rest | 0,2286626 | Ja |
| service-rest | 0,9095663 | Ja |
| hente/levere-rest | 0,9936670 | Ja |
| fritid-rest | 0,9601166 | Ja |
| hytte/ferie/natur-rest | 0,7795653 | Ja |
| besøk-rest | 0,1738553 | Ja |

I følge Dip Test er det bare i segmentene arbeid-virke, skole-virke, hente/levere-virke og arbeid-rest at det er støtte for multimodalitet. Dette stemmer med det vi observerer av de empiriske fordelingene i figur 3.1. De fire fordelingene er alle bimodale, med en topp om morgen og en topp om ettermiddagen.

Fra intuisjon, og ut fra erfaringer med første versjonen av døgnmodeller, forventer vi at pendlerstrømmen for en gitt relasjon forklarer størrelsesforholdet mellom morgenreiser og ettermiddagsreiser. Ut fra pendlerdata (se avsnitt 2.2) beregnes en relativ pendlerindeks som representerer pendlerstrømmen.¹⁶ En verdi på 1 betyr at pendlerstrømmen er balansert, dvs. at det bor omtrent like mange personer ved startstasjon og jobber ved endestasjon som det jobber ved startstasjon og bor ved endestasjon. En verdi mindre (større) enn 1 betyr at det bor forholdsvis færre (flere) ved startstasjon og jobber ved endestasjon. Figur 3.2 viser empiriske fordelinger for ulike størrelsesordener av den relative pendlerindeksen (horisontalt) og reisehensikter (vertikalt). I figuren er heltrukne linjer virkedøgn og stiplede linjer restdøgn.

¹⁶ Teknisk sett beregnes den relative pendlerindeksen som raten mellom pendlerindeks mellom startstasjon og sluttstasjon og pendlerindeks mellom sluttstasjon og startstasjon. For å unngå ekstreme verdier som oppstår ved tallverdier som er null eller nær 0 settes en minimumsverdi på 0.05 og en maksimalverdi på 200 for den relative pendlerindeksen.



Figur 3.2: Empiriske døgnfordelinger i estimeringsgrunnlaget for ulike segmenter og intervaller for relativ pendlerindeks. Heltrukne linjer er virkedøgn og stiplede linjer restdøgn. Merk at vi bruker logaritmen til den relative pendlerindeksen i figuren.

For arbeidsreiser (øverste rad i figur 3.2) ser vi en klar sammenheng mellom pendlerindeks og døgnfordeling: Når den relative pendlerindeksen mindre enn 1 (logaritmen er negativ) så er det en høyest topp om ettermiddagen. Det vil si at når flere bor ved endestasjonen og jobber ved startstasjonen enn omvendt vil de fleste observerte ankomsttidene (ved endestasjon) i RVU være om ettermiddagen (det vil si arbeidsreisen hjem). Eksempler på slike stasjonsrelasjoner er Oslo S \rightarrow Moss og Lillestrøm \rightarrow Hakadal.

Når den relative pendlerindeksen er større enn 1 så er det en størst topp om morgenen. Flere bor ved startstasjon og pendler til endestasjonen og i RVU observeres ankomsttid for arbeidsreisene på disse strekningene stort sett om morgenen. Eksempler på relasjoner med denne egenskapen er Melhus \rightarrow Trondheim S og Kambo \rightarrow Nationaltheatret.

Når pendlerindeksen er nær 1 (logaritmen nær 0) er det mer balanse mellom morgen- og ettermiddagstoppe. Dette ser vi i det midterste panelet i den øverste raden (intervallet $[-0.589, 0.589]$). At morgentoppen er noe høyere observerte vi også sist, og det skyldes måten reisehensikt er definert i RVU der reiser hjem fra arbeid blir til handel/service, fritid etc. hvis man gjør et stopp på vei hjem.

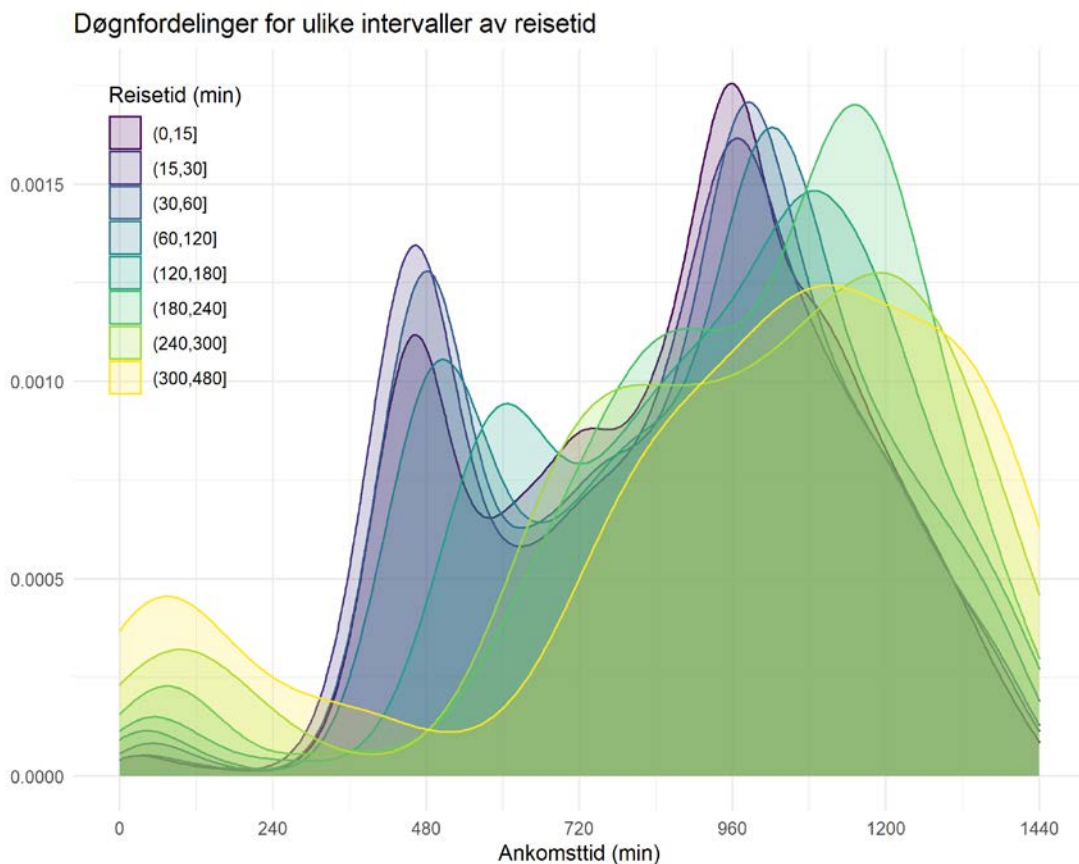
Skolereiser ser ut til å påvirkes noe av pendlerindeksen, men estimeringsgrunnlaget her er tynt sammenliknet med arbeidsreisene (se antall reiser i tabell 3.7). Sammenhengen er veldig svak eller ikke-eksisterende for de øvrige reisebensikter.

Det er naturlig at reisetiden mellom togstasjoner påvirker ønsket ankomsttid ved endestasjonen. Dette fant vi i datagrunnlaget som ble benyttet til estimeringen av forrige versjon av døgnfordelingene og det samme viser RVU 2019.

Tabell 3.9 viser antall observasjoner i åtte intervaller av reisetid (togtid). Figur 3.3 viser empiriske fordelinger over ankomsttid for de samme intervallene av reisetid.

Tabell 3.9: Reisetid og antall observasjoner i RVU.

| Reisetid (min) | Antall |
|----------------|--------|
| (0,15] | 13585 |
| (15,30] | 5092 |
| (30,60] | 5532 |
| (60,120] | 2798 |
| (120,180] | 1332 |
| (180,240] | 469 |
| (240,300] | 307 |
| (300,480] | 308 |



Figur 3.3: Empiriske døgnfordelinger for ulike intervaller av reisetid.

Lengre reisetid forskyver fordelingene mot seinere ankomsttider. De aller lengste reisene (reisetid > 180 minutter/3 timer) følger ikke den typiske fordelingen med en morgen- og ettermiddagstopp, men har i stedet en bredere topp om ettermiddagen/kvelden og en liten topp etter midnatt. Dette kan forklares ved at det er færre arbeidsreiser og flere fritidsreiser (fritid, hytte/ferie/natur og besøk) blant de lange reisene enn blant de kortere reisetidene (se tabell 3.10).

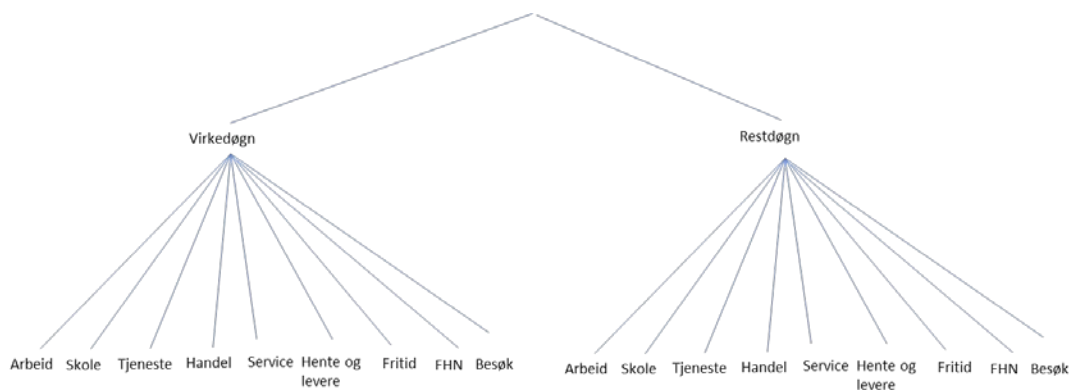
Tabell 3.10: Prosentandel reiser med ulike hensikter.

| Hensikt | Andel reiser med reisetid | |
|-------------------|---------------------------|-----------|
| | ≤ 180 min | > 180 min |
| arbeid | 31,3 | 20,2 |
| skole | 1,4 | 0,6 |
| tjeneste | 4,4 | 6,2 |
| handel | 16,2 | 10,9 |
| service | 9,2 | 6,5 |
| hente/levere | 10,3 | 6,3 |
| fritid | 11,4 | 19,3 |
| hytte/ferie/natur | 3,6 | 14,0 |
| besøk | 12,2 | 16,0 |

4 Modell for døgntype- og reisehensikt

4.1 Estimeringsmodellen

Modellen for reisehensikt og døgntype ble i utgangspunktet lagt opp som en nested logitmodell som illustrert i figur 4.1.



Figur 4.1: Opprinnelig modellstruktur for døgntype og reisehensikt (FHM: Ferie/hytte/natur).

I den endelige modellen er skalaparameterne satt fast til 1, slik at den nestede logitmodellen forenkles til en multinomisk logitmodell.¹⁷

For hvert av de 18 alternativene spesifiseres nyttefunksjoner. Disse funksjonene ble omtalt som genereringsfunksjoner i Flügel et al 2017 siden det økonomiske nyttekonseptet ikke har en intuitiv tolkning her.

Prosessen for utvalget av forklaringsvariabler baserte seg på

- 1) Uttesting av en modell som tilsvarende modellen fra versjon 1.
- 2) Uttesting av flere/nye forklaringsvariabler. For indeksverdier basert på sonedata brukte vi et kriterium om at alle parameterne skal være positive (for å gjøre tolkningen enklere).
- 3) Uttesting av hvilke parametere som burde få separate verdier for tog og «ikke-tog» og for restdøgn og virkedøgn.
- 4) Simulering av markedsandeler gitt (test)modell for å sammenligne predikerte andeler versus andeler i datagrunnlaget. Basert på dette ble det inkludert noen flere parametere som forbedrer modelltreff. Vedlegg 7.2 viser en sammenligning for den endelige estimeringsmodellen.

¹⁷ Tilsvarende nested logitmodell ga verdier for skalaparameterne på 1,03 (virkedøgn) og 1,0 (restdøgn).

Tabell 4.1 dokumenterer forklaringsvariabler og viser i hvilke nyttefunksjoner disse inngår. G står for «generisk» og betyr at forklaringsvariabelen inngår med samme verdi i nyttefunksjonen til både virkedøgn (V) og restdøgn (R). «V/R» betyr at forklaringsvariabelen inngår for begge døgn typer, men med ulike parameterverdier.

Tabell 4.1: Forklaringsvariabler i estimeringsmodellen for døgn type og reisehensikt.

| | arbeid | skole | tjeneste | handle | service | hente/ levere | fritid | FHN* | besøk |
|----------------------------------|--------|-------|----------|--------|---------|------------------|--------|------|-------|
| Konstantledd | R | V/R | V/R | V/R | V/R | V/R | V/R | V/R | V/R |
| dummy tog | R | V/R | V/R | V/R | V/R | V/R | V/R | V/R | V/R |
| reisetid (min) | | G | | G | G | G | V/R | G | V/R |
| reisetid ^2 (1000 min) | | G | | G | G | G | V/R | G | V/R |
| reisetid (min) - tog | | | G | | | | | | |
| reisetid ^2 (1000 min) -tog | | | G | | | | | | |
| reisetid (min) - ikke tog | G | | G | | | | | | |
| reisetid ^2 (1000 min) -ikke tog | G | | G | | | | | | |
| dummy turiststasjon | | | | | | | G | G | |
| dummy by_by | G | | | | | | | | |
| dummy landlig | | | | | | | G | G | |
| dummy flyplass | | G | | | | V/R | | | |
| Andel arbeidspl. in A31VH** | | | | G | | | | | |
| Andel arbeidspl. in A32VH** | | | | G | | | | | |
| Andel arbeidspl. in A33VH** | | | | | | | G | G | |
| Andel arbeidspl. in A43TJE** | | | | | G | | | | |
| Andel arbeidspl. in A71HSOS** | | | | | G | G | | | |
| Andel arbeidspl. in A72HSOS** | | | | | | G | | | |
| Total Arbeidspl. / befolkning** | G | | G | | | | | | |
| Antall i barneskole / bef.** | | | | | | G | | | |
| Antall i skole totalt / bef.** | | G | | | | | | | |
| Antall i Universitet / bef.** | | G | | | | | | | |
| Befolkning /total arb.pl. ** | | | | | | | | | G |
| Antall hytter /befolkning** | | | | | | | | G | |
| Antall hoteller /befolkning** | | | | | | | | G | |

*Ferie/hytte/natur

** log-transformert; variabler er forklart i mer detalj i og under Tabell 4.2.

Merk at konstantleddet og «dummy tog» er normalisert til 0 for arbeidsreiser i virkedøgn. Når det gjelder effekten av reisetid så er det reisetid med tog for arbeidsreiser normalisert. Det innebærer at parameterne til de andre reisehensikter skal se i forhold til dette. Positive (negative) verdier betyr at markedsandelen i forhold til arbeidsreiser vokser (faller) med reisetid. En direkte tolkning vanskeliggjøres allikevel pga. parameteren til kvadratene av reisetid.

Modellen er estimert med maksimum likelihood-metoden i pakken PandalBiogeme.¹⁸

Forklaringskraften (adjusted rho square) til modellen er 0.141. Det er noe lavere enn ved første versjon av modellen. Grunnen til det ligger trolig i at denne estimeringsmodellen bruker observasjoner fra flere transportmidler (og stort sett bruke generiske parametere), mens estimeringsmodellen for første versjon bare ble estimert på togreiser.

Parameterestimaterne og robuste T-verdier er dokumentert i tabell 4.2.

¹⁸ <https://transp-or.epfl.ch/documents/technicalReports/Bier20.pdf>

Tabell 4.2: Estimeringsresultater for modell for døgnstype og reisebent.

| Variabelnavn | Forklaring | Value | Rob. t-test |
|----------------------------|--|---------|-------------|
| ASC02_skole_vd | Konstantledd - skole - virkedøgn | -0,9613 | -3,9344 |
| ASC02_skole_vd_tog | dummy tog - skole - virkedøgn | 0,2706 | 2,1861 |
| ASC03_tje_vd | Konstantledd - tjeneste - virkedøgn | -2,1164 | -47,6583 |
| ASC03_tje_vd_tog | dummy tog - tjeneste - virkedøgn | -0,1709 | -1,0035 |
| ASC04_handle_vd | Konstantledd - handle - virkedøgn | 1,7534 | 8,5025 |
| ASC04_handle_vd_tog | dummy tog - handle - virkedøgn | -1,6948 | -10,4985 |
| ASC05_serv_vd | Konstantledd - service - virkedøgn | -0,9654 | -7,4055 |
| ASC05_serv_vd_tog | dummy tog - service - virkedøgn | -1,3843 | -8,4102 |
| ASC06_hentlev_vd | Konstantledd - hente/levere - virkedøgn | -0,1596 | -0,8378 |
| ASC06_hentlev_vd_tog | dummy tog - hente/levere - virkedøgn | -2,2346 | -9,3747 |
| ASC07_fritid_vd | Konstantledd - fritid - virkedøgn | -0,2057 | -1,3912 |
| ASC07_fritid_vd_tog | dummy tog - fritid - virkedøgn | -0,7202 | -5,8895 |
| ASC08_hytteferie_vd | Konstantledd - ferie/hytte/natur - virkedøgn | -2,0417 | -6,9830 |
| ASC08_hytteferie_vd_tog | dummy tog - ferie/hytte/natur - virkedøgn | -1,0410 | -5,0358 |
| ASC09_besok_vd | Konstantledd - besøk - virkedøgn | -1,7433 | -46,8831 |
| ASC09_besok_vd_tog | dummy tog - besøk - virkedøgn | -0,8772 | -6,1750 |
| ASC10_arb_restd | Konstantledd - arbeid - restdøgn | -1,6562 | -63,6526 |
| ASC10_arb_restd_tog | dummy tog - arbeid - restdøgn | -0,1676 | -1,6972 |
| ASC11_skole_restd | Konstantledd - skole - restdøgn | -2,9429 | -11,5741 |
| ASC11_skole_restd_tog | dummy tog - skole - restdøgn | 0,2723 | 1,0049 |
| ASC12_tje_restd | Konstantledd - tjeneste - restdøgn | -3,7295 | -51,6231 |
| ASC12_tje_restd_tog | dummy tog - tjeneste - restdøgn | -0,7103 | -2,3449 |
| ASC13_handle_restd | Konstantledd - handle - restdøgn | 1,1032 | 5,3246 |
| ASC13_handle_restd_tog | dummy tog - handle - restdøgn | -2,2194 | -8,5483 |
| ASC14_serv_restd | Konstantledd - service - restdøgn | -1,9137 | -14,4838 |
| ASC14_serv_restd_tog | dummy tog - service - restdøgn | -1,8235 | -6,3212 |
| ASC15_hentlev_restd | Konstantledd - hente/levere - restdøgn | -1,0807 | -5,6838 |
| ASC15_hentlev_restd_tog | dummy tog - hente/levere - restdøgn | -2,6426 | -6,4147 |
| ASC16_fritid_restd | Konstantledd - fritid - restdøgn | -0,8155 | -5,4287 |
| ASC16_fritid_restd_tog | dummy tog - fritid - restdøgn | -1,0286 | -7,1574 |
| ASC17_hytteferie_restd | Konstantledd - ferie/hytte/natur - restdøgn | -2,0976 | -7,1285 |
| ASC17_hytteferie_restd_tog | dummy tog - ferie/hytte/natur - restdøgn | -1,2265 | -5,3016 |
| ASC18_besok_restd | Konstantledd - besøk - restdøgn | -1,8212 | -49,1039 |
| ASC18_besok_restd_tog | dummy tog - besøk - restdøgn | -1,3224 | -7,9807 |
| b_A31VH_handle | andel arbeidspl. i detaljhandel med varer som kjøper ofte - handel (log-transformert) | 0,3483 | 5,8895 |
| b_A32VH_handle | andel arbeidspl. i detaljhandel med varer som kjøper sjelden - handel (log-transformert) | 0,4896 | 12,5620 |
| b_A33VH_fri | andel arbeidsplasser salg av mat/overnatting - fritid (log-transformert) | 0,3222 | 7,1084 |
| b_A33VH_hytfer | andel arbeidsplasser salg av mat/overnatting - hytter (log-transformert) | 0,1035 | 1,3523 |
| b_A43TJE_ser | Andel arbeidsplasser i Reisebyrå, turistrelaterte tjenester, post, bank, utleie, etc. - service | 0,0173 | 1,7067 |
| b_A70HSOS_ser | Helse og sosial sektor som ikke inngår under - service (log-transformert) | 0,1496 | 3,2764 |
| b_A71HSOS_hentlev | Alminnelige sykehus (ikke sykehjem/psykiatri), legetjeneste, poliklinikker, tannhelse, barnehager/parker, etc. - hente/levere (log-transformert) | 0,0255 | 2,0199 |
| b_A71HSOS_ser | Alminnelige sykehus (ikke sykehjem/psykiatri), legetjeneste, poliklinikker, tannhelse, barnehager/parker, etc. - service (log-transformert) | 0,0136 | 0,8877 |
| b_A72HSOS_hentlev | SFO og Fritidsklubber - hente/levere (log-transformert) | 0,3434 | 6,6460 |
| b_BSK_hentlev | Barneskole per befolkning - hente/levere (log-transformert) | 0,0214 | 0,4142 |
| b_Bef_bes | Befolkning per arbeidsplass - besøk (log-transformert) | 0,2758 | 6,4613 |
| b_Stotal_sko | Skoleplasser total per befolkning (log-transformert) | 0,5934 | 4,3629 |

| Variabelnavn | Forklaring | Value | Rob. t-test |
|---------------------------|--|---------|-------------|
| b_Uni_sko | Universitetsplasser per befolkning (log-transformert) | 0,0189 | 7,7715 |
| b_arbtot_arb | total arbeidsplasser per befolkning - tjeneste (log-transformert) | 0,2075 | 6,7073 |
| b_arbtot_tje | total arbeidsplasser per befolkning - arbeid (log-transformert) | 0,3253 | 5,7811 |
| b_byby_arb | dummy relasjoner innad/mellom byer - arbeid | 0,1475 | 5,4270 |
| b_flyplass_hentelev | dummy flyplass - hente/levere - virkedøgn | 1,1336 | 7,5453 |
| b_flyplass_hentelev_restd | dummy flyplass - hente/levere - restdøgn | 1,7725 | 10,2713 |
| b_flyplass_tje | dummy flyplass - tjeneste | 0,4122 | 2,3125 |
| b_hot_hytfer | hoteller per befolkning - ferie/hytte/natur | 0,0132 | 2,1351 |
| b_hyt_hytfer | hytter per befolkning - ferie/hytte/natur | 0,0832 | 2,7065 |
| b_landlig_fri | dummy landlige togstasjoner - fritid | 0,1120 | 0,9344 |
| b_landlig_hytfer | dummy landlige togstasjoner - ferie/hytte/natur | 0,2188 | 1,2062 |
| b_togtid_bes | reisetid (min) - besøk - virkedøgn | 0,0054 | 2,7453 |
| b_togtid_bes_restd | reisetid (min) - besøk - restdøgn | 0,0069 | 3,4855 |
| b_togtid_fri | reisetid (min) - fritid - virkedøgn | 0,0029 | 1,4670 |
| b_togtid_fri_restd | reisetid (min) - fritid - restdøgn | 0,0084 | 4,2747 |
| b_togtid_handle | reisetid (min) - handle | 0,0002 | 0,0816 |
| b_togtid_hentlev | reisetid (min) - hente/levere | 0,0013 | 0,6486 |
| b_togtid_hytfer | reisetid (min) - ferie/hytte/natur | 0,0095 | 4,7464 |
| b_togtid_ikketog_arb | reisetid (min) - for andre transportmidler enn tog - arbeid | 0,0039 | 1,9992 |
| b_togtid_ikketog_tje | reisetid (min) - for andre transportmidler enn tog - tjenester | 0,0078 | 3,8974 |
| b_togtid_ser | reisetid (min) - service | 0,0029 | 1,4550 |
| b_togtid_sko | reisetid (min) - skole | 0,0006 | 0,3073 |
| b_togtid_sq_bes | reisetid ² / 1000 (min) - besøk - virkedøgn | 0,0008 | 0,1497 |
| b_togtid_sq_bes_restd | reisetid ² / 1000 (min) - besøk - restdøgn | -0,0008 | -0,1620 |
| b_togtid_sq_fri | reisetid ² / 1000 (min) - fritid - virkedøgn | 0,0034 | 0,6544 |
| b_togtid_sq_fri_restd | reisetid ² / 1000 (min) - fritid - restdøgn | -0,0021 | -0,4063 |
| b_togtid_sq_handle | reisetid ² / 1000 (min) - handle | 0,0056 | 1,0835 |
| b_togtid_sq_hentlev | reisetid ² / 1000 (min) - hente/levere | 0,0039 | 0,7576 |
| b_togtid_sq_hytfer | reisetid ² / 1000 (min) - ferie/hytte/natur | -0,0034 | -0,6483 |
| b_togtid_sq_ikketog_arb | reisetid ² / 1000 (min) - for andre transportmidler enn tog - arbeid | 0,0013 | 0,2514 |
| b_togtid_sq_ikketog_tje | reisetid ² / 1000 (min) - for andre transportmidler enn tog - tjenester | -0,0010 | -0,1978 |
| b_togtid_sq_ser | reisetid ² / 1000 (min) - service | 0,0023 | 0,4379 |
| b_togtid_sq_sko | reisetid ² / 1000 (min) - skole | 0,0049 | 0,9500 |
| b_togtid_sq_tog_tje | reisetid ² / 1000 (min) for tog - tjeneste | -0,0020 | -0,2582 |
| b_togtid_tog_tje | reisetid (min) for tog - tjeneste | 0,0082 | 2,3644 |
| b_turisme_fri | dummy turiststasjoner - fritid | 0,3024 | 2,8245 |
| b_turisme_hytfer | dummy turiststasjoner - Ferie/hytte/natur | 1,2996 | 11,0371 |

Vi ser at flere parameter er statistisk signifikant. Nedenfor omtaler vi signifikante effekter angående indeksvariabler som er basert på sonedata:

- Andel arbeidsplasser i detaljhandel øker sannsynligheten for reiser med reisehensikt Handel (*b_A31VH_handle*, *b_A32VH_handle*)
- Andel arbeidsplasser på alminnelige sykehus, legetjeneste, poliklinikker, tannhelse, barnehager/parker øker sannsynligheten for reiser med reisehensikt Hente/levere (*b_A71HSOS_hentlev*)¹⁹
- Andel arbeidsplasser på øvrig helse og sosial sektor øker sannsynligheten for reiser med reisehensikt Service (*b_A70HSOS_ser*)

¹⁹ Variabelen har også en forventet positiv effekt for Service, men denne effekten er ikke signifikant

- Andel arbeidsplasser i SFO og Fritidsklubber øker sannsynligheten for reiser med reisehensikt Hente/levere ($b_{A72HSOS_hentlev}$)
- Forholdstallet befolkning/arbeidsplasser øker sannsynligheten for reiser med reisehensikt Besøk (b_{Bef_bes})
- Skoleplasser og universitetsplasser (i forhold til befolkningstallet) øker sannsynligheten for reiser med reisehensikt Skole (b_{Stotal_sko} og b_{Uni_sko})
- Forholdstallet arbeidsplasser/befolkning øker sannsynligheten for reiser med reisehensikt Arbeid og Tjeneste (b_{arbtot_arb} og b_{arbtot_tje})
- Antall hytter og hoteller (relativ til befolkning) øker sannsynligheten for reiser med reisehensikt Ferie/hytte/natur (b_{hot_hytfer} , b_{hyt_hytfer})

Videre er flere av parametrene til dummyvariabler som karakteriserer togstasjonsrelasjoner også statistisk signifikant.

- Turiststasjoner øker sannsynligheten for reiser med reisehensikt Fritid og Ferie/hytte/natur ($b_{turisme_fri}$ og $b_{turisme_hytfer}$) Effekten er klar sterke for Ferie/hytte/natur.
- Flyplassstasjoner øker sannsynligheten for reiser med reisehensikt Tjeneste og Hente/levere ($b_{flyplass_tje}$ og $b_{flyplass_hentelev}$). Sistnevnte effekter er spesielt høyt i restdøgn ($b_{flyplass_hentelev_restd}$)
- Relasjoner fra og til byer/byområder øker sannsynlighet reiser med reisehensikt Arbeid (b_{byby_arb})

4.2 Kalibrering og implementering

Modellen for reisehensikt og døgntype er implementert i en Excel-fil («Hensiktsmodell_V2.xlsx»). Filen er oppbygd i flere ark:

- «Resultater» inneholder predikerte andeler for restdøgn/virkedøgn, reisehensiktsfordeling gitt døgntype og en samlet reisehensiktsfordeling for alle togstasjonsrelasjoner (N=99857)²⁰.
- «Inndata» inneholder dataverdier til alle forklaringsvariabler.
- «Parameter» inneholder verdier til alle modellparameterne. Parameterne farget i grå er ikke i bruk i modellen (dette er parametere som eksklusivt gjelder andre transportmidler enn tog).
- «Kalibrering» inneholder et opplegg for etterkalibrering. Dette er nærmere omtalt nedenfor.
- «Beregninger» inneholder verdier til nyttefunksjoner (med underliggende beregninger) og ubetingete valgsannsynligheter for alle 18 alternativer (2 døgntyper ganget med 9 reisehensikter).
- «Sammenligning» inneholder en sammenligning med kundeundersøkelsen. Som nevnt ovenfor er en direkte sammenligning ikke mulig, men nivået på aggregerte andeler for arbeid, tjeneste og fritid kan brukes i kalibreringen.

Kalibreringsarket er lagt opp slik at brukerne selv kan justere hvordan modellen blir kalibrert. Dette gjøres ved å sette verdier i celleområde B2:E14 («aktive kalibreringsparametere»). De opprinnelige kalibreringsparameterne satt av TØI er dokumentert under

²⁰ Dette inkluderer relasjoner med identisk start- og sluttstasjon. Disse har ingen reell tolkning og brukes ikke videre i Trenklin.

«opprinnelige kalibreringsparametere» (celleområde I2:L14). I filen overlevert til Jernbanedirektoratet er aktive kalibreringsparametere satt likt som de opprinnelige kalibreringsparametere.

Oppsettet er lagt opp til 32 (2x2x2x4) kalibreringsparametere som er inndelt etter 4 dimensjoner:

1. Restdøgn eller virkedøgn
2. Arbeids- eller tjenestereiser
3. Faste effekter eller distanseavhengige effekter
4. Type relasjon (alle relasjoner, relasjoner til/fra flyplasser, relasjoner fra/til Oslo Lufthavn, relasjoner fra/til turisttogstasjoner)

Faste effekter forskyver konstantleddene i nyttefunksjonene mens distanseavhengige effekter forskyver parameteren relatert til variabel togtid.

Merk at kalibreringsparameterne gitt døgntype ikke er gjensidig utelukkende og det inngår «summen av effekter» i beregningen. F.eks. vil den faste effekten av relasjonen Geilo – Oslo Lufthavn være summen av alle verdier for alle 4 typer relasjoner.

Målet med den opprinnelig kalibreringen var å utligne effekten av at noen togreiser (spesielt tilbringerreiser til flyplasser) ikke er registrert som togreiser i RVU (se diskusjon i avsnitt 3.2). Vi tror at dette medførte at den ikke-kalibrerte modellen hadde for lav andel tjenestereiser til flyplasser (og for høy andel arbeidsreiser) og at effekten av reiselengde på tjenestereiser er overdrevet.

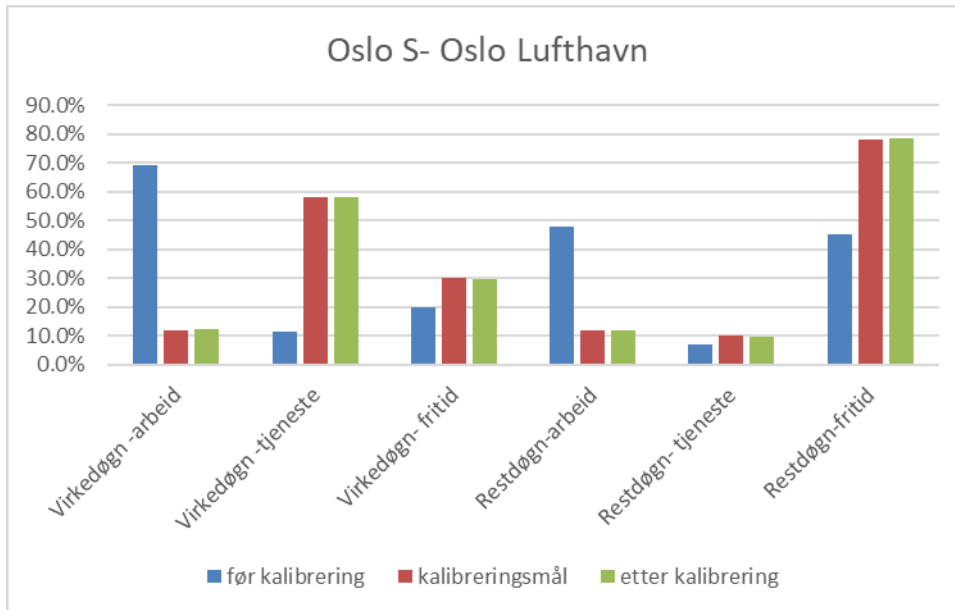
Vi har også en formodning om at respondenter i RVU ikke klarer å skille godt nok mellom tjenestereiser og arbeidsreiser (noen respondenter svarer arbeidsreiser for reiser som egentlig er tjenestereiser og vice-verca). Vi har derfor også valgt å justere distanseeffekten for arbeidsreiser. Dette fører til lavere andel arbeidsreiser for lange distanser sammenlignet med RVU (og sammenlignet med første versjon av hensiktsmodellen). Etter kalibreringen virker resultatene for arbeidsreiser på lange relasjoner å være mer konsistente med resultater fra kundeundersøkelsen.

Videre har vi satt kalibreringsverdier slik at modellens predikasjon mellom Oslo S og Oslo Lufthavn tilsvare (tilnærmet) kalibreringsmålet vi fikk fra Jernbanedirektoratet (figur 4.2). Kalibreringsmålet ble tidligere etablert for anvendelse i Trenklin, basert på resultater av Ellis og Nordheim (2018).

| | Arbeid | Fritid | Forretning |
|-----------|--------|--------|------------|
| Virkedøgn | 12% | 30% | 58% |
| Restdøgn | 12% | 78% | 10% |
| Totalt | 12% | 40% | 48% |

Figur 4.2: Kalibreringsmål for reisebensi- og døgfordeling til Oslo Lufthavn (Kilde: Jernbanedirektoratet)

Figur 4.3 dokumenterer treff mot kalibreringsmålet etter justering av kalibreringsparametere.



Figur 4.3: Kalibrering av bensiktsfordelinger for relasjonen Oslo S - Oslo Lufthavn.

Etter avtale med Jernbanedirektoratet, inneholder de opprinnelige kalibreringsparameterne ingen justering for stasjoner fra/til turisttogstasjoner. Dette fordi vi ikke har datagrunnlag for denne kalibreringen på nåværende tidspunkt.

4.3 Utvalgte resultater

4.3.1 Døgntype

Predikert andel restdøgn ligger for de fleste relasjoner mellom 10 og 50% (tilsvarende vil virkedøgnandelen variere mellom 50 og 90%).

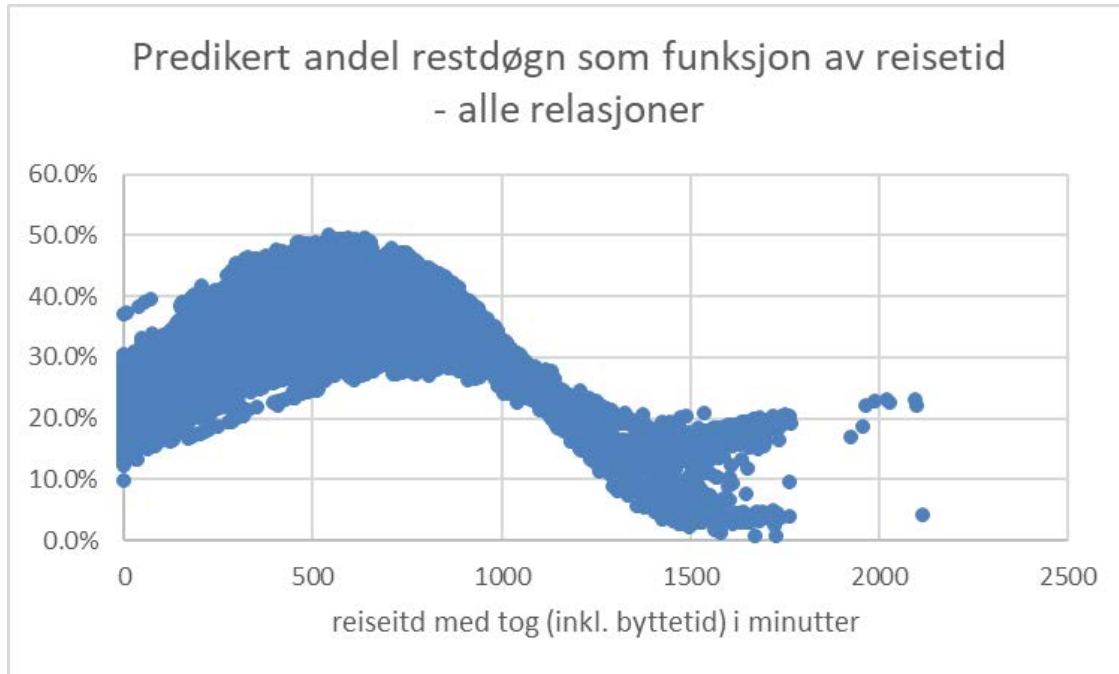
Tabell 4.3 viser predikert andel restdøgn for noen utvalgte togstasjonsrelasjoner.

Tabell 4.3: Predikert restdøgnandel for utvalgte togstasjonsrelasjoner.

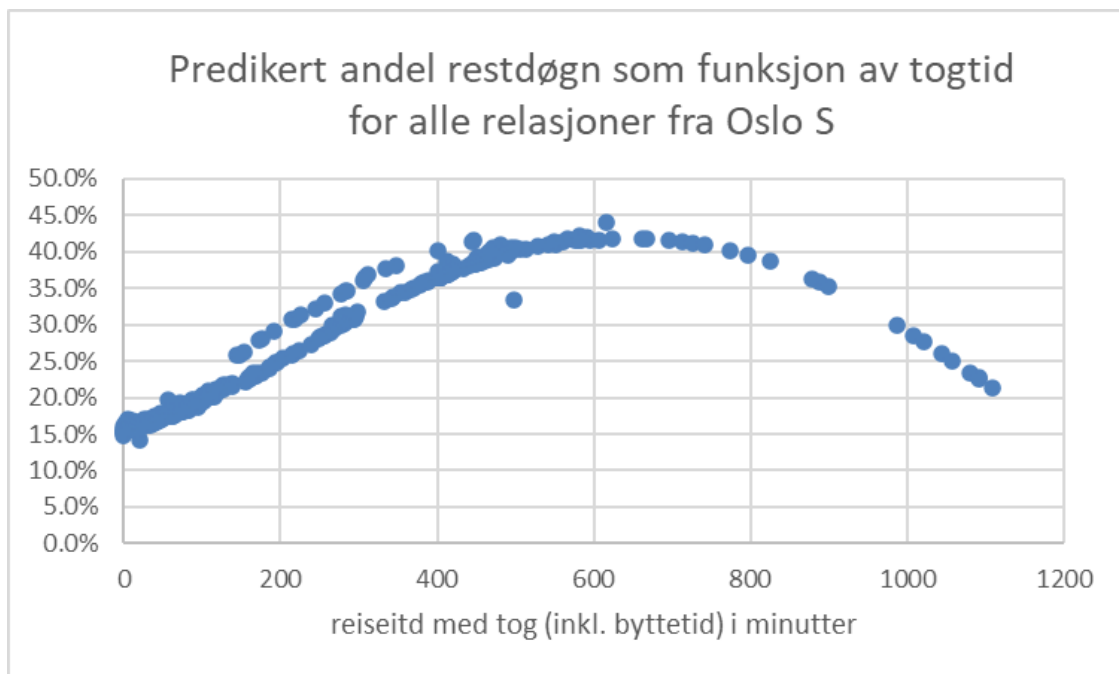
| Til stasjon | Fra Oslo S | Fra Trondheim S | Fra Flå |
|---------------------|------------|-----------------|---------|
| Tønsberg stasjon | 19,0% | 43,0% | 31,9% |
| Torp stasjon | 17,0% | 34,4% | 30,8% |
| Lillestrøm stasjon | 16,3% | 40,3% | 31,5% |
| Lillehammer stasjon | 21,6% | 33,0% | 38,5% |
| Oslo lufthavn | 14,2% | 22,1% | 26,1% |
| Bergen stasjon | 38,8% | 39,3% | 40,0% |
| Geilo stasjon | 30,8% | 44,7% | 28,4% |
| Trondheim lufthavn | 33,4% | 13,2% | 43,0% |

Den mest sentrale forklaringsvariabelen for andel restdøgn er reisetid. For en gitt startstasjon predikerer modellen forholdsvis lite variasjon i andel restdøgn på tvers av endestasjoner med lignende reisetid. Predikerte andeler for restdøgn øker med reisetid innenfor vanlige reisetider (under 10 timer). Ved reisetid over 10 timer er vi utenfor modellens estimeringsgrunnlaget og andel restdøgn reduseres igjen, trolig fordi effekten av kvadratene til reisetiden her slår inn for mye. Dette er et noe uheldig resultat (som er diskutert videre i avsnitt 6.2). I praksis vil dette trolig ha liten betydning siden disse relasjoner har veldig lite

passasjergrunnlag, men vi anbefaler at resultater brukes med forsiktighet hvis modellen brukes for langdistanseforbindelser som er lengre enn Oslo-Bergen eller Oslo-Trondheim. Figur 4.4 og figur 4.5 viser modellens predikerte sammenheng mellom andel restdøgn og reisetid.



Figur 4.4: Modellpredikasjon for restdøgn for alle relasjoner.



Figur 4.5: Modellpredikasjon for restdøgn for alle relasjoner som starter ved Oslo S.

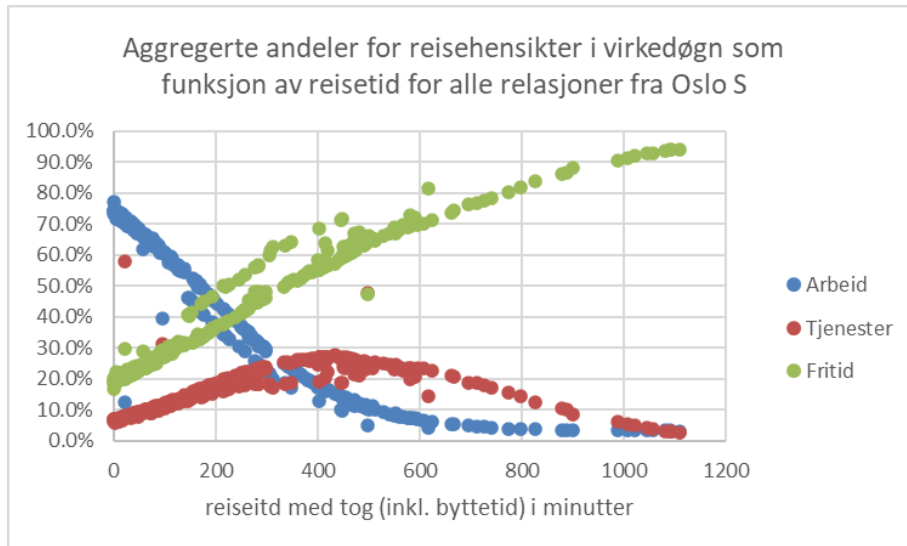
4.3.2 Reisehensikt

Tabell 4.4 viser noen modellresultater for aggregerte reisehensikter for utvalgte togstasjonsrelasjoner (tilsvarende relasjoner i tabell 4.3).

Tabell 4.4: Predikerte reisehensiktsfordelinger for utvalgte togstasjonsrelasjoner.

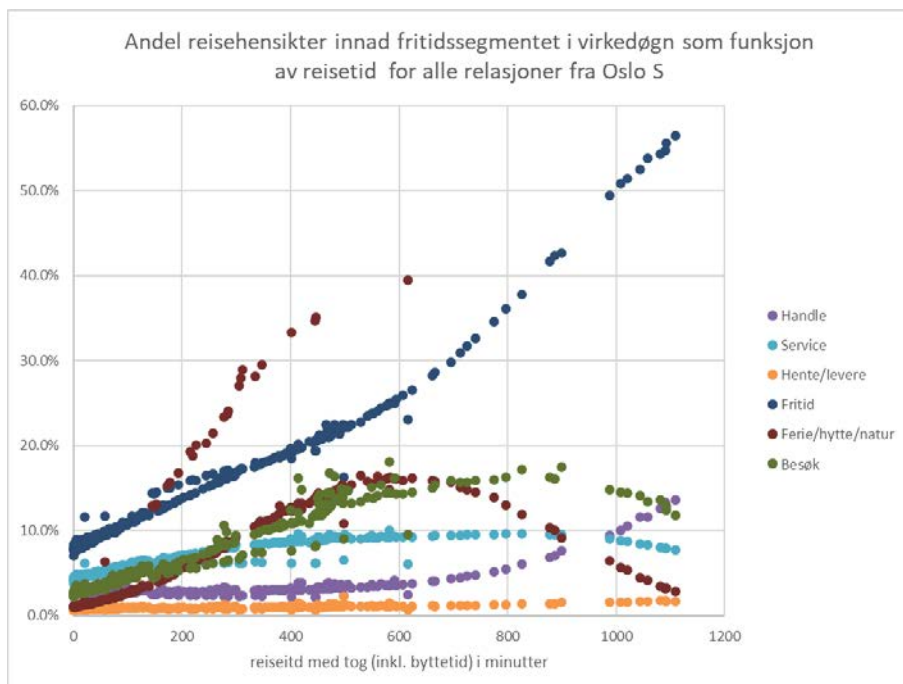
| | Virkedøgn etter kalibrering | | | Restdøgn etter kalibrering | | |
|-----------------------------|-----------------------------|------------|----------|----------------------------|------------|----------|
| | % arbeid | % tjeneste | % fritid | % arbeid | % tjeneste | % fritid |
| Fra Oslo S til: | | | | | | |
| Tønsberg stasjon | 64,8% | 9,0% | 26,2% | 40,2% | 4,4% | 55,3% |
| Torp stasjon | 39,7% | 31,3% | 29,1% | 22,7% | 9,6% | 67,7% |
| Lillestrøm stasjon | 73,6% | 6,3% | 20,0% | 53,8% | 3,8% | 42,4% |
| Lillehammer stasjon | 55,0% | 13,1% | 31,8% | 27,8% | 5,5% | 66,7% |
| Oslo lufthavn stasjon | 12,2% | 58,0% | 29,8% | 11,9% | 9,8% | 78,3% |
| Bergen stasjon | 15,8% | 20,4% | 63,9% | 3,8% | 3,7% | 92,5% |
| Geilo stasjon | 34,3% | 15,8% | 49,9% | 9,6% | 4,1% | 86,2% |
| Trondheim lufthavn stasjon | 4,8% | 47,9% | 47,3% | 1,3% | 6,7% | 92,0% |
| Fra Trondheim S til: | | | | | | |
| Tønsberg stasjon | 6,5% | 16,9% | 76,6% | 1,3% | 2,6% | 96,1% |
| Torp stasjon | 3,3% | 48,6% | 48,1% | 0,8% | 6,5% | 92,7% |
| Lillestrøm stasjon | 13,7% | 20,3% | 66,1% | 3,0% | 3,5% | 93,5% |
| Lillehammer stasjon | 27,8% | 18,9% | 53,3% | 8,3% | 4,4% | 87,2% |
| Oslo lufthavn stasjon | 1,1% | 72,4% | 26,5% | 0,6% | 7,2% | 92,2% |
| Bergen stasjon | 4,8% | 5,9% | 89,3% | 0,8% | 1,2% | 97,9% |
| Geilo stasjon | 2,5% | 10,7% | 86,8% | 0,4% | 1,5% | 98,1% |
| Trondheim lufthavn stasjon | 54,0% | 27,6% | 18,4% | 35,7% | 12,8% | 51,5% |
| Fra Flå til: | | | | | | |
| Tønsberg stasjon | 31,3% | 11,1% | 57,6% | 11,2% | 2,8% | 86,0% |
| Torp stasjon | 17,4% | 28,8% | 53,9% | 5,9% | 4,6% | 89,6% |
| Lillestrøm stasjon | 33,1% | 10,2% | 56,6% | 12,2% | 2,6% | 85,2% |
| Lillehammer stasjon | 16,9% | 12,0% | 71,1% | 4,4% | 2,2% | 93,4% |
| Oslo lufthavn stasjon | 3,4% | 52,5% | 44,1% | 1,6% | 4,2% | 94,2% |
| Bergen stasjon | 14,1% | 13,1% | 72,8% | 3,4% | 2,3% | 94,3% |
| Geilo stasjon | 39,1% | 6,4% | 54,5% | 16,8% | 1,9% | 81,3% |
| Trondheim lufthavn stasjon | 0,8% | 12,1% | 87,1% | 0,2% | 1,1% | 98,7% |

Resultatene virker stort sett logiske. Som for andelen av døgntype i forrige avsnitt, er resultatene mest preget av reisedistanse. Det ser vi tydelig i figur 4.6 som viser fordelinger for aggregerte reisehensikter for virkedøgn for alle relasjoner fra Oslo S.



Figur 4.6: Modellpredikasjon for aggregerte reisehensiktsfordelinger for alle relasjoner som starter ved Oslo S. Virkedøgn.

I figur 4.6 ser vi at det er forholdsvis lite variasjon på tvers av til-stasjoner som har lignende reisetid²¹. Største unntaket er flyplassene som har høy andel tjenestereiser og turiststasjoner som har høyest andel fritidsreiser. Hvis vi deler opp Fritid i de 6 undersegmenter, så ser vi at det er spesielt Ferie/hytte/natur som forklarer dette (figur 4.7).



Figur 4.7: Modellpredikasjon for reisehensikter innenfor fritidssegmentet for alle relasjoner som starter ved Oslo S. Virkedøgn.

Dette er ikke opplagt hvorfor Fritid skal ha størst predikert andel blant de alle lengste relasjoner. Igjen, tyder dette på at modellen oppfører seg noe rart for de lengste distanser.

²¹ Dette til tross for at mange andre forklaringsvariabler har signifikant effekt i estimeringsmodellen (se avsnitt 4.1).

5 Modell for døgnfordeling

5.1 Estimeringsmodellen

Vi benytter samme parametriske modell for døgnfordelingene som i TØI-rapport 1558: En *mixture of linear regressions* der vi lar forventningsverdien for hver komponent avhenge av forklaringsvariabelen reisetid. Modellen er beskrevet i kapittel 5.2 i Flügel et al. 2017 og vi henviser til den rapporten for alle detaljer.

Kort oppsummert er en finite mixture of normals en vektet sum av normalfordelinger der λ beskriver vektingen mellom komponentene, μ er forventningsverdien og σ standardavviket i normalfordelingene.

Vi benytter en mixture bestående av to komponenter, og har dermed at døgnfordelingen kan skrives som

$$f(t) = \lambda_1 N(t; \mu_1, \sigma_1^2) + \lambda_2 N(t; \mu_2, \sigma_2^2), \quad \lambda_1 + \lambda_2 = 1, \quad -\infty \leq t \leq \infty. \quad (1)$$

Videre lar vi forventningsverdiene være beskrevet av regresjonslikningen $\mu = \beta_0 + \beta_1 \times \text{reisetid}$. Det er dette som gjør modellen til en mixture of linear regressions.

For segmentet arbeidsreiser legger vi til et estimeringslag til for kunne parametriske beskrive effekten pendlerindeksen har på vekten mellom morgen- og ettermiddagstoppen, jf. de empiriske fordelingene over ankomsttid i Figur 3.2. Vi lar

$$\lambda_1 = \gamma_0 + \gamma_1 \log(\rho)$$

der ρ er den relative pendlerindeksen for togstasjonsrelasjonen (se kapittel 2.2).

Parameterne γ_0 og γ_1 finnes ved en bootstrap-inspirert algoritme som sorterer observasjonene etter stigende pendlerindeks og tilpasser en standard mixture of normals til subsett av data i en iterativ prosess.

Vi estimerer ett sett av parameterverdier for hvert segment og benytter R-pakken `mixtools`²² og implementasjonen av EM-algoritmen i funksjonene `mixtools::regmixEM` og `mixtools::normalmixEM`.

Mixture-fordelingen i (1) er kontinuerlig, mens i Trenklin benyttes en døgnfordeling diskretisert til hvert minutt i døgnet (1440 minutter). Døgnmodellene gjøres diskrete ved å evaluere (1) for $t = 0, 1, \dots, 1439$ og dele verdiene på en normaliseringskonstant²³ slik at fordelingen summeres til én.

Den oppdaterte versjonen av døgnmodellen skiller seg fra den i Flügel et al. 2017 på følgende punkter:

²² Tatiana Benaglia, Didier Chauveau, David R. Hunter, Derek Young (2009). `mixtools`: An R Package for Analyzing Finite Mixture Models. *Journal of Statistical Software*, 32(6), 1-29. URL <http://www.jstatsoft.org/v32/i06/>.

²³ Normaliseringskonstanten er $k = \sum_{t=0}^{1439} f(t)$.

- 1) Nye fordelinger er estimert på bakgrunn av både bil- og togreiser. Første versjon benyttet kun togreiser.
- 2) På grunn av lite data ble det for første versjon foretatt noen skjønnsmessige vurderinger, især for tjenestereiser. Fordelingen for tjenestereiser ble satt lik fordelingen for arbeidsreiser, men med en tredje komponent lagt til for å fange opp reiser midt på dagen. Med flere reiser i estimeringsgrunnlaget, har vi valgt å ikke legge til en tredje komponent for tjenestereiser i denne versjonen. Vi rapporterer også parameterestimater uten noen form for justering/kalibrering.
- 3) Første versjon består av døgnfordelinger for virkedøgn, for reisehensiktene: arbeid, tjeneste og andre reiser. Den nye versjonen er mer detaljert, og består av ni hensikter for virke- og restdøgn. Vi benytter to komponenter for alle hensikter for konsistens.
- 4) I første versjon valgte vi å fiksure μ_1 (etter estimeringen) slik at denne lokaliseringsparameteren ikke var avhengig av reisetid. I denne versjonen rapporterer vi alle parametere slik at det er mulig å la morgenkomponenten avhenge av reisetid, eller fiksure den i etterkant, slik som sist.
- 5) Notasjonsforskjeller mellom versjonene:
 - a. V1: $\mu_1, \mu_2 = \alpha + \beta \times reisetid$
 - b. V2: $\mu_1 = \beta_{0,1} + \beta_{1,1} \times reisetid, \mu_2 = \beta_{0,2} + \beta_{1,2} \times reisetid$

5.2 Estimeringsresultater

Tabell 5.1 og tabell 5.2 viser parameterverdiene for døgnfordelingene i alle segment for henholdsvis virkedøgn og restdøgn.

Tabell 5.1: Parameterestimater døgnfordelinger. Virkedøgn.

| | arbeid | skole | tjeneste | handel | service | hente / levere | fritid | hytte / ferie / natur | besøk |
|----------|--------|--------|----------|--------|---------|----------------|---------|-----------------------|---------|
| lambda.1 | 0,55* | 0,41 | 0,48 | 0,04 | 0,08 | 0,23 | 0,34 | 0,08 | 0,17 |
| lambda.2 | 0,45* | 0,59 | 0,52 | 0,96 | 0,92 | 0,77 | 0,66 | 0,92 | 0,83 |
| beta0.1 | 442,87 | 454,73 | 601,79 | 799,23 | 612,49 | 515,17 | 777,31 | 555,75 | 734,31 |
| beta1.1 | 0,70 | 0,95 | 0,20 | -2,54 | -1,90 | -0,28 | -0,65 | -1,73 | -1,37 |
| beta0.2 | 965,44 | 818,91 | 928,26 | 942,23 | 870,11 | 1038,66 | 1153,70 | 1005,48 | 1083,46 |
| beta1.2 | 0,80 | 0,74 | 0,65 | 0,78 | 0,57 | 0,35 | 0,22 | 0,26 | 0,08 |
| sigma.1 | 81,68 | 30,18 | 152,51 | 282,00 | 85,66 | 144,87 | 285,72 | 230,22 | 340,79 |
| sigma.2 | 138,19 | 222,53 | 183,13 | 187,51 | 223,24 | 163,42 | 122,21 | 208,16 | 187,91 |

*) Disse verdiene skal erstattes med andre parametere slik at vektene påvirkes av pendlerindeksen. Se tabell 5.3.

Tabell 5.2: Parameterestimer døgnfordelinger. Restdøgn.

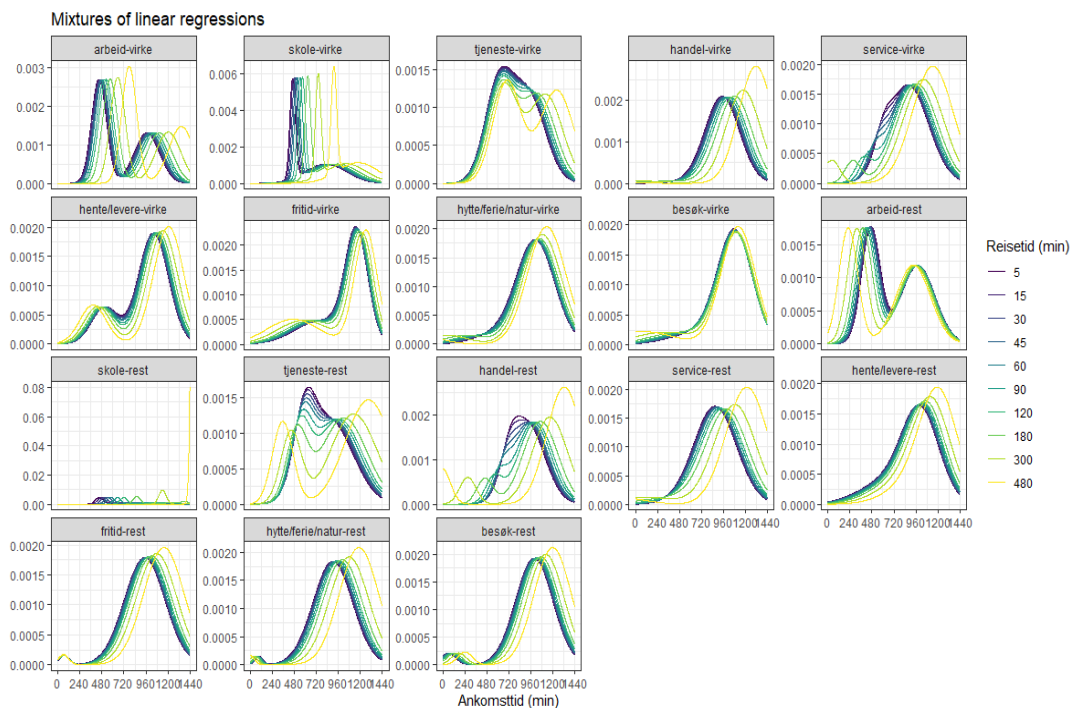
| | arbeid | skole [†] | skole-virke | tjeneste | handel | service | hente /levere | fritid | hytte/ ferie/ natur | besøk |
|----------|--------|--------------------|-------------|----------|--------|---------|---------------|--------|------------------------|--------|
| lambda.1 | 0,45* | 0,44 | 0,41 | 0,26 | 0,13 | 0,09 | 0,28 | 0,02 | 0,02 | 0,04 |
| lambda.2 | 0,55* | 0,56 | 0,59 | 0,74 | 0,87 | 0,91 | 0,72 | 0,98 | 0,98 | 0,96 |
| beta0.1 | 472,03 | 437,00 | 454,73 | 594,27 | 739,83 | 782,36 | 639,52 | 67,98 | 95,71 | 55,48 |
| beta1.1 | -0,52 | 2,35 | 0,95 | -0,50 | -1,56 | -2,21 | 0,75 | -0,01 | -0,20 | 0,39 |
| beta0.2 | 983,81 | 743,21 | 818,91 | 861,05 | 908,19 | 866,76 | 1006,76 | 942,73 | 895,51 | 992,50 |
| beta1.2 | -0,11 | 3,53 | 0,74 | 0,89 | 0,86 | 0,73 | 0,42 | 0,44 | 0,60 | 0,43 |
| sigma.1 | 102,67 | 40,17 | 30,18 | 112,63 | 90,18 | 326,03 | 307,65 | 56,70 | 42,71 | 81,29 |
| sigma.2 | 187,46 | 181,28 | 222,53 | 250,89 | 191,02 | 228,28 | 201,03 | 221,75 | 215,06 | 202,93 |

*) Disse verdiene skal erstattes med andre parametere slik at vektene påvirkes av pendlerindeksen. Se tabell 5.3.

†) Døgnfordelingen for skolereiser restdøgn er estimert på bakgrunn av 51 observasjoner. Dette er for få datapunkter til å gi en pålitelig fordeling over ønsket ankomsttid. Vi anbefaler å erstatte skole-rest med skole-virke.

Legg merke til at vi her har estimert fordelinger for arbeidsreiser uten å ta høyde for den relative pendlerindeksen. Dette kommer vi tilbake til i neste avsnitt.

Figur 5.1 viser døgnfordelinger med parameterverdiene fra tabellene over og for ulike reiselengde.



Figur 5.1: Estimerte døgnfordelinger for ulike segment og varierende reisetid. (Skole-rest erstattes med skole-virke.)

Fordi forventningsverdien er en lineær funksjon av reisetid blir fordelingen over ønsket ankomsttid forskjøvet mot seinere tidspunkt på dagen med økende reisetid.

Fordelingen for skolereiser restdøgn skiller seg ut ved at β_1 er høy for begge komponenter (hvh. 2,35 og 3,53) og fordelingen i Figur 4.8 (første kolonne, tredje rad fra toppen) viser at økende reisetid forskyver døgnfordelingen til urealistisk seine tidspunkt. Denne fordelingen

er estimert basert på svært få observasjoner og vi anbefaler å erstatte den med fordelingen for skolareiser virkedøgn.

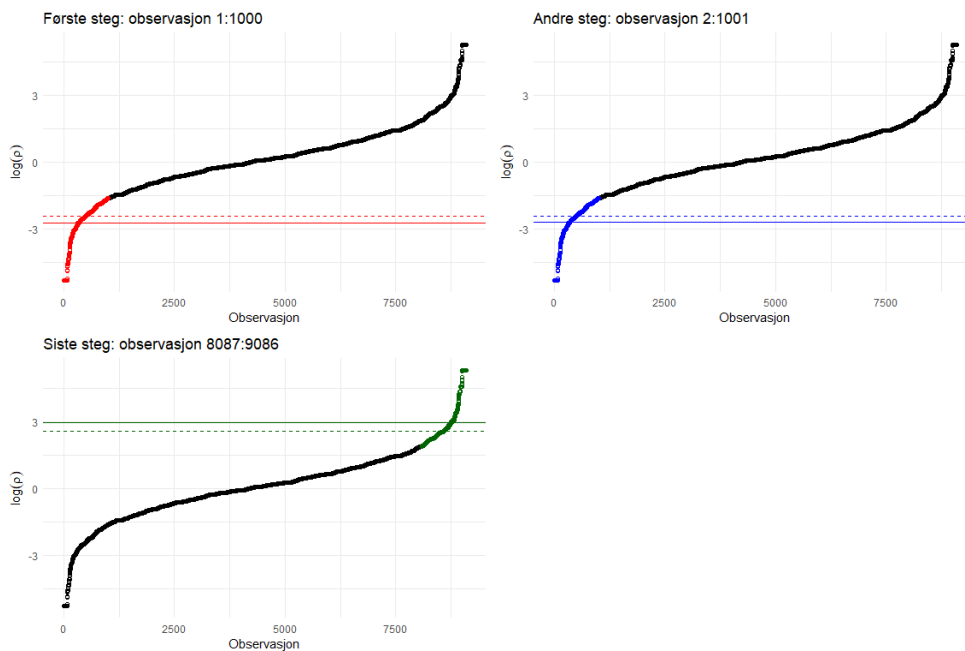
Merk at også andre fordelinger kan bli forskjøvet for mye mot høyre ved lange reisetider. Se siste avsnitt i kapittel 5.3 for forslag til hvordan dette kan håndteres i implementeringen i Trenklin.

Fordelingene for arbeidsreiser som vises i figur 5.1 skal ikke benyttes i den endelige versjonen, da mixture vektene erstattes av parametere som tar høyde for hvilken vei pendlerstrømmen går, slik det blir beskrevet i neste avsnitt.

Arbeidsreiser og relativ pendlerindeks

Den relative pendlerindeksen angir hvilken vei arbeidsreisestrømmen går. Tilsvarende angir vektene λ_1 og λ_2 om tyngdepunktet i mixture fordelingen er om morgenen eller om ettermiddagen. Vi benytter dette til å finne en empirisk sammenheng mellom pendlerindeks og mixture vektene.

Vi estimerer sammenhengen basert på alle arbeidsreiser (virke- og restdøgn). Dette gir et estimeringsdatasett på 9086 reiser. Datasettet blir sortert etter pendlerindeksen og så benyttes et subsett på n_{est} ²⁴ reiser til å estimere vektene med `mixtools::normalmixEM`²⁵. Algoritmen itererer gjennom subsett suksessivt, slik som vist i figur 5.2.

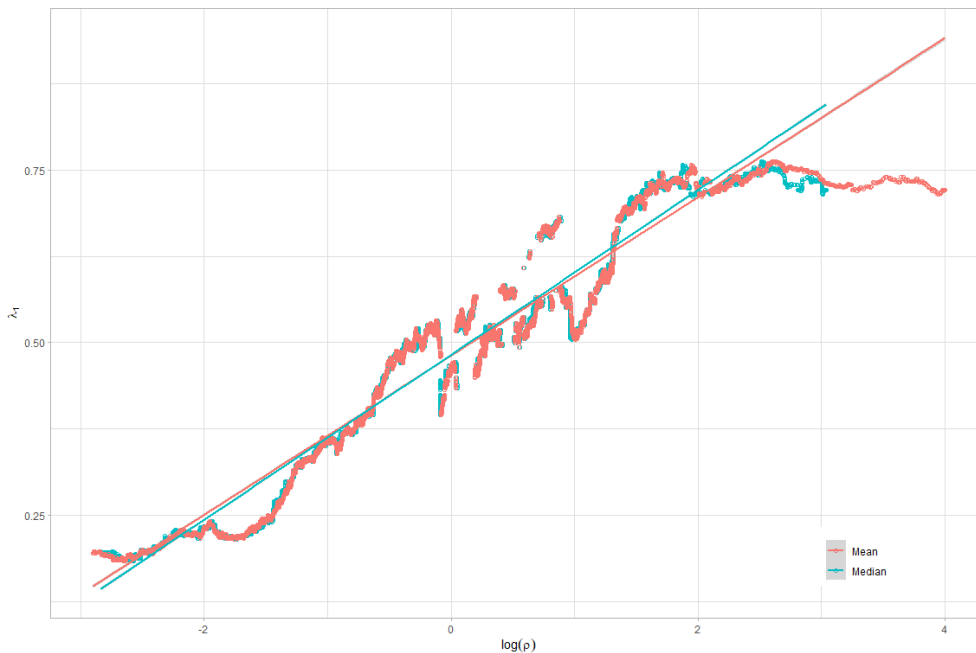


Figur 5.2: Algoritmen itererer gjennom subsett av data. Fargede prikker viser $n_{est} = 1000$ observasjoner i første (røde), andre (blå) og siste (grønne) trinn i algoritmen. Heltrukken horisontal linje er gjennomsnittet av $\log(\rho)$ og stiplede linje er medianen.

I figur 5.3 vises den empiriske sammenhengen mellom $\log(\rho)$ og λ_1 når vi benytter både gjennomsnittsverdien (mean) og medianverdien.

²⁴ Vi har testet med ulike verdier av n_{est} . I Figur 4.9 vises $n_{est} = 1000$, men i resultatene som presenteres er $n_{est} = 600$. Estimaten er stabile for ulike størrelser på subsettet.

²⁵ Vi tar ikke hensyn til reisetid i denne estimeringen.



Figur 5.3: Empirisk sammenheng mellom pendlerindeks og første mixture vekt.

Når pendlerindeksen øker, øker også vekten på morgenkomponenten. Denne egenskapen så vi også i figur 3.2. Tabell 5.3 viser estimat med standardfeil for γ_0 og γ_1 sammen med verdiene vi fant i Flügel et al. 2017.

Vi har tilpasset to varianter av modellen:

- A. Alle parametere estimeres.
- B. Forventning (μ_1, μ_2) og standardavvik (σ_1, σ_2) anses som konstante og er satt til estimatene fra en tilpassing til hele datasettet.

Tabell 5.3: Parameterestimer for første mixture vekt når gjennomsnitt og median av $\log(\rho)$ benyttes. Kolonnen til høyre viser estimat fra første versjon av døgnmodellene.

| | | | Gjennomsnitt | | Median | | 2017 |
|----------|-----------|------------|--------------|--------------|---------|--------------|------|
| | | | Estimat | Standardfeil | Estimat | Standardfeil | |
| A | Intercept | γ_0 | 0,4807 | 0,0005136 | 0,4823 | 0,0004716 | 0,4 |
| | log_rho | γ_1 | 0,1149 | 0,0003990 | 0,1193 | 0,0003776 | 0,2 |
| B | Intercept | γ_0 | 0,5115 | 0,0008902 | 0,5134 | 0,0008537 | |
| | log_rho | γ_1 | 0,1362 | 0,0006916 | 0,1414 | 0,0006835 | |

Det er ikke noen stor forskjell på om vi benytter gjennomsnittsverdien eller medianen, men de estimerte parameterverdiene blir litt høyere når forventning og standardavviket er fiksert (B). Avrundet til ett desimal blir verdiene like og vi benytter i modellene at:

$$\gamma_0 = 0,5 \text{ og } \gamma_1 = 0,1.$$

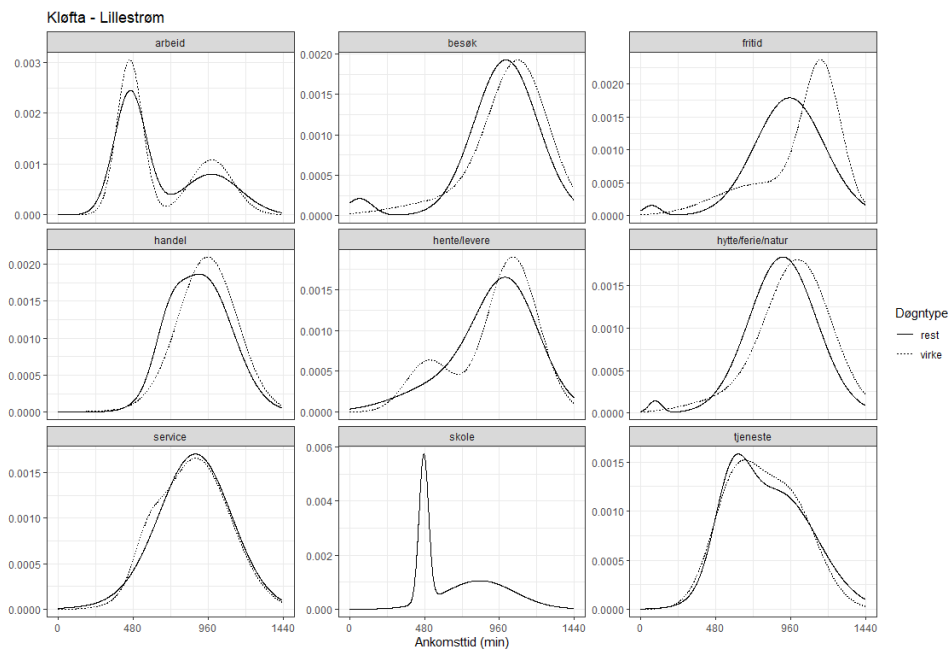
Dette gir symmetri mellom morgen- og ettermiddagskomponenten når den relative pendlerindeksen er 1. Videre er økningen i vekten på morgenkomponenten (morgenrushet) noe svakere med økende pendlerindeks enn det vi fant i forrige versjon, men dette skyldes nok at pendlerindeksen beregnes på en ny måte slik at estimatene ikke er direkte sammenliknbare.

For å sikre at mixture vektene er mellom 0 og 1, benytter vi følgende likning for λ_1

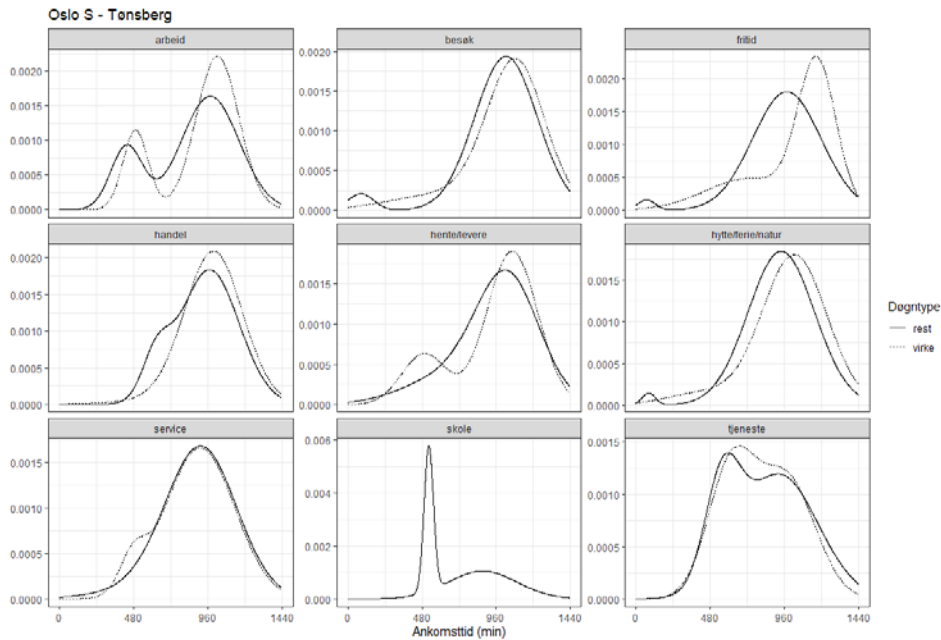
$$\lambda_1 = \min \left\{ 1, \max \{ 0, 0.5 + 0.1 \log \rho \} \right\}, \lambda_2 = 1 - \lambda_1.$$

5.3 Implementering og validering

Når døgnfordelingene skal benyttes i Trenklin vil de vektet sammen basert på andeler estimert med hensiktsmodellen (se kapittel 4). Hvordan hver enkelt døgnfordeling gitt segment ser ut er dermed mindre viktig enn denne vektede hovedfordelingen. Vi viser noen eksempler for utvalgte stasjonsrelasjoner.



Figur 5.4: Døgnfordelinger for ulike segmenter på relasjonen Kløfta–Lillestrøm.

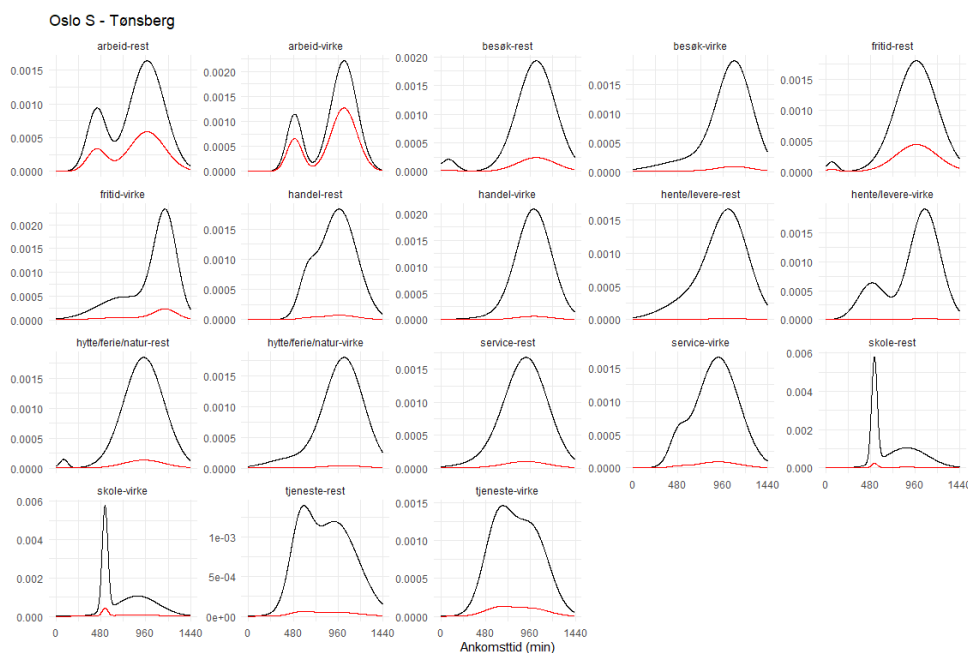


Figur 5.5: Døgnfordelinger for ulike segmenter på relasjonen Oslo S-Tønsberg.

Døgnfordelingene for alle reisehensikter bortsett fra arbeid, påvirkes kun av reisetiden. Fordelingen for arbeidsreiser påvirkes i tillegg av den relative pendlerindeksen og vi ser at på relasjonen Oslo S → Tønsberg er ettermiddagstoppen større enn morgentoppen, mens det er omvendt på relasjonen Kløfta → Lillestrøm.

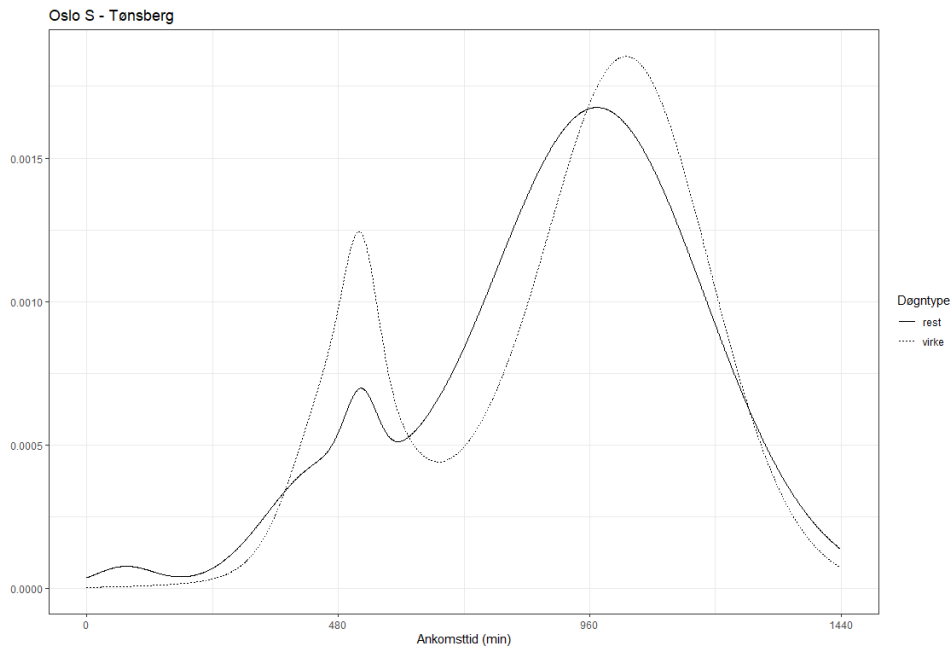
Variasjon mellom stasjonsrelasjoner oppnås ved ulik vektning mellom de 18 døgnfordelingene basert på estimerte hensiktsandeler (se 6.3). Dette betyr at relasjoner med samme reisetid og pendlerindeks, likevel kan få forskjellige (total) døgnfordelinger.

Figur 5.6 viser originale og vektete døgnfordelinger. Originalfordelingene er i svart, mens de røde fordelingene er vektet (multiplisert med) andeler fra reisehensiktsmodellen.



Figur 5.6: Døgnfordelinger for ulike segmenter på relasjonen Oslo-S-Tønsberg. Svarte kurver er estimert fordeling, mens røde kurver er fordelingen vektet med andel fra reisehensiktsmodellen.

Figur 5.6 viser hvordan arbeidsreiser har en høyere andel enn øvrige reisehensikter (se tabell 4.4). Summert over alle hensikter gir dette en total døgnfordeling på relasjonen, slik som vist i figur 5.7.

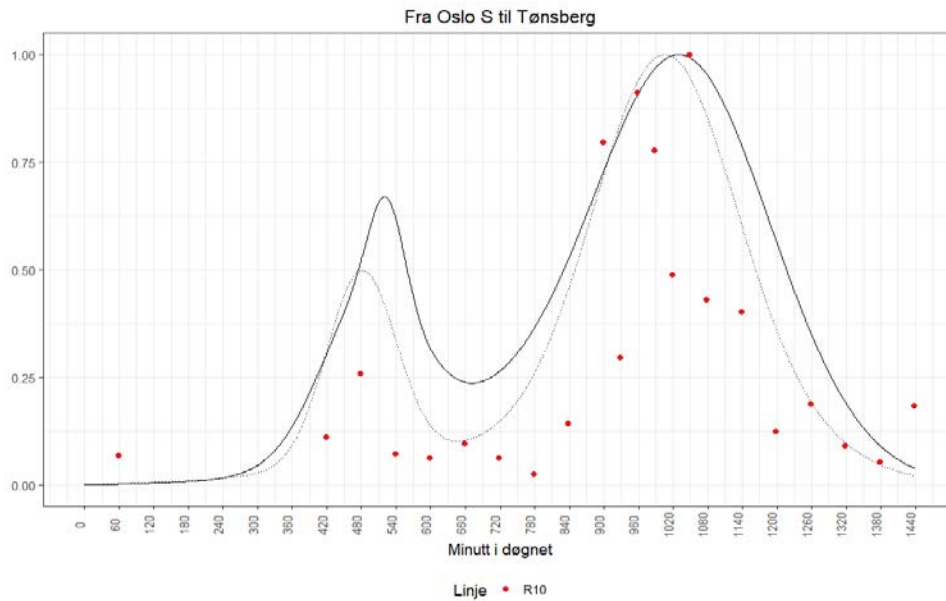


Figur 5.7: Døgnfordeling for alle hensikter på relasjonen Oslo-S–Tønsberg.

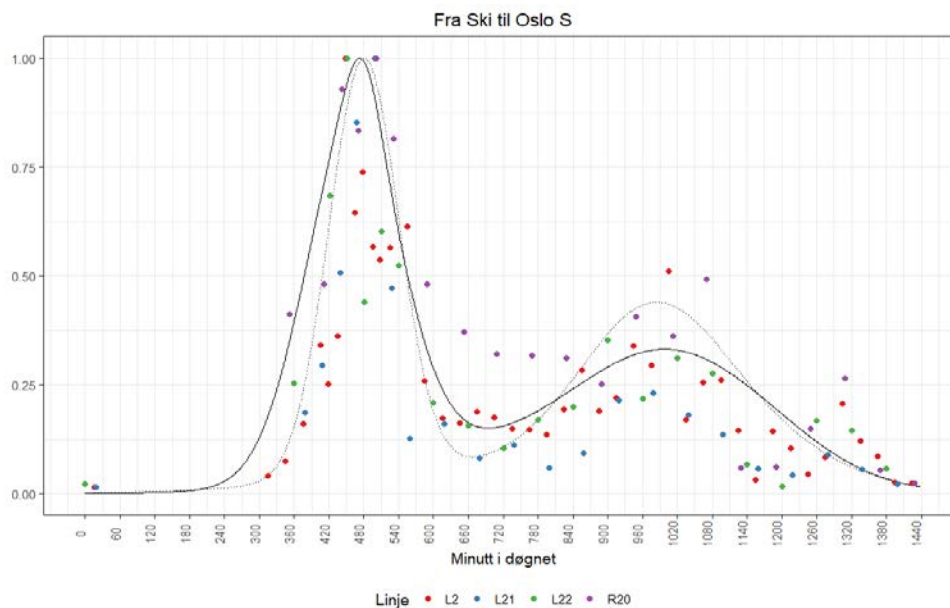
Validering

I Flügel et al. 2017 ble estimerte døgnfordelinger validert mot transformerte passasjertellinger (APC) fra tog. Det har ikke vært ressurser til å benytte nye APC data på en tilsvarende måte i dette prosjektet, men vi har gjort noen sammenlikninger mellom ny og gammel modell for utvalgte relasjoner:

Figur 5.8 viser en sammenlikning på relasjonen Oslo S–Tønsberg og figur 5.9 viser tilsvarende for relasjonen Ski –Oslo S.



Figur 5.8: Validering av døgnfordeling (alle hensikter, virkedøgn) mot transformerte passasjertellinger for relasjonen Oslo S–Tønsberg. Versjon 1: stiplet linje, versjon 2: heltrukken linje. (Fordelingene og passasjertellingene er normert ved å dele på respektive maksimumsverdier).



Figur 5.9: Validering av døgnfordeling (alle hensikter, virkedøgn) mot transformerte passasjertellinger for relasjonen Ski–Oslo S. Versjon 1: stiplet linje, versjon 2: heltrukken linje. (Fordelingene og passasjertellingene er normert ved å dele på respektive maksimumsverdier).

Det er ganske godt samsvar mellom versjon 1 og 2. En av utfordringene med å validere mot APC data er at disse ikke finnes på relasjonsnivå, det er derfor vanskelig å vite om versjon 1 eller 2 best representerer fordelingen over ønsket ankomsttid. Siden vi i versjon 2 har benyttet både tog- og bilreiser vil den nyeste modellen være mindre styrt av eksisterende togtilbud.

Leveranse

Fra prosjektet leveres døgnmodellen som:

1. Excel-fil med parametere for fordelinger i alle segment (18 parametersett).
2. Oppdaterte verdier for pendlerindeks på alle stasjonsrelasjoner.
3. R-script for visualisering av døgnfordelingene.

Vi avslutter med en viktig bemerkning:

Siden sammenhengen mellom lokasjonsparameterne (μ_1 , μ_2) og reisetid er lineær, er det mulig å få fordelinger som får urealistiske verdier på lange relasjoner. I Flügel et al. 2017 ble dette løst for arbeidsreiser ved at vi i) fikserte μ_1 og ii) la en begrensning på hvor høy μ_2 kunne være. Dette ble skjønnsmessig vurdert og i denne versjonen har vi valgt en tilnærming basert på mindre manuell justering. Vi har heller ikke hatt ressurser til samme grad av validering som sist. Figur 4.15 antyder at en fiksering av μ_1 kunne være passende, videre bør man vurdere om det skal legges inn øvre/nedre grenser for lokasjonsparameterne, især for arbeidsreiser. Fordelingene er estimert på reiser kortere enn åtte timer og bør brukes med varsomhet for de lange relasjonene.

6 Diskusjon

6.1 Usikkerhet

Som generelt usikkerhetsmoment må det sies at fordelingen er estimert på – og dermed gjenspeiler - «pre-korona» reiseadferd. Varige endringer etter pandemien, som økt hjemmekontor, kan ha betydelig påvirkning på hensiktsfordeling og døgnfordelingen.

Modellen skal gjenspeile systematiske forskjeller mellom togstasjonsrelasjoner. Det forventes ikke at modellen klarer å fange opp effekter som er spesifikke for enkelte togstasjoner. Det hadde heller ikke vært mulig med et relativt lite datagrunnlag og et stort antall relasjoner modellen skal anvendes på. Modellen vil derfor ikke treffe på alle togstasjonsrelasjoner.

Modellen for reisehensikt/døgnstype er estimert på et datagrunnlag for togreiser inntil 20 timer (med veldig få observasjoner over 10 timer), mens modellen for døgnfordeling er estimert for reiser inntil 8 timer. For veldig lange reiser (over 8 timer) er det stor usikkerhet om modellen gir fornuftige svar. En kjent svakhet er prediksjon av andel restdøgn for disse lange turene. Dette er kort omtalt også i neste avsnitt.

Lokasjonsparameterne i døgnfordelingene er lineære funksjoner av reisetid. På kortere reiser ser dette ut til å gi realistiske effekter, men på lange reiser vil fordelingene forskyves mot urealistisk seint tidspunkt på døgnet. En måte å håndtere dette på er å legge inn øvre begrensninger for lokasjonsparameterne i implementeringen i Trenklin, enten basert på kalibrering mot APC eller skjønnsmessige vurderinger.

6.2 Kjente svakheter og forbedringsmuligheter

At vi var nødt til å etterkalibrere fordelingen for reisehensikt er ikke optimalt. I retrospekt burde vi ha forsøkt å omdefinere tilbringerreiser med tog som egne helreiser i datagrunnlaget. Dette var dessverre ikke mulig med RVU-datasettet som vi fikk tilgang til.²⁶ Det hadde også krevet mye datakodning, spesielt med tanke på stedsfesting av tilbringerreiser. Vi anbefaler likevel at man omdefinere dette ved etablering av versjon 3, i alle fall for tilbringerreiser til flyplasser.

Andel arbeidsreiser kan framstå som høy for turiststasjoner med relativt kort reiseavstand (for lengre relasjoner vil distanseeffekten inntre og «presse» andel arbeidsreiser ned). Dette kan løses ved ytterligere etterkalibrering, men krever at man da har et datagrunnlag å kalibrere mot.

Sammenhengen mellom restdøgn og reisedistanse fungerer dårlig for veldig lange relasjoner (over 8 timer reisetid med tog; se figur 4.4). Grunnen til det ligger trolig i kvadrattermen til reisetiden. Vi anbefaler derfor ikke å bruke modellen til å splitte opp reiser i restdøgn og virkedøgn for veldig lange relasjoner. En mulighet kan ligge i å sette parameterverdier til

²⁶ Datasettet inkluderte åttisifrede grunnkretser men ingen personvariabler og bare sentrale variabler angående reise (ingen variabler som delreiser).

kvadrattermene til 0 og bruke opplegget for distanse-avhengig kalibrering til å justere tilbake effektene for reiser under 8 timer.

I løpet av prosjektet prøvde vi å generere døgnfordelinger basert på maskinlæring. Dette gjorde vi ved å trene dype nevrale nettverk til å predikere ankomsttid som beregnet i estimeringsgrunnlaget basert på RVU. En slik tilnærming kan sees som noe som står mellom ikke-parametriske modeller og strukturerte modeller med få (tolkbare) parametere. Med dype nevrale nettverk hadde man fått en modell bestående av veldig mange (men ikke-tolkbare) parametere. Vi mener at dette er en spennende tilnærming som kunne passe fint til denne type problemstilling. Å gjøre det på en måte som man faktisk kan anbefale fordelinger på, krever dog mer tid og ressurser enn det som var tilgjengelig innenfor dette arbeidet.²⁷

De endelige døgnfordelingene prosjektet leverer er basert på tilnærmingen som ble etablert i Flügel et al. 2017. Fordelen med en vektet sum av to normalfordelinger er at disse modellene har parametere som er lette å tolke. Ulempen er at de ikke er fleksible nok til å gjenskape egenskaper vi ser i de empiriske fordelingene: Et eksempel er at komponenten som beskriver ettermiddagsrushet for arbeidsreiser får høy varians, trolig siden den inkluderer reiser om formiddagen/ettermiddagen. En relativ enkelt utvidelse og oppdatering av modellene kan være å innføre flere komponenter i mixture modellene, for eksempel tre eller fire for å la to komponenter beskrive morgen-/ettermiddagsrush, mens øvrige komponenter kan beskrive reiser utenom rush.

6.3 Anbefaling for implementering av døgnfordelinger i Trenklin

For å implementere døgnmodellen i Trenklin anbefaler vi følgende steg:

- 1) Legg inn 9 matriser med oppsplitting i reisehensikter (for døgntypen som skal analyseres) basert på (eksogen) reisehensiktsmodell (i dagens versjon er det 3 slike matriser).
- 2) Legg inn 9 sett av parametere for døgnfordeling²⁸ i parameter-arket (i dagens versjon er det 3 slike sett av parametere).
- 3) Endre de interne beregninger i VBA-koden slik at de implisitte døgnfordelinger for Arbeid (som definert i Trenklin) er en vektet sum av døgnfordelingen av Arbeid og Skole (som definert i hensiktsmodellen), og slik at Fritid (som definert i Trenklin) er en vektet sum av Handle, Service, Hente/levere, Fritid, Ferie/hytte/natur og Besøk. Vektene beregnes ut fra fordelingene fra steg 1).
- 4) For tjenestereiser kan man bruke dagens beregningsopplegg.

De resulterende fordelinger fra steg 3) vil dermed fange opp forskjeller i detaljerte reisehensikter på tvers av relasjoner. Effekten for arbeidsreiser forventes å være nokså liten i og med at skolereiser utgjør en liten andel av Arbeid og fordi arbeidsreiser og skolereiser har nokså like døgnfordelinger i utgangspunktet.

²⁷ Som nevnt i avsnitt 1.3, har vi i dette prosjektet brukt denne tilnærmingen for å sjekke om andre typer parametere (basert på sonedata) hadde forbedret predikeringen av døgnfordelinger. Konklusjonen var at pendlerindeksen fanger opp sentrale effekter og at effekten av sonedata hadde liten effekt på predikerings-evne. Dette førte til at vi ikke prøvde å inkludere flere forklaringsvariabler i de parametriske døgnfordelingsmodeller.

²⁸ Disse er overlevert Jernbanedirektoratet i det egen Excel-fil. Som nevnt i avsnitt 5, burde man vurdere om det skal legges inn øvre/nedre grenser for lokasjonsparametere, især for arbeidsreiser.

Referanser

- Flügel, S., Ingebrigtsen, R. og Hulleberg, N. (2017). Modellering av reisehensikts- og døgnfordeling for togreiser. TØI rapport 1558/2017.
- Flügel, S. og Hamre, T. (2019) En ny modell for forskyvning av reisetidspunkt i regionale transportmodeller. TØI-rapport 1727/2019.
- Flügel S., Ukkonen A., Hulleberg, N., Østli, V. (2020). Verktøy for framskriving og scenarioanalyse av togreiser. TØI rapport 1804/2020.
- Grue, B., Landa-Mata, I. og Flotve, B. L. (2021). Den nasjonale reisevaneundersøkelsen 2018/19 - nøkkelrapport. TØI rapport 1835/2021.
- Hansen W., Gundersen, F. og Jordbakke, G.N. (2019). Høyfrekvent ferjetilbud. TØI-rapport 1732/2019.
- Ranheim, P. (2017). Trenklin versjon 3 - Dokumentasjon og brukerveiledning, Jernbanedirektoratet.
- Ukkonen, A. og Flügel, S., (2020). Videreutvikling av Trenklin (versjon 3.2): Tilpasning av trenselsfunksjoner og valg av antall togsett per avgang. TØI-Arbeidsdokument 51614.

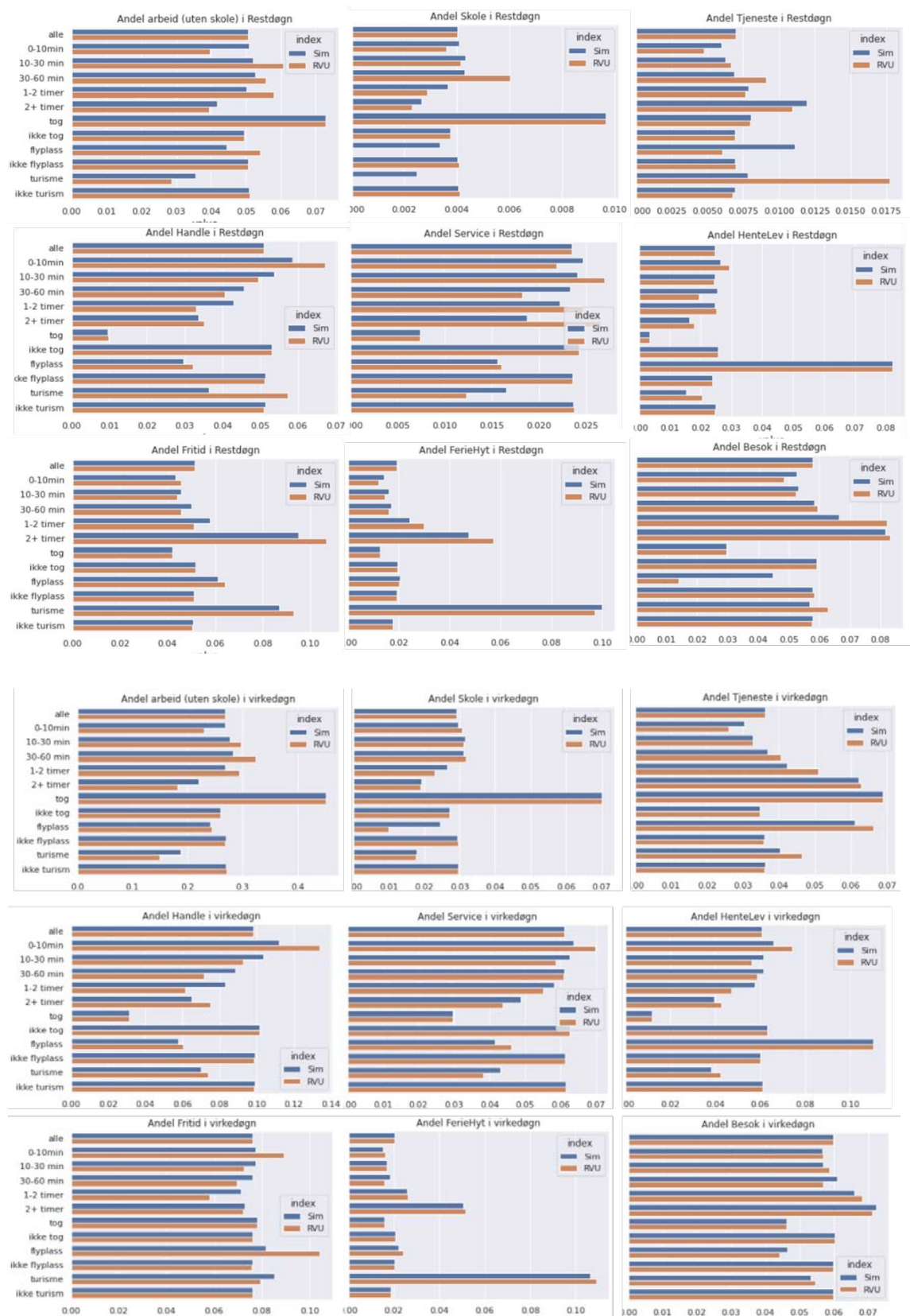
Vedlegg

Vedlegg 1.1 Togstasjoner ved flyplass og turistdestinasjon

Denne tabellen er tatt fra Flügel et al 2021.

| | Flyplass | Turisme | | Flyplass | Turisme |
|------------------------|----------|---------|--------------------|----------|---------|
| Oslo lufthavn stasjon | | 1 | Mjølfjell stasjon | | 1 |
| Torp stasjon | | 1 | Myrdal stasjon | | 1 |
| Trondheim lufthavn st. | | 1 | Oppdal stasjon | | 1 |
| Bjørli stasjon | | | Otta stasjon | | 1 |
| Finse stasjon | | 1 | Rena stasjon | | 1 |
| Flå stasjon | | 1 | Ringebu stasjon | | 1 |
| Geilo stasjon | | 1 | Røros stasjon | | 1 |
| Gol stasjon | | 1 | Upsete stasjon | | 1 |
| Grua stasjon | | 1 | Ustaoset stasjon | | 1 |
| Hallingskeid stasjon | | 1 | Vaksdal stasjon | | 1 |
| Haugastøl stasjon | | 1 | Vieren stasjon | | 1 |
| Hjerkinn stasjon | | 1 | Voss stasjon | | 1 |
| Hunderfossen stasjon | | 1 | Ørneberget stasjon | | 1 |
| Kongsvoll stasjon | | 1 | Ål stasjon | | 1 |
| Ljosanbotn stasjon | | 1 | Åndalsnes stasjon | | 1 |
| Meråker stasjon | | 1 | | | |

Vedlegg 1.2 Sammenligning av simulerte reisehensiktsandeler (før kalibreringen) og RVU-andeler



Transportøkonomisk institutt (TØI) Stiftelsen Norsk senter for samferdselsforskning

TØI er et anvendt forskningsinstitutt, som mottar basisbevilgning fra Norges forskningsråd og gjennomfører forsknings- og utredningsoppdrag for næringsliv og offentlige etater. TØI ble opprettet i 1964 og er organisert som uavhengig stiftelse.

TØI utvikler og formidler kunnskap om samferdsel med vitenskapelig kvalitet og praktisk anvendelse. Instituttet har et verrfaglig miljø med rundt 90 høyt spesialiserte forskere.

Instituttet utgir tidsskriftet Samferdsel på internett og driver også forskningsformidling gjennom TØI-rapporter, artikler i vitenskapelige tidsskrifter, samt innlegg og intervjuer i media. TØI-rapportene er gratis tilgjengelige på instituttets hjemmeside www.toi.no.

TØI er partner i CIENS Forskningscenter for miljø og samfunn, lokalisert i Forskningsparken nær Universitetet i Oslo (se www.ciens.no). Instituttet deltar aktivt i internasjonalt forsknings-samarbeid, med særlig vekt på EUs rammeprogrammer.

TØI dekker alle transportmidler og temaområder innen samferdsel, inkludert trafiksikkerhet, kollektivtransport, klima og miljø, reiseliv, reisevaner og reiseetterspørsel, arealplanlegging, offentlige beslutningsprosesser, næringslivets transporter og generell transportøkonomi.

Transportøkonomisk institutt krever opphavsrett til egne arbeider og legger vekt på å opptre uavhengig av oppdragsgiverne i alle faglige analyser og vurderinger.

Besøks- og postadresse:

Transportøkonomisk institutt
Gautstadalléen 21
NO-0349 Oslo

22 57 38 00
toi@toi.no
www.toi.no