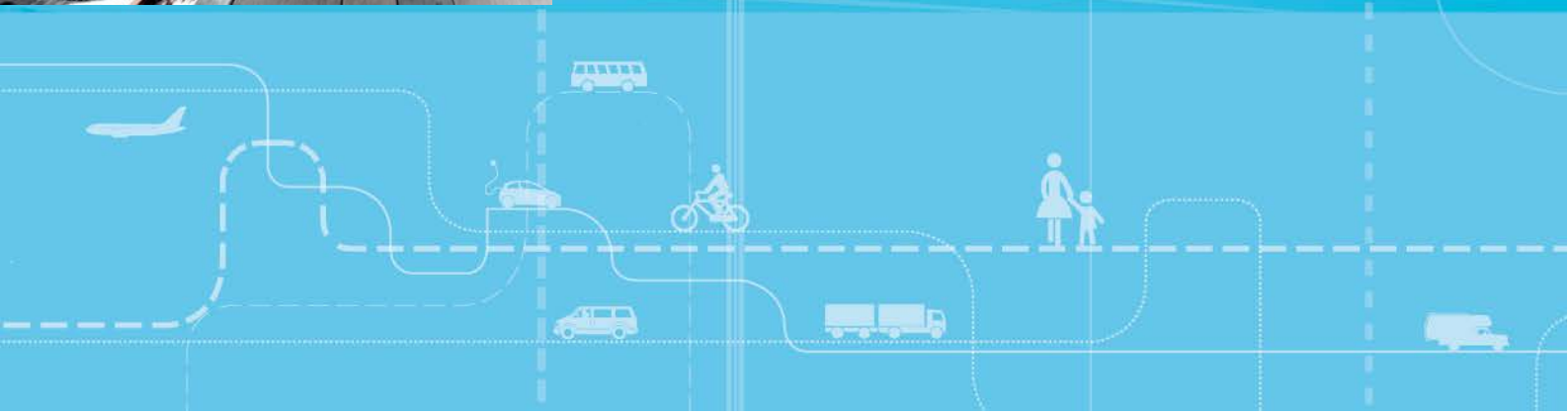
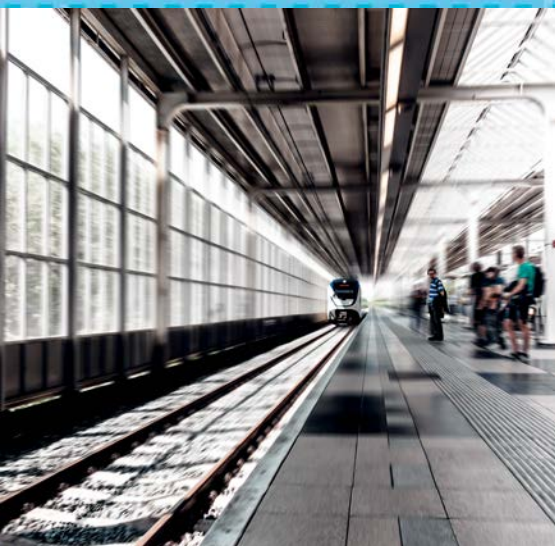


Modellering av reisehensikts- og døgnfordelinger for togreiser



Modellering av reisehensikts- og døgnfordelinger for togreiser

Stefan Flügel, Rikke Ingebrigtsen, 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: Modellering av reisehensikts- og døgnfordelinger for togreiser

Title: Modelling distributions of trip purpose and preferred arrival time for train trips

Forfattere: Stefan Flügel,
Rikke Ingebrigtsen,
Nina Hulleberg

Authors: Stefan Flügel,
Rikke Ingebrigtsen,
Nina Hulleberg

Dato: 03/2017

Date: 03/2017

TØI-rapport: 1558/2017

TØI Report: 1558/2017

Sider: 81

Pages: 81

ISBN elektronisk: 978-82-480-1898-8

ISBN Electronic: 978-82-480-1898-8

ISSN: 0808-1190

ISSN: 0808-1190

Finansieringskilde: Jernbanedirektoratet (tidligere Jernbaneverket)

Financed by: Norwegian Rail Authorities

Prosjekt: 4356- Fordeling Trenklin

Project: 4356- Fordeling Trenklin

Prosjektleder: Stefan Flügel

Project Manager: Stefan Flügel

Kvalitetsansvarlig: Anne Madslie

Quality Manager: Anne Madslie

Fagfelt: Transportmodeller

Research Area: Transportmodeller

Emneord: Reisehensiktsfordeling

Keywords: Trip purpose

Døgnfordeling

Preferred arrival time

Togtransport

Train transport

Modellering

Modelling

Sammendrag:

På oppdrag fra Jernbaneverket (JBV) har Transportøkonomisk institutt (TØI) utarbeidet reisehensikts- og døgnfordelingsmodeller for personreiser med tog i Norge. Rapporten dokumenterer inndata, metodeutvikling og empiriske analyser som TØI har utført i forbindelse med prosjektet. Reisehensiktsfordelingen ble etablert med statistiske modeller av typen logitmodell der den avhengige variabelen er reisehensikt av enkle turer i RVU 2013-2014. Modellen bruker reisetid, antall og type arbeidsplasser, befolkning, type stasjon, og antall hytter og hoteller som forklaringsvariabler i modellen. Døgnfordelingene modelleres som 'mixtures of linear regressions' med normalfordelte komponenter. For arbeidsreiser finner vi en sammenheng mellom pendlerstrømmen og formen på døgnfordelingen, det vil si om hovedvekten av de reisende ankommer endestasjonen om morgenen eller om ettermiddagen.

Summary:

Commissioned by the Norwegian Rail Administration (JBV), the Institute of Transport Economics (TOI) has established models for trip purpose- and preferred arrival time distributions for personal train trips in Norway. The report documents the input data, methodology development and empirical analyzes that TOI has done in connection with the project. Distributions over trip purposes were established with logit models where the dependent variable is the travel purpose of trips reported in the National travel survey (NTS 2013-2014). The model uses travel time, number and type of jobs and population around train stations, type of train stations (e.g. station close to airports) and the number hotels as explanatory variables. Distributions of preferred arrival time are modeled as mixtures of linear regressions with normally distributed components. For work trips we find a correlation between commuting patterns and the shape of distributions.

Language of report: Norwegian

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

*Institute of Transport Economics
Gaustadalleen 21, 0349 Oslo, Norway
Telefon 22 57 38 00 - www.toi.no*

Forord

Rapporten dokumenterer metodeutvikling og empiriske analyser som Transportøkonomisk institutt (TØI) har jobbet med i forbindelse med å etablere såkalte reisehensikt- og døgnfordelinger for personreiser med tog. Disse fordelingene brukes i Jernbanedirektoratets modellverktøy Trenklin, for å fordele togturer mellom to togstasjoner etter reiseformål og ønsket ankomsttid. Modellen predikerer for eksempel hvor høy andel arbeidsreiser det er fra Oslo S til Ski stasjon, og hvordan ankomsttiden (på Ski stasjon) for disse arbeidsreisene fordeler seg over døgnet. Denne informasjon er viktig for samfunnsøkonomiske beregninger, siden tidsverdien avhenger av reisehensikt, og trengsel ombord på togene avhenger av når reiseetterspørselen er størst.

Arbeidet er utført av tre forskere ved TØI. Stefan Flügel har vært prosjektleder og utviklet reisehensiktmodellen. Rikke Ingebrigtsen har utviklet modeller for døgnfordelinger. Nina Hulleberg har hovedsakelig jobbet med databearbeiding og har bidratt i rapportskrivning og diskusjoner underveis.

Modellen er utviklet på oppdrag fra Jernbaneverket (fra 2017 Jernbanedirektoratet). Kontaktperson hos oppdragsgiver har vært Jon-Kristian Ryan Hovland, som sammen med Patrick Ranheim og Malene Nerland har bidratt til prosjektet. Vi takker for gode innspill og godt samarbeid gjennom prosjektet.

Forskningsleder Anne Madslien ved TØI har vært ansvarlig for kvalitetssikring av rapporten. Sekretær Trude Rømming har hjulpet til med ferdigstilling av rapporten.

Oslo, mars 2017

Transportøkonomisk institutt

Gunnar Lindberg
Direktør

Anne Madslien
Forskningsleder

Innhold

Sammendrag

Summary

1	Introduksjon.....	1
1.1	Bakgrunn.....	1
1.2	Reisehensikt.....	1
1.3	Ønsket ankomsttid.....	3
2	Oversikt metodetilnærming.....	6
2.1	Arbeidsopplegg.....	6
2.2	Oversikt: data og dataflyt.....	7
2.3	Prinsipper ved modellering.....	8
2.4	Jernbaneverkets innspill til prosjektet.....	9
3	Datafangst og databearbeiding.....	10
3.1	Beskrivelse av «rådata».....	10
3.2	Fra grunnkrets til togstasjoner.....	14
3.3	Opplegg for å konvertere snitt-tellinger til OD-matriser.....	17
4	Etablering av reisehensiktsmodell.....	20
4.1	Definisjon av reisehensikter og fordeling i RVU.....	20
4.2	Statistisk modell.....	22
4.3	Forklaringsvariabler i modellen.....	22
4.4	Estimering.....	25
4.5	Illustrering av resultater.....	28
4.6	Implementert modell.....	30
4.7	Kjente svakheter og forbedringsmuligheter.....	31
5	Etablering av døgnfordelinger.....	32
5.1	Definisjon av døgnfordeling.....	32
5.2	Statistisk modell for døgnfordelingene.....	33
5.3	Modelltilnærming og datagrunnlag.....	37
5.4	Fordelinger basert på RVU.....	44
5.5	Mixture parametere fra telldata.....	55
5.6	Oppsummering av døgnfordelinger.....	58
5.7	Validering av de endelige døgnfordelingene mot transformerte passasjertellinger.....	62
5.8	Implementering av modell.....	65
5.9	Svakheter ved døgnfordelingene og behov for nye data.....	65
6	Referanser.....	67
	Vedlegg.....	69
	Vedlegg 1: Stasjoner som ikke er kodet inn i RTM og/eller ikke finner nærmeste grunnkrets.....	70
	Vedlegg 2: Døgnfordelinger over avreisetidspunkt med ulike transportmidler.....	71
	Vedlegg 3: Sammenligning av ankomsttidsvariabler.....	74
	Vedlegg 4: Validering av døgnfordelinger for utvalgte stasjonsrelasjoner.....	76

Sammendrag

Modellering av reisehensikts- og døgnfordelinger for togreiser

TØI rapport 1558/2017

Forfattere: Stefan Flügel, Rikke Ingebrigtsen, Nina Hulleberg

Oslo 2017 81 sider

Hvorfor personer reiser med tog og når på dagen de reiser kan antas å variere med egenskapene til start- og endestasjon for reisen, for eksempel antall og type arbeidsplasser i nærheten av stasjonene, og egenskaper ved selve reisen (reisetid mellom start- og endestasjon). Vi har identifisert hvordan reisehensikt (hvorfor) og ankomsttid (når) fordeler seg for ulike togstasjonsrelasjoner, og har etablert statistiske modeller som kan predikere disse fordelingene gitt forklaringsvariabler. Vi finner at reisehensikt i stor grad forklares av reisetid: korte reiser har betydelig høyere andel arbeidsreiser enn lange reiser. Døgnfordelingene er sterkt avhengig av hvor mange personer som pendler i den ene retningen sammenlignet med den motsatte retningen.

Bakgrunn

På oppdrag fra Jernbaneverket (JBV, nåværende Jernbanedirektoratet) har Transportøkonomisk institutt (TØI) utarbeidet reisehensikts- og døgnfordelingsmodeller for personreiser med tog i Norge. Rapporten dokumenterer inndata, metodeutvikling og empiriske analyser som TØI har utført i forbindelse med prosjektet.

Fordelingene som er utviklet skal brukes i modellverktøyet Trenklin, for å fordele togreiser mellom togstasjoner etter reiseformål og ønsket ankomsttid. I samfunnsøkonomiske analyser av jernbaneprosjekter er reisehensiktsfordelingen av betydning for korrekt verdsetting av tilbudsendringer. Det skilles mellom arbeids-, forretningsreiser og øvrige reiser i modellen. I tillegg er reiseetterspørselens fordeling over døgnet viktig for gode etterspørselsberegninger av tiltak, og spesielt for verdsetting av tiltak som bidrar til å redusere trengsel ombord på togene.

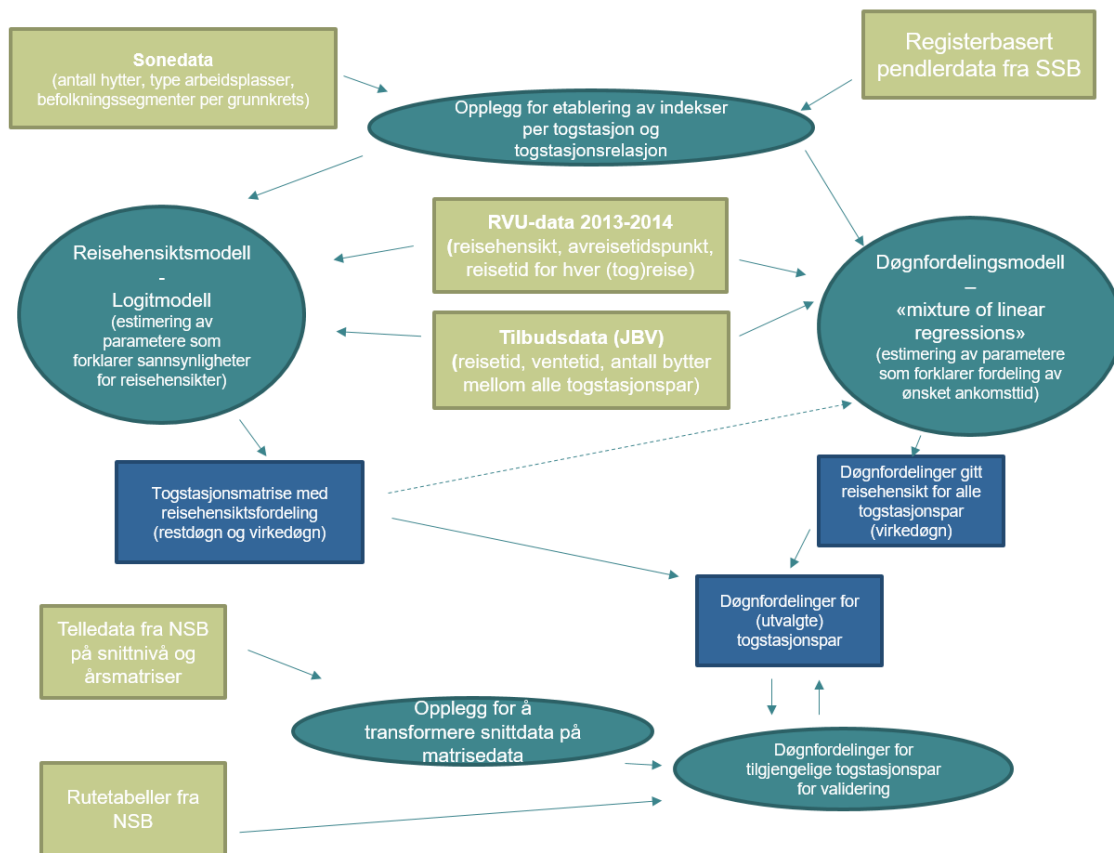
Data

Vi bruker en rekke ulike eksterne datakilder i dette prosjektet:

- RVU-data (2013/14): rapportert reisehensikt og avreisetidspunkt for personreiser, stedsfesting på grunnkrets nivå.
- Sonedata som i regionale transportmodeller (RTM) for 2014. Dette er opprinnelig data fra SSB: informasjon om hytter/hoteller, ulike type arbeidsplasser og ulike befolkningssegmenter på grunnkrets nivå.
- Telledata fra NSB (2014) (passasjertellinger for utvalgte toglinjer) og årsmatriser, ikke differensiert etter reisehensikt.
- Togstasjonsmatriser med togtilbud (ombordtid, reisetid, ventetid, antall bytter) levert av Jernbaneverket, gjelder år 2014.
- Rutetabeller for 2014, levert av Jernbaneverket.
- Pendlerstatistikk fra SSB på grunnkretsrelasjoner, registerdata fra 2014.

En viktig del av prosjektet har bestått i å tilrettelegge dataene slik at de kan brukes som grunnlag for estimering av fordelingene. En viktig kilde i prosjektet er RVU data fra 2013-2014 som gir oss informasjon om formål og ankomsttid ved reisen. Mens RVU er data på individnivå, opererer Trenklin på togstasjonsnivå. Dette legger føringer for hvilke typer forklaringsvariabler som skal brukes for å forklare fordelinger. Vi er i dette prosjektet ikke interessert i hvordan reisehensikt og ankomsttid varierer med sosiodemografiske variabler (kjønn, husholdningsstørrelse, inntekt osv.), men hvordan den varierer med egenskaper ved togstasjoner. Dette innebærer at vi må koble sammen individdata med geografiske data som har et mer aggregert nivå.

Figur S1 viser en oversikt over inndata og hvordan disse henger sammen med de ulike stegene for å etablere modellene.



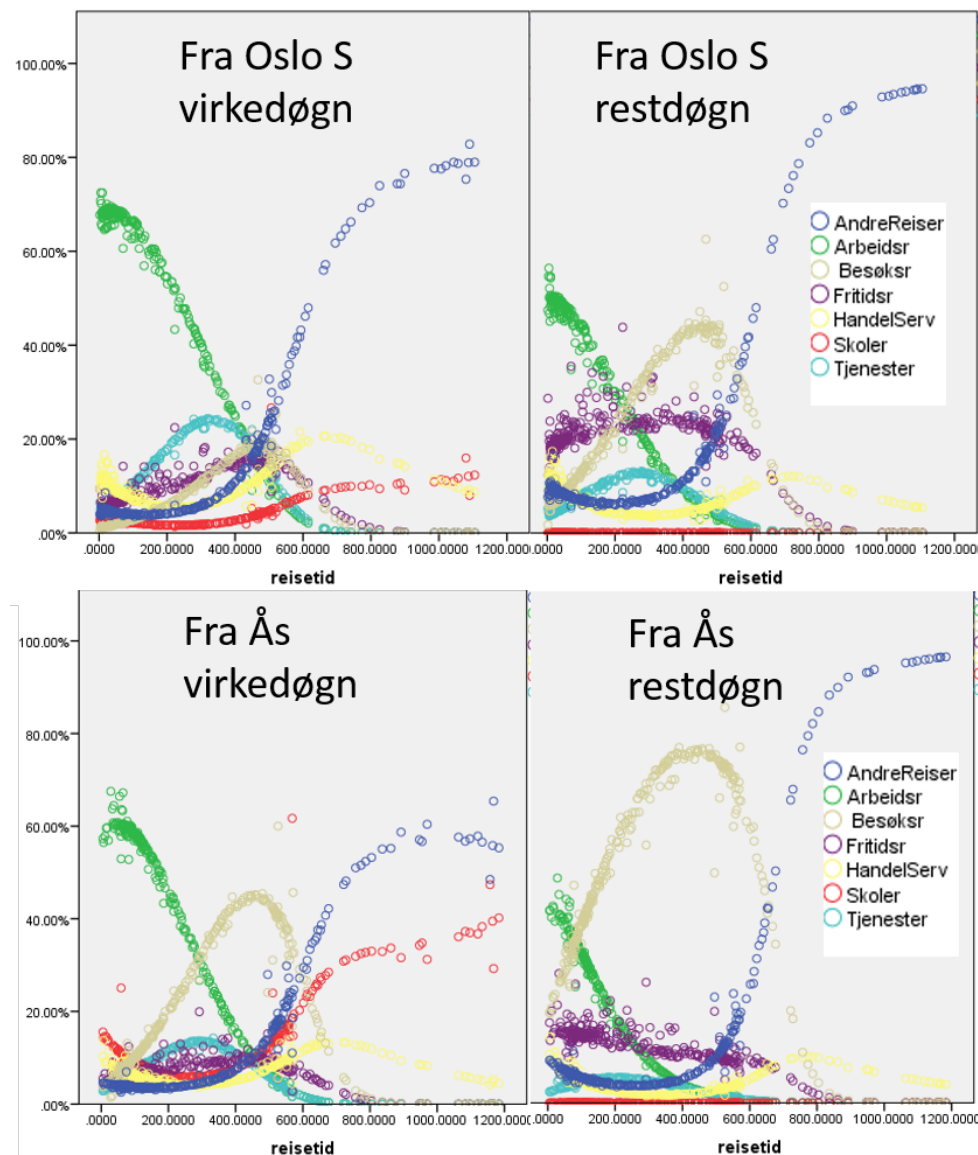
Figur S1: Dataflyt (lysegrønn: eksterne inndata, grønn: interne beregninger, blå: utdata). Stiplet pil indikerer uttesting, men ikke brukt i endelig modell.

Reisehensiktsfordeling

Reisehensikt ved en (tog)reise beskriver formålet med reisen. Reiseformålet er typisk ikke observert i trafikk- eller passasjertellinger og må etterspørres i reisevaneundersøkelser. En reisehensiktsfordeling beskriver relative andeler av trafikken som utgjøres av hver reisehensikt, og skal summeres til 100 prosent over alle hensikter. Det antas at reisehensiktsfordelingen er forskjellig for ulike stasjonspar, og fordelingen vil dermed være avhengig av hvilke type reiser som genereres ved startstasjon, og hvilke reiser som attraheres ved endestasjon. Videre antas reisehensiktsfordelingen å være ulik for virkedøgn (vanlige arbeidsdager) og restdøgn (helg og fellesferie).

Reisehensiktsfordelingen er etablert med hjelp av statistiske modeller av type logitmodell der den avhengige variabelen er reisehensikt for enkeltturer i RVU 2013-2014. Det brukes reisetid, antall og type arbeidsplasser, befolkning, type stasjon, og antall hytter og hoteller som forklaringsvariabler i modellen. Andelen arbeidsreiser er klart høyest på korte relasjoner og avtar mer eller mindre lineært med avstand. For reiser over 700 km er andelen arbeidsreiser tilnærmet null. Andelen tjenestereiser med tog øker først, og avtar igjen etter rundt 200-400 km (avhengig av relasjon og type døgn). Andelen til de andre reisehensiktene er noe mer varierende, men i sum vil disse øke med distanse og vil være tilnærmet 100 prosent for de lengste relasjonene.

Figur S2 viser andelene til reisehensiktsfordeling mot reisetid (i minutter) for alle stasjonspar med henholdsvis Oslo S og Ås som startstasjoner.



Figur S2: Reisehensiktsandeler etter reisetid (i minutter). Hvert punkt representerer en endestasjon med henholdsvis Oslo S og Ås som startstasjon.

Tabell S1 og S2 viser predikerte reisehensiktsandeler for utvalgte stasjonspar.

Tabell S1: Predikerte andeler for Trenklins hovedreisehensikter på utvalgte strekninger/ stasjonspar.

Predikerte andeler for virkedøgn i %			Til stasjon			
			Lillehammer	Lillestrøm	Oslo S	Trondheim S
Fra stasjon	Lillehammer	Arbeid		58	58	30
		Tjeneste		8	12	14
		Andre reiser		34	29	56
	Lillestrøm	Arbeid	58		71	21
		Tjeneste	8		5	12
		Andre reiser	34		24	67
	Oslo S	Arbeid	59	71		13
		Tjeneste	12	5		20
		Andre reiser	29	24		66
	Trondheim S	Arbeid	32	20	13	
		Tjeneste	14	12	20	
		Andre reiser	54	68	68	

Tabell S2: Predikerte andeler for Trenklins hovedreisehensikter på utvalgte strekninger/ stasjonspar.

Predikerte andeler for virkedøgn i %			Til stasjon			
			Bergen	Drammen	Geilo	Oslo S
Fra stasjon	Bergen	Arbeid		24	44	16
		Tjeneste		16	11	26
		Andre reiser		60	45	57
	Drammen	Arbeid	22		47	72
		Tjeneste	14		11	7
		Andre reiser	65		42	21
	Geilo	Arbeid	39	47		44
		Tjeneste	13	11		17
		Andre reiser	49	42		39
	Oslo S	Arbeid	14	72	45	
		Tjeneste	22	7	17	
		Andre reiser	64	21	39	

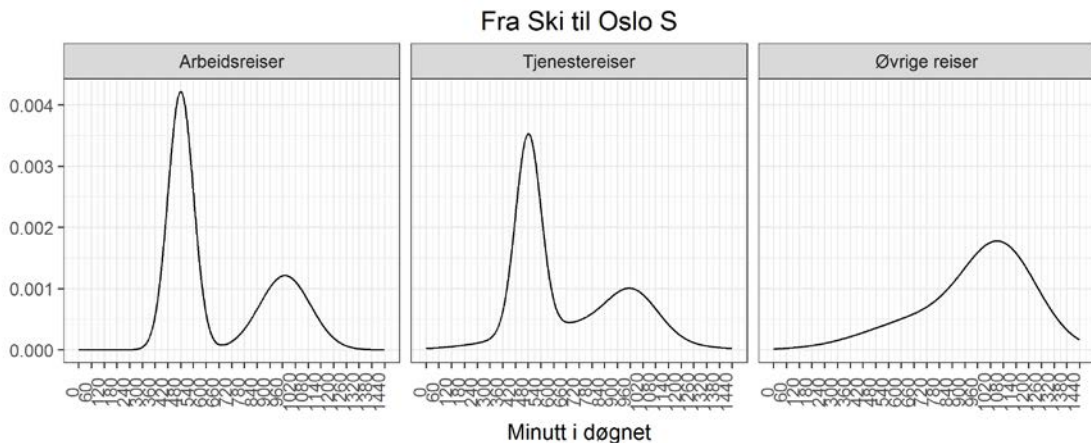
Modellen er implementert i Excel der nye fordelinger kan beregnes ved endring av inndata. I rapporten påpekes ulike forbedringsmuligheter for videreutvikling av modellen.

Døgnfordeling

En døgnfordelingsmodell beskriver hvordan etterspørselen etter togreiser fordeler seg over døgnet. I Trenklin benyttes fordelinger over ønsket ankomsttid ved endestasjon gitt startstasjon og reisehensikt. Døgnfordelingene modelleres som *mixtures of linear regressions* med normalfordelte komponenter. Togreiser fra den nasjonale reisevaneundersøkelsen for 2013–14 er benyttet som inndata for ønsket ankomsttid, mens reisetid og den relative pendlerstrømmen mellom start- og endestasjon beskriver formen på fordelingene. Døgnfordelingene etableres kun for virkedøgn.

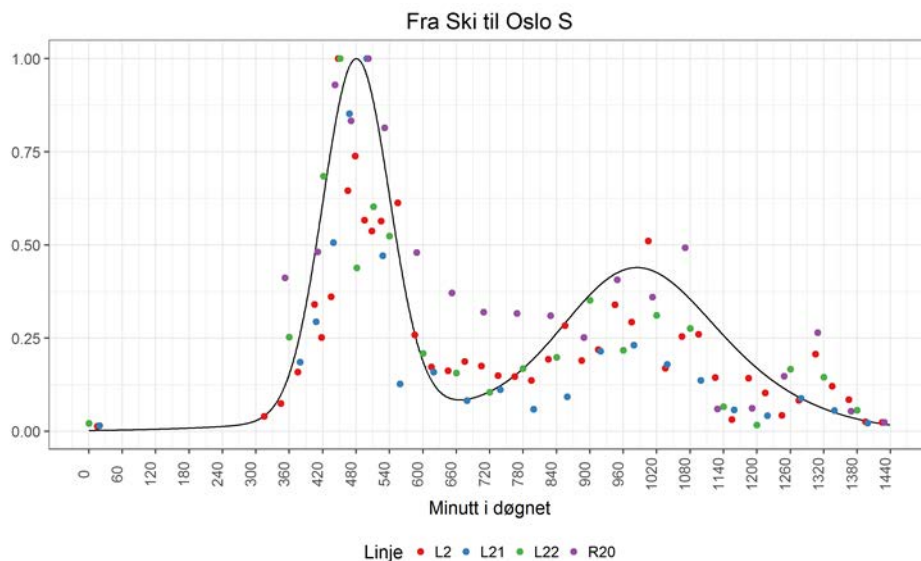
Mangelen på data med informasjon om både reisehensikt og ankomsttid for togreiser har gjort det utfordrende å etablere døgnfordelingene. For arbeidsreiser var det tilstrekkelig med togreiser i RVU til å finne en sammenheng mellom pendlerstrømmen og formen på døgnfordelingen, det vil si om hovedvekten av de reisende ankommer endestasjonen om morgenen eller om ettermiddagen. For de to andre reisehensiktene, og især for tjenestereiser, er datagrunnlaget tynt og det er knyttet større usikkerhet til formen på disse fordelingene.

Figur S3 viser døgnfordelinger over ankomsttid ved Oslo S for reiser fra Ski stasjon.



Figur S3: Døgnfordeling over ankomsttid ved Oslo S for reiser som starter ved Ski stasjon for de tre reisehensiktene arbeidsreise, tjenestereise og øvrige reiser.

I Figur S4 er døgnfordelingene og andelene fra hensiktsfordelingen (65 prosent arbeidsreiser, 5 prosent tjenestereiser og 30 prosent øvrige reiser) kombinert for å sammenligne fordelingen over ankomsttid uavhengig av reisehensikt med observasjoner basert på faktiske passasjertellinger fra NSB. Den heltrukne linjen er den estimerte fordelingen, mens punktene angir ankomsttidsfordelingen for toglinjene L2, L21, L22 og R20 på Østfoldbanen. Passasjertellingene er transformert til OD-format fra netto av- og påstigende passasjerer.



Figur S4: Sammenligning mellom døgnfordeling for alle reisebensikter og transformerte passasjertellinger for stasjonsrelasjonen Ski – Oslo S. Fordelingen og passasjertellingene er normert ved å dele på respektive maksimumsverdier for visualiseringens skyld.

Figurene over viser at hovedvekten av (arbeids)reiser fra Ski ankommer Oslo S om morgenen (480 er kl. 8.00). Det er relativt godt samsvar mellom den estimerte fordelingen og passasjertellingene, men det ser ut til at etterspørselen midt på dagen estimeres som noe lavere enn den faktisk er. Dette skyldes trolig at det er få observasjoner av tjenestereiser og øvrige reiser med tog i RVU, samt at reisebensikt er definert slik at en reise fra arbeid til butikk/barnehage/treningscenter etc. regnes som en øvrig reise og at dette gir en etterspørselstopp etter denne typen reiser rundt kl. 16, når arbeidsdagen er over (ved normal arbeidstid).

Summary

Modelling distributions of trip purpose and preferred arrival time for train trips

TOI Report 1558/2017

Authors Stefan Flügel, Rikke Ingebrigtsen, Nina Hulleberg
Oslo 2017 81 pages Norwegian language

Why and when people are traveling by train can be assumed to vary with the characteristics of the departure and destination station (for example, number and type of jobs near the stations) and characteristics of the journey (journey time between the start and end station). We have identified how travel purpose (why) and arrival time (when) are distributed over various pairs of train stations and have built statistical models that can predict these distributions given explanatory variables.

Commissioned by the Norwegian Rail Administration (JBV), the Institute of Transport Economics (TOI) has established models for trip purpose and preferred arrival time distributions for personal train trips in Norway. The report documents the input data, methodology development and empirical analyzes that TOI has done in connection with the project.

The distributions are to be used in the modeling tool Trenklin which operates on a train station relation level. The model do therefor not explain how trip purpose vary on an individual level (by characteristics of the travelers) but by characteristics related to the train stations.

Distributions over trip purposes were established with logit models where the dependent variable is the travel purpose of trips reported in the National travel survey (NTS 2013-2014). The model uses travel time, number and type of jobs and population around train stations, type of train stations (e.g. station close to airports) and the number of hotels as explanatory variables. We find that trip purpose is largely explained by travel time. For example short journeys have an higher proportion of commuting trips than long journeys.

Distributions of preferred arrival time are modeled as mixtures of linear regressions with normally distributed components. For work trips we find a correlation between commuting patterns and the shape of distributions. For instance, the morning peaks will be greater for trips into Oslo where relative many working places are located.

1 Introduksjon

1.1 Bakgrunn

Rapporten er knyttet til prosjektet «Døgn- og reisehensiktsmodeller for persontransport» på oppdrag fra Jernbaneverket (JBV).

Arbeidsoppgavene i dette prosjektet er beskrevet i konkurransegrunnlaget:

- «Første del av oppdraget går ut på å innhente og sammenstille relevante data for å kunne estimere sammenhenger mellom karakteristika ved stasjonsrelasjoner, fordeling av reisehensikt og døgnfordeling. En sentral del av arbeidet blir å finne egnede funksjonsformer for døgn- og hensiktsfordeling, definere variabler som skal inngå i modellen og estimere parameterne ved bruk av statistiske metoder på egnede data.»
- «Andre del av oppdraget går ut på å utvikle en brukerorientert modell som benytter de estimerte funksjonene på inngangsdata.»

Modellen skal brukes som utgangspunkt for å etablere sentrale inndata i JBVs transportmodell Trenklin (p.t. versjon 2.8, Ranheim (2016), Flügel og Hulleberg (2016)). Mer spesifikt skal resultatet fra prosjektet være:

- Reisehensiktsfordeling mellom alle togstasjoner i Norge for virkedøgn og restdøgn
- Døgnfordeling over ønsket ankomsttid på virkedøgn mellom alle togstasjoner i Norge gitt reisehensikt (arbeidsreiser (inkl. skolereiser), tjenestereiser og andre reiser)

De to neste avsnittene beskriver nærmere hva vi mener med «reisehensikt» og «ønsket ankomsttid», og begrunner hvorfor det er viktig å ha kunnskap om deres fordeling for ulike togreiser.

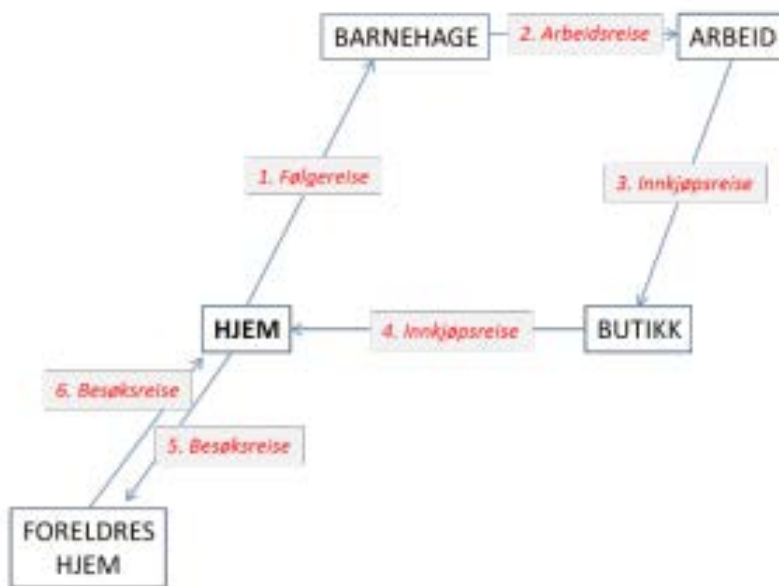
1.2 Reisehensikt

Reisehensikt ved en (tog)reise beskriver formålet med reisen. Et formål er typisk ikke observert i trafikk- eller passasjertellinger og må etterspørres i reisevaneundersøkelser. Formål kategoriseres i grupper, og den mest disaggregerte kategoriseringen av reisehensikter i den nasjonale reisevaneundersøkelsen (RVU) er som gitt i Figur 1.1.

- 1 = "Arbeidsreise (Reise til/fra arbeid)"
- 2 = "Skole (reise til/fra skole)"
- 3 = "Tjenestereise (Reise i arbeid)"
- 4 = "Innkjøp av dagligvarer"
- 5 = "Andre innkjøp (alle andre innkjøp)"
- 6 = "Service/div ærend (bank/post, reisebyrå etc.)"
- 7 = "Medisinske tjenester (lege/sykehus, tannlege)"
- 8 = "Hente/bringe/følge barn til/fra"
- 9 = "Hente/bringe/følge barn til/fra sports- og"
- 10 = "Andre hente-/bringe-/følgereiser"
- 11 = "Besøk (privat besøk hos familie, venner, sykebesøk)"
- 12 = "Kino, teater, konsert, utstilling mv"
- 13 = "Kafé, restaurant, pub mv"
- 14 = "Fotballkamp, sportsarrangement mv som tilskuer"
- 15 = "Organiserte fritidsaktiviteter, musikk, idrett, trening,"
- 16 = "Gikk/syklet/jogget en tur/skitur/luftet hund"
- 17 = "Reiste til fritidsbåt/marina"
- 18 = "Båttur"
- 19 = "Hyttetur"
- 20 = "Andre ferie-/helgereiser"
- 21 = "Annet formål"

Figur 1.1: Disaggregert kategorisering av reisehensikter i den nasjonale reisevaneundersøkelsen (RVU).

Det er viktig å merke seg at flere reisehensikter er overlappende, for eksempel vil en reise fra arbeidsplass til barnehage både være «arbeidsreise (reise til/ fra arbeid)» og «hente barn». Regelen i RVU er at det er dit du skal (destinasjon) som bestemmer reisehensikten, bortsett fra for reiser hjem til eget bosted hvor reisehensikt vil defineres ut fra den aktivitet man gjorde på startsted av reisen (i en reisekjede vil dette være siste destinasjon før man kommer hjem). Figur 1.2 illustrerer dette.



Figur 1.2: Reisehensikt gitt ved dit du skal og hvor du kommer fra (kilde: Hjorthol mfl. 2014).

En reisehensiktsfordeling beskriver relative andeler for de ulike reisehensikter, og skal summeres til 100% over alle hensikter. Reisehensiktsfordelinger er viktige inndata i Trenklin og andre transportmodeller, samt i verktøy for nytte-kostnadsanalyser. Det er fordi det brukes ulike tidsverdier for ulike reisehensikter, og fordi reiseetterspørselsfordelingen varierer med reisehensikt. Tidsverdien for tjenestereiser er rundt 8,5 ganger så høy som tidsverdien for fritidsreiser. Tidsverdiene er en sentral størrelse for adferdsmodelleringen i Trenklin og vil i stor grad påvirke beregningen av trafikantnytte gitt en endring i transporttilbudet (f. eks. redusert reisetid eller ventetid). Det er derfor viktig at man treffer reisehensiktsfordelingen i utgangssituasjonen best mulig.

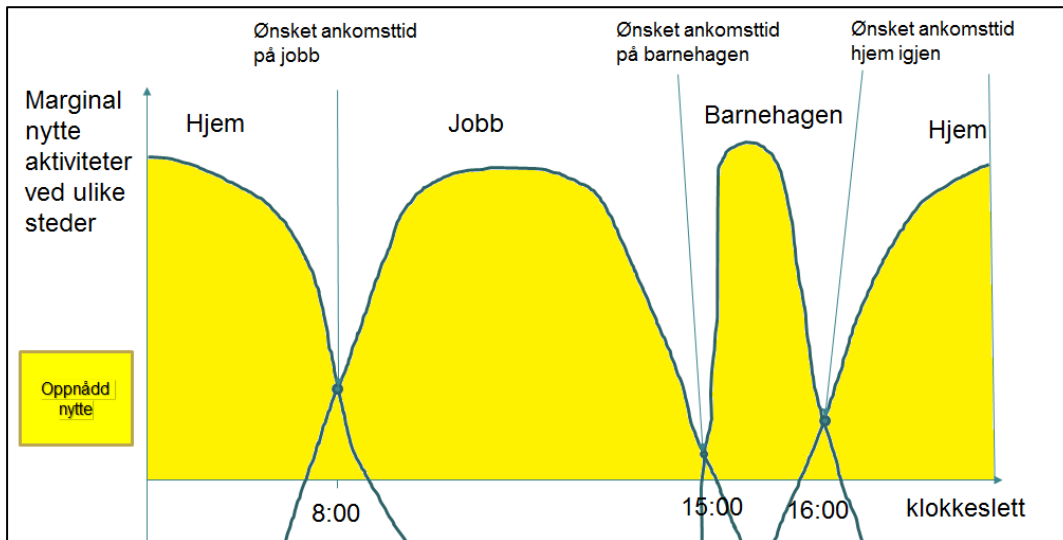
Reisehensiktsfordelingen antas å variere med togstasjonsrelasjon (par av startstasjon og endestasjon) og er avhengig av hvilke type reiser som genereres ved startstasjon og attraheres ved endestasjon. En stasjon omgitt av mange arbeidsplasser vil attrahere mange arbeidsreiser. Siden turer fra arbeid til hjem også regnes som arbeidsreiser, vil den samme stasjonen også genere mange arbeidsreiser i motsatt reiseretning.

Videre antas reisehensiktsfordelingen (eller kort «hensiktsfordelingen») å være ulik for virkedøgn (vanlige arbeidsdager) og restdøgn (helg og fellesferie). Vi beregner derfor to separate fordelinger; én for virkedøgn og én for restdøgn.

1.3 Ønsket ankomsttid

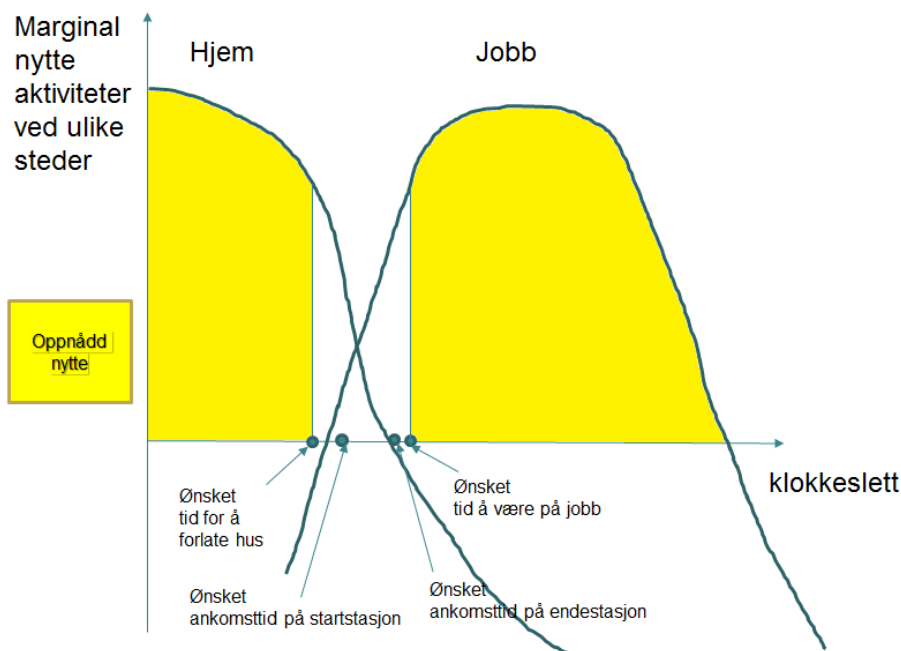
Ved siden av reisehensikt er vi i dette prosjektet interessert i *når* togreisende ønsker å ankomme endestasjonen. Ønsket ankomsttid (ved endestasjon) vil i stor grad avhenge av når man ønsker å utføre en planlagt aktivitet. Ønsker man for eksempel å se en kinofilm som begynner klokken 20:00, vil man gjerne å ankomme kinoen 19:55. Gitt at det tar fem minutter å gå mellom togstasjonen og kinoen, er det ønskelig å være på endestasjonen 19:50.

Det er mulig å beskrive konseptet med ønsket ankomsttid med utgangspunkt i mikroøkonomisk teori. Man antar at det til enhver aktivitet finnes en marginal nytte ved å utføre den gitte aktiviteten ved ethvert klokkeslett. Dette er illustrert i Figur 1.3. Når man ikke tar hensyn til reisetid (altså i en verden der man kan teleportere seg fra sted til sted) så vil ønsket ankomsttid være det tidspunktet den marginale nytten ved stedet du er/aktiviteten du utfører, er i ferd med å bli lavere enn nytten der du skal. Man kan maksimere den totale nytten (illustrert som det gule arealet under kurvene) ved å alltid være der hvor den marginale nytten er størst (for eksempel 8:00 ved jobb, 15:00 ved barnehagen og 16:00 ved bosted).



Figur 1.3: Ønsket ankomsttid i en hypotetisk verden med teleportering.

I virkeligheten kan man ikke oppnå nytten som illustrert i Figur 1.3 siden man må bruke tid på å reise, og reisetiden vil typisk ikke ha like høy nytteverdi som selve (hoved)aktiviteten. Siden reisetiden gir lavere nytte, kan vi normalisere nytten ved reise til 0, og med en gitt reisetid (her antatt 30 minutter) vil man oppnå nytten som illustrert i Figur 1.4 (viser bare overgangen mellom hjem og jobb).

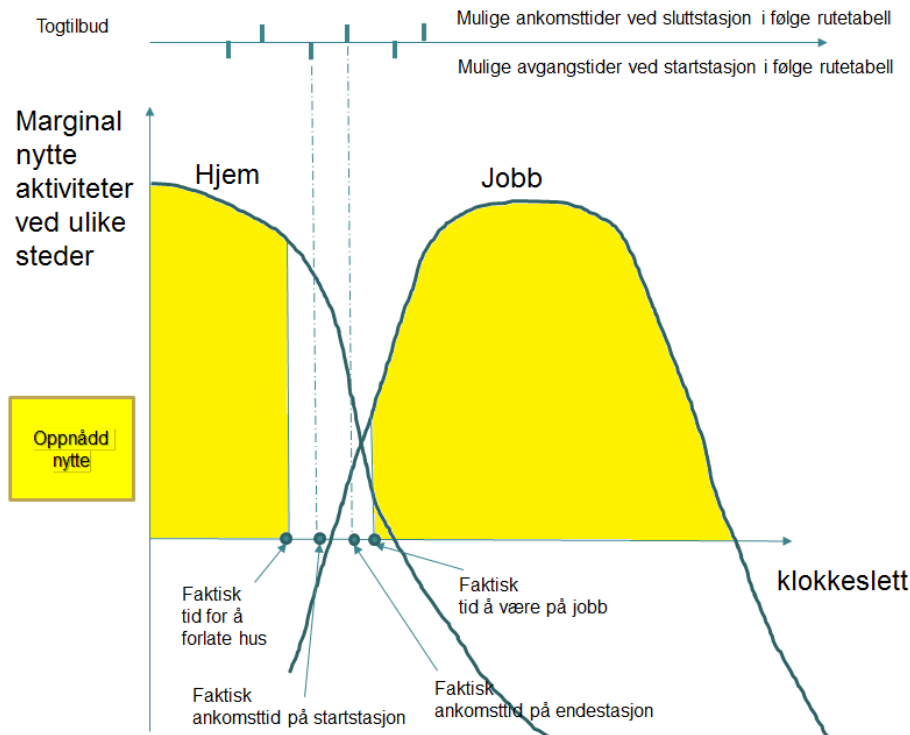


Figur 1.4: Ønsket ankomsttid gitt reisetid (i dette eksempelet 10 minutters tilbringertid til stasjonen, 15 minutter på tog, samt 5 minutters reisetid fra togstasjon til arbeidsplass).

Vi må skille mellom ønsket ankomsttid ved aktivitet og ønsket ankomsttid ved endestasjon. Det er ønsket ankomsttid ved endestasjon vi prøver å modellere i dette prosjektet.

En utfordring er at ønsket ankomsttid ikke er observert for togreiser. For bilreiser (og andre transportmidler der man kan velge avreisetidspunkt selv) vil man – i fravær av uventede køsituasjoner – kunne anta at den observerte ankomsttiden vil samsvare med

ønsket ankomsttid. For togreiser (og andre reiser der transportmiddelet følger en rutetabell) vil man typisk ha en «skjult ventetid» som oppstår hvis man må ankomme endestasjonen før eller etter den ønskede ankomsttiden. Dette er illustrert i Figur 1.5, der den reisende må forlate hjemmet tidligere enn ønsket.



Figur 1.5: Faktisk ankomsttid ved valg av togavgang.

Figur 1.5 viser at ønsket ankomsttid vil påvirke valg av togavgang. I dynamiske transportmodeller som Trenklin, der valg av togavgang predikeres i modellen, trenger man derfor et anslag på hvor mange personer som ønsker å ankomme til ethvert tidspunkt på en endestasjon. Med andre ord må man anta en døgnfordeling over ønsket ankomsttid ved endestasjonen.

Gitt en slik fordeling og en antakelse om hvor høyt togpassasjerer verdsetter reduksjon i (skjult) ventetid kan man beregne nytteverdien av tilbudsendringer. For eksempel vil en økning av avgangsfrekvens føre til at noen personer kan reise på et mer gunstig tidspunkt som reduserer den skjulte ventetiden og øker den oppnådde nytten.

Illustrasjonen viser også at ønsket ankomsttid per person vil variere med reisehensikt. Typisk vil arbeidsreiser (til/fra jobb) foregå tidligere på dagen enn fritidsreiser (til/fra trening, kino o.l.). Fra Figur 1.4 ser man også at ønsket ankomsttid må antas å variere med reiselengde, noe som motiverer antagelsen om ulike fordelinger per togstasjonsrelasjon (og ikke per endestasjon isolert sett).

Siden reisehensikten «arbeidsreiser» spesifiseres for både reise «til jobb» og «fra jobb til hjemsted» (se avsnitt 1.2), vil den ønskede ankomsttiden for arbeidsreiser både være relevant om formiddagen (som typisk «til jobb») og om ettermiddagen (som typisk «fra jobb»). Derfor vil døgnfordeling over ønsket ankomsttid avhenge av pendlerstrømmen mellom start- og endestasjon. Det vil si at en relasjon som Langhus–Oslo S typisk vil ha en høy andel arbeidsreiser om morgenen, mens relasjonen Oslo S–Langhus vil ha mange arbeidsreiser med ønsket ankomsttid om ettermiddagen.

Videre i rapporten refererer vi ofte til «døgnfordeling over ønsket ankomsttid ved endestasjon» som bare «døgnfordeling».

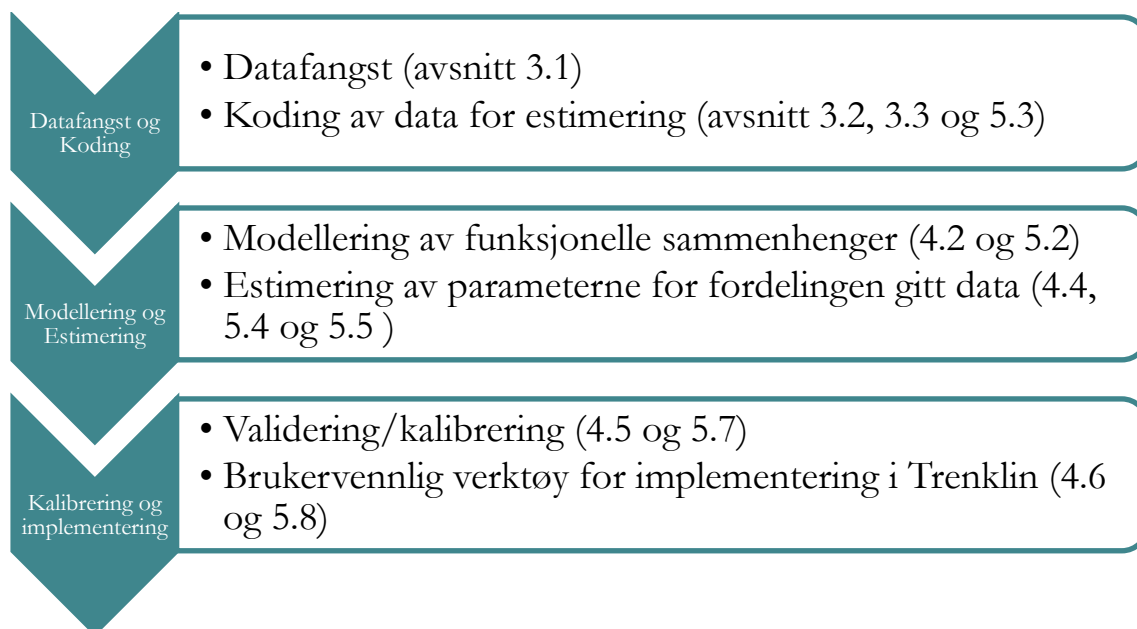
2 Oversikt metodetilnærming

I dette kapittelet gir vi en oversikt og redegjørelse av den metodiske tilnærmingen vi har valgt å bruke for å løse oppgaven.

2.1 Arbeidsopplegg

Arbeidsopplegget la opp til en ren empirisk framgangsmåte. Vi har ikke funnet litteratur som kunne gi oss teoretiske sammenhenger mellom forklaringsvariabler og reisehensikts- og/eller døgnfordeling. Det virker også å være lite empiriske erfaringstall fra den nasjonale og internasjonale litteraturen som man kunne anta å være overførbare til vårt case. Vi har derfor prioritert å samle inn relevante norske data, og bruker disse som utgangspunkt for å lage egne modeller og verktøy. Med begrepet modell mener vi matematiske sammenhenger, mens et verktøy er en fil eller et program som anvender modellen til å produsere/predikere fordelingene.

Arbeidsopplegget til prosjektet kan deles i tre hovedgrupper: 1. datafangst og -koding, 2. modellering og estimering og 3. kalibrering og implementering. Dette er illustrert i Figur 2.1 som også angir avsnitt i rapporten hvor deloppgavene er beskrevet nærmere.



Figur 2.1: metodisk arbeidsopplegg i prosjektet.

2.2 Oversikt: data og dataflyt

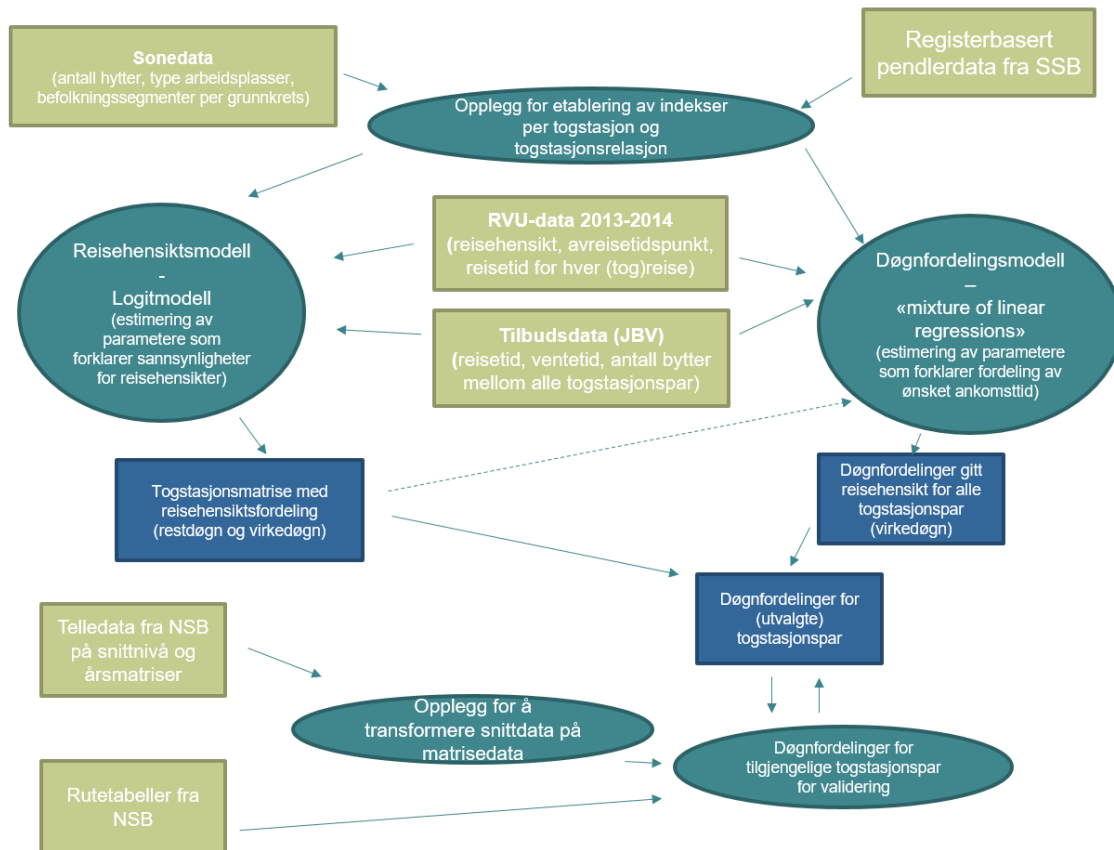
Vi bruker en rekke ulike eksterne datakilder for estimering av fordelingene:

- RVU-data (2013/14): rapportert reisehensikt og avreisetidspunkt for personreiser, stedsfesting på grunnkrets nivå.
- Sonedata som i bruk i regionale transportmodeller (RTM) for 2014. Dette er opprinnelig data fra SSB: informasjon om hytter/hoteller, ulike type arbeidsplasser og ulike befolkningssegmenter på grunnkrets nivå.
- Telledata fra NSB (2014) (passasjertellinger for utvalgte toglinjer) og årsmatriser, ikke differensiert etter reisehensikt.
- Togstasjonsmatriser med togtilbud (ombordtid, reisetid, ventetid, antall bytter) levert av JBV, gjelder år 2014.
- Rutetabeller for 2014, levert av Jernbaneverket.
- Pendlerstatistikk fra SSB på grunnkretsrelasjoner, registerdata fra 2014.

Alle datakilder er såkalte sekundære datakilder som har blitt innsamlet for generelle formål og ikke i forbindelse med dette prosjektet. Dette innebærer at flere datakilder må bearbeides/omkodes for å passe til formålet i prosjektet. Dette er knyttet til formatet som forventes på sluttproduktet, dvs. fordelinger på togsstasjonsrelasjonsnivå (matriseform). For et mest mulig konsistent opplegg transformeres formatet på inndata før estimering. Vi har benyttet tre typer omkoding i prosjektet:

1. Transformasjon av data på grunnkrets nivå til data på togstasjonsnivå (se avsnitt 3.2)
2. Transformasjon av data på snitt til data på matrisenivå (se avsnitt 3.3)
3. Omregning av avreisetidspunkt og total reisetid til aktivitet (fra RVU) til ankomsttid ved endestasjon (se avsnitt 5.3.1)

Figur 2.2 viser dataflyten i prosjektet.



Figur 2.2: Dataflyt (lysegrønn: eksterne inndata, grønn: interne beregninger, blå: utdata). Stiplet pil indikerer uttesting, men ikke brukt i endelig modell.

2.3 Prinsipper ved modellering

Det er noen prinsipper som vi legger til grunn ved modellering og estimering med utgangspunkt i hvordan hensikts- og døgnmodellen skal brukes i praksis. I denne sammenheng er det viktig å merke seg at vi – ved bruk av RVU – observerer reisehensikt og ankomsttid på individnivå mens Trenklin opererer på togstasjonsnivå.

Dette legger føringer for hvilke typer forklaringsvariabler som skal brukes for å forklare fordelinger. Vi er i dette prosjekt ikke interessert i hvordan reisehensikt og ankomsttid varierer med sosiodemografiske variabler (kjønn, husholdningsstørrelse, inntekt osv.), men hvordan den varierer med egenskaper ved togstasjoner. Dette innebærer at vi må koble sammen individdata (der vi observerer formål og ankomsttid for reisen) med geografiske data som har et mer aggregert nivå.

I dette prosjektet skal vi lage generiske modeller, det vil si vi skal finne matematiske sammenhenger som gjelder uavhengig av den spesifikke togstasjonen (men som fører til en variasjon blant alle stasjonspar). Teknisk sett betyr det at vi ikke modellerer eller estimerer stasjonsspesifikke konstanter. Heller enn å bruke en parameter for Oslos S beskriver vi type stasjoner (i dette tilfelle «storbyssenter»). Grunnen til at vi lager generiske modeller er i hovedsak at modellen skal kunne brukes for å predikere fordelinger for nye togstasjoner.

En annen grunn til å bruke generiske modeller er at det finnes over hundre tusen togstasjonsrelasjoner i Norge, og vi har altfor lite data til å estimere veldig spesifikke sammenhenger.

Det er derfor ikke hensiktsmessig (eller mulig) å lage modeller som er tilpasset enhver togstasjonsrelasjon. Prinsippet i modellutviklingen er at vi beskriver logiske sammenhenger på et mer overordnet nivå. Slike generiske modeller vil naturligvis bomme på enkelte relasjoner, men vil ved riktig spesifisering produsere en ønsket og fornuftig variasjon i prediksjonene. Vi oppfatter det som tilstrekkelig med tanke på hva som er etterspurt ved bruk av modellene i Trenklin.

Et annet prinsipp ved valg av modell er at det bør være relativt enkelt å kalibrere modellen i lys av nye og mer representative data. Valg av en logitmodell for reisehensiktfordeling følger dette prinsippet. En stor fordel med vanlige logitmodeller (multinomiske logitmodeller) er at parameterestimaterne til alle variabler unntatt konstantledd vil være konsistent estimert selv om utvalget ikke er tilfeldig trukket fra populasjonen (Manski og Lerman 1977).

2.4 Jernbaneverkets innspill til prosjektet

Jernbaneverket (JBV), ved Jon-Kristian Hovland og Patrick Ranheim, har deltatt aktivt i prosjektet og har bidratt til gjennomføringen ved å

- Levere noe av inndata (telldata og rutetilbudsdata)
- Generelle kommentarer og konkrete innspill i totalt 4 prosjektmøter
- Uttesting av reisehensiktsmodell
- Innspill til hvordan modellen bør implementeres

Første versjon av reisehensiktsmodellen ble presentert for JBV i begynnelsen av november 2016. JBV testet modellen og formulerte konkrete innspill til hvordan modellen kan/bør endres/forbedres. Det førte bl.a. til at vi innførte retningsbalanse i reisehensiktsmodellen. Det vil si at vi har tvunget modellen til å produsere (tilnærmet) samme hensiktsfordeling for – for eksempel - Oslo S–Ski som for Ski-Oslo S.

Når det gjelder døgnfordelingene har vi i løpet av prosjektets gang vært innom flere metodetilnæringer. Den største utfordringen har vært mangel på data som beskriver både reisehensikt og reisetidspunkt mellom togstasjoner. I dialog med JBV har ambisjonsnivået for modellkompleksiteten blitt senket etter hvert. Den primære datakilden vi benytter oss av er RVU, men telldata er benyttet til validering og uttesting for relasjoner JBV var spesielt interessert i.

3 Datafangst og databearbeiding

3.1 Beskrivelse av «rådata»

3.1.1 RVU-data (2013-2014)

RVU er en nasjonal reisevaneundersøkelse som har blitt gjennomført regelmessig siden 1985 (fra og med år 2015 samles den inn kontinuerlig).

I RVU 2013/14 er det blitt utført rundt 61 400 intervjuer og det er registrert over 200 000 reiser. Svarprosent ligger på 20%. Av det totale frafallet skyldes 2/3 problemer med å oppnå kontakt eller tekniske problemer, mens 1/3 av frafallet skyldes at personer ikke ønsker å delta (Hjorthol mfl. 2014).

Utvalget består av bosatte i Norge som er minst 13 år. 10 000 respondenter er tilfeldig trukket, mens de resterende respondentene er rekruttert via såkalte regionale tilleggsutvalg. En oversikt over de 24 regionale tilleggsutvalgene finnes i Hjorthol mfl. (2014). Disse virker å være jevnt fordelt over landet.

3.1.2 Sonedata

Med sonedata mener vi variabler som beskriver karakteristika ved grunnkretser i Norge. Disse kommer opprinnelig fra SSB, men er for dette prosjektet hentet ut fra RTM-systemet hvor variablene ligger i bearbeidet form.

Vi kan inndelesonedata i 3 grupper: informasjon om arbeidsplasser (totalt og oppdelt i kategorier), befolkningsdata (totalt og oppdelt i kjønn og ulike aldersgrupper) og øvrige sonedata.

Type arbeidsplasser

Kategoriseringen av arbeidsplasser som brukes i RTM-systemet skiller seg fra NACE-grupperingen (EUs standard for næringsgruppering). Kategoriseringen er tilpasset de ulike reisehensiktene. En av de viktigste forskjellene mellom RTM-gruppene og NACE-gruppene er at det i RTM-systemet ble etablert en egen kategori tilpasset reisehensikten hente/levere (Rekdal mfl. 2013). Tabell 3.1 viser inndelingen og tilordning i RTM-systemet ut fra NACE-kodingen, mens Tabell 3.2 gir en mer verbal beskrivelse av type arbeidsplasser.

Tabell 3.1: Inndeling av type arbeidsplasser etter NACE-kodene (tabell 2.1 i Rekdal mfl. 2013).

NR	Hovedkategori	Kategori	Beskrivelse	NACE2002	Nytt navn
1	1	Primær		01-05	A10PRI
2	2	Sekundær		10-45, 60-62, 90	A20SEK
3	3	Verksted og handel		50-55	A30VH
4	3	Ekskl.	Høyfrekvent VH (privat)	Deler av 52	A31VH
5	3	Ekskl.	Lavfrekvent VH (privat)	Deler av 50, 52	A32VH
6	3	Ekskl.	Fritid	Deler av 55	A33VH
7	3	Ekskl.	Service (privat)	Deler av 52	A34VH
8	4	Tjenester		63-74, 91-99	A40TJE
9	4	Ekskl.	HentLev/Fritid/Service (privat)	93.04	A41TJE
10	4	Ekskl.	HentLev/Fritid	92.13, 92.32, 92.33, 92.61 92.621, 92.622, 92.629 92.721, 92.722, 92.729	A42TJE
11	4	Ekskl.	Service (privat)	63, 64, 65, 71, 93.01 93.02, 93.05	A43TJE
12	4	Ekskl.	Fritid	91.31, 91.32, 91.33, 92.12 92.34, 92.51, 92.521, 92.53 92.71	A44TJE
13	5	Offentlig adm.		75	A50OFF
14	6	Undervisning	HentLev	80	A60UND
15	7	Helse og sosial		85	A70HSOS
16	7	Ekskl.	Service (privat), HentLev	85.111, 85.112, 85.121 85.122, 85.124, 85.125 85.13, 85.142, 85.327 85.328	A71HSOS
17	7	Ekskl.	HentLev	85.331, 85.332	A72HSOS
18	7	Øvrig Helse sos		(residual)	A73REST
19		I alt		00-99	A009TOT

Tabell 3.2 : Type arbeidsplasser i sonedata i RTM-modellen (basert på Tabell 10.2 i Rekdal m. f. 2013); totalt antall i tabellen gjelder år 2009/2010.

kode	Kommentar	Antall
A10PRI	Primærnæringer	24211
A20SEK	Sekundærnæringer	557942
A30VH	Verksted & varehandel som ikke inngår under (agentur og engros, etc.)	123844
A31VH	Detaljhandel med varer vi kjøper ofte (matvarer, vinmonopol,)	69452
A32VH	Detaljhandel med varer vi kjøper sjelden (biler, klær, kjøkken, møbler, etc.)	167056
A33VH	Salg av mat/drikke/overnatting (hoteller, cafeer, gatekjøkken, etc.)	71977
A34VH	Reparasjoner/verksteder, skomaker, etc.	1080
A40TJE	Tjenester som ikke inngår under	371651
A41TJE	Helsestudio-, massasje- og solstudiovirksomhet	5400
A42TJE	Kino, kunst, fornøyelser, idrett, kultur, fritid	14708
A43TJE	Reisebyrå, turistrelaterte tjenester, post, bank, utleie, etc.	69860
A44TJE	Museer, biblioteker, politiske org., religiøse org., interesseorg., etc.	24181
A50OFF	Offentlig administrasjon	131380
A60UND	All undervisning	188719
A70HSOS	Helse og sosial sektor som ikke inngår under	187238
A71HSOS	Alminnelige sykehus (ikke sykehjem/psykiatri), legetjeneste, poliklinikker, tannhelse, barnehager/parker, etc.	187472
A72HSOS	SFO og Fritidsklubber	14608
A73REST	Hjemmehjelp, laboratorier, asylmottak	85050
Total		2295829

I forbindelse med estimeringsarbeidet har vi testet hvilke av disse variablene som bidrar til å forklare reisehensikts- og døgnfordelingene. Ikke alle variabler er tatt med i modellene (se detaljer i avsnitt 4.3).

Befolkningsdata

Befolkningsdata kommer oppsplittet i kjønn og aldersgrupper (5-årsintervaller). I modellen for reisehensikt inngår befolkningen totalt, uten oppsplitting i aldersgrupper. Variablene brukes for å beskrive andel besøksreiser mellom to togstasjoner og er utgangspunkt for koding av variabler som identifiserer stasjoner med «landlig»/spredt befolkning.

Øvrige sonedata

Øvrige sonedata som har blitt testet ut i dette arbeidet er antall hoteller (NUMHOTS), antall hytter og fritidshus (HYTFRIT), antall elevplasser i grunnskolen (GSKOLE), antall elevplasser i videregående skole (VGSKOLE) og studieplasser på universitet/høyskolenivå (UHSKOLE).

3.1.3 Togtilbud for år 2014

Data om togtilbudet i 2014 (og 2015) er levert fra Jernbaneverket. En nasjonal Trenklin-modell har blitt brukt for å etablere gjennomsnittlig ombordtid, tid ved bytting, antall omstigninger og ventetid mellom alle togstasjoner i Norge (inkl. noen svenske stasjoner).

I analysene videre har vi slått sammen ombordtid og ventetid ved bytte til en variabel «reisetid».

Siden RVU-observasjoner gjelder 2013 og 2014 bruker vi tilbudsdata for 2014 (ikke 2015) i estimeringen. Det er relativt lett å legge inn nye reisetider i den implementerte modellen om det er ønsket i framtiden.

3.1.4 NSB-Telledata

I dette prosjektet har vi hatt tilgang til to typer telledata; passasjertellinger for utvalgte toglinjer, samt NSBs landsdekkende matriser (årsmatriser).

Passasjertellinger

NSB teller alle av- og påstigninger to ganger i året, en gang på høsten og en gang på våren. Tabell 3.3 viser hvilke strekninger vi har mottatt telledata for, på hvilket nivå tellingene forekommer, og om dataene også inneholder tidsinformasjon (rutetabell).

Tabell 3.3: Oversikt over mottatte telledata.

Toglinje	Av- og påstigninger per stasjon	Snitt-tellinger	Rutetabell	Annet format
L1: Spikkestad - Lillestrøm		X	X	
L2: Lysaker - Ski		X	X	
L12: Kongsberg – Eidsvoll Verk		X	X	
13: Eidsvoll – Drammen (Kongsberg)		X	X	
L14: Asker – Kongsvinger		X	X	
L21: Skøyen – Moss		X	X	
L22: (Gulskogen) Skøyen – Mysen (Rakkestad)		X	X	
R10: Skien – Lillehammer		X	X	
R20: Oslo S – Halden		X	X	
Jærbanen (L59)	X	X	X	
Nabotåget	X			
Trønderbanen	X			
Vossebanen				X
Ombordtellinger Østlandet				X

Telledata er konfidensielle og kan ikke rapporteres i absolutte tall.

Årsmatriser

NSB utarbeider også egne trafikkmatriser som viser det totale antallet togreiser i løpet av et år. Denne statistikken er fordelt på toglinjer og stasjonsrelasjonsnivå, i motsetning til tellingene i forrige avsnitt som er på stasjons- eller snittnivå. Trafikkmatrisene er basert på uttak fra NSBs billettstatistikk, samt bearbeiding av av- og påstigningstillinger i områder hvor den vanlige billettstatistikken ikke er tilstrekkelig.

Bearbeidingen resulterer i tre trafikkmatriser per linje:

- Interntrafikk, reiser mellom stasjoner som ligger langs den utvalgte linjen.
- Eksternttrafikk, reiser mellom to stasjoner hvor en stasjon ligger på den aktuelle linjen, mens den andre stasjonen ligger langs en annen linje.
- Gjennomgående trafikk, reiser mellom stasjoner langs andre linjer.

I dette prosjektet har vi hatt tilgang til NSBs landsdekkende matriser for 2013, noe som samsvarer bra med tidspunkt for RVU-dataene.

3.1.5 Pendlerstatistikk

I tilknytning til et annet prosjekt har TØI bestilt data fra SSB om pendling for år 2014. Dataene tar utgangspunkt i registerbasert sysselsettingsstatistikk. En pendler er her definert som en person med forskjellig bosteds- og arbeidsgrunnkrets. Det merkes at

- Det mangler informasjon om arbeidsplasser som ikke er tilknyttet en bedrift (selvstendig næringsdrivende). Her defineres arbeidsgrunnkrets som «missing value».
- For noen yrker er det antatt at arbeidsplass ligger ved bostedsgrunnkrets (sjøfolk, militær). Disse observasjonene er ikke relevant for dette prosjektet.
- Det er generell usikkerhet knyttet til bedrifter uten fast arbeidssted (anleggsarbeid, transportnæring osv.)

Datasettet inndeles i grupper for pendling etter ulike arbeidsforhold: «1-19 timer per uke», «20-29 timer» og «30 +timer». I døgnmodellen har vi brukt alle jobbkategorier som forklaringsvariabel.

3.2 Fra grunnkrets til togstasjoner

Stedsfesting i RVU, sonedata fra RTM og pendlerstatistikk er inndata på grunnkretsnivå og må - som nevnt i avsnitt 2.2 - konverteres til togstasjonsnivå før estimering.

I arbeidet med prosjektet ble det vurdert ulike metoder for å koble grunnkretser til togstasjoner. Jernbaneverket har blant annet utarbeidet en oversikt over hvilke grunnkretser som befinner seg innenfor ulik radius (luftlinje) fra stasjonen. Fordelen med denne tilnærmingen er at det er en veldig enkel metode, og man trenger kun informasjon om grunnkretsplasseringen og stasjonsplasseringen. Ulempen er at metoden ikke tar hensyn til hvor befolkningstyngdepunktet i grunnkretsen ligger. Metoden tar heller ikke hensyn til distanse langs vei, noe som trolig er avgjørende for mange ved valg av stasjon. Til slutt har det blitt gjort endringer i grunnkretsnummer siden 2013 som vi ville måtte ta hensyn til.

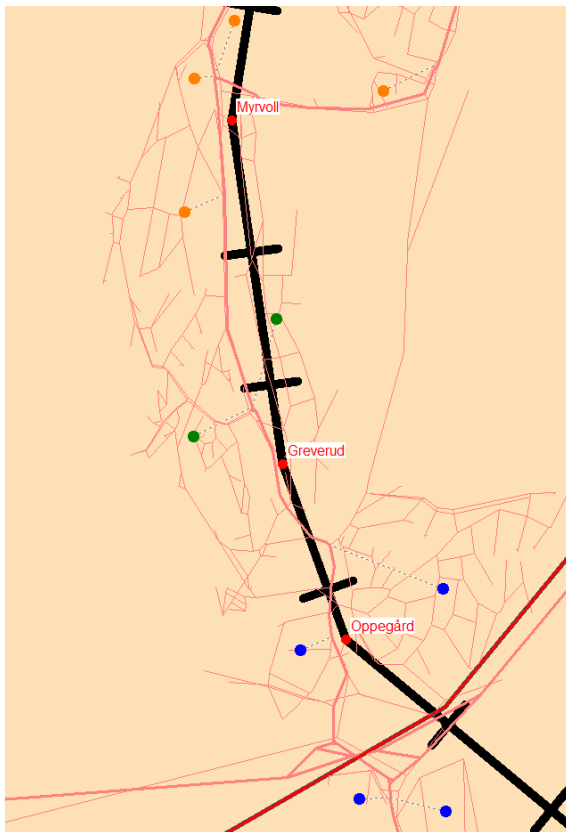
På bakgrunn av dette valgte vi å utarbeide et eget opplegg for å koble grunnkretser til togstasjoner. Avsnitt 3.2.1 presenterer dette opplegget, mens avsnitt 3.2.2 beskriver en beregningsmåte for indeksverdier basert på sonedata.

3.2.1 Opplegg for å koble grunnkretser til togstasjoner

I regional transportmodell (RTM) benyttes data om demografi og arbeidsplasser i hver grunnkrets (sone) som grunnlag for å beregne transporttetterspørselen. Befolkning og arbeidsplasser i den enkelte sone plasseres som en forenkling i et geografisk tyngdepunkt i sonen (sonesentroide). Dette geografiske tyngdepunktet ble ved etableringen av RTM forsøkt plassert i grunnkretsens bebyggelsesmessige eller transportmessige tyngdepunkt. I ettertid har plasseringene blitt redigert ved endring av grunnkretsstruktur. Det er derfor stor grunn til å tro at grunnkretsene ligger fornuftig plassert i RTMs nettverk.

I RTM¹ beregnes det LoS-matriser (Level-of-Service-matriser) mellom hvert sonepar. Siden RTM har grunnkrets som soneenhet, betyr dette at vi får grunnkrets-til-grunnkrets matriser som inneholder informasjon om distanse, reisetid o.l. mellom alle par av grunnkretser. Ved å sette enhver togstasjon som er kodet inn i nettverket til å være en sone, vil man i tillegg kunne beregne LoS mellom stasjoner, samt grunnkrets-til-stasjonsverdier.

En slik LoS-beregning gir oss distansen (langs veg i nettverket) mellom alle grunnkretser og alle stasjoner (togstasjoner som er kodet inn i nettverket). For hver grunnkrets ønsker vi å finne nærmeste togstasjon, og dette kan vi gjøre ved hjelp av LoS-beregningene. Se Figur 3.1 for eksempel på hvordan grunnkretser fordeler seg langs Østfoldbanen. For oversikt over fordeler og ulemper ved valgt tilnærming, se Tabell 3.4.



Figur 3.1: Eksempel på hvordan grunnkretser fordeler seg til togstasjoner langs Østfoldbanen. De oransje grunnkretsene har Myrvoll som nærmeste stasjon, de grønne Greverud og de blå grunnkretsene har Oppegård som nærmeste stasjon.

¹ RTM bruker Cube Voyager, et av flere programmer for transportplanlegging i Cube, som er en kommersielt tilgjengelig programpakke.

Tabell 3.4: Fordeler og ulemper ved valgt tilnærming for kobling mellom grunnkretser og togstasjoner.

Fordel	Ulempe
Grunnkretsen er plassert der hvor tyngdepunktet i grunnkretsen er	Ikke alle stasjoner er kodet inn
Realistisk at det er distanse langs vei som bestemmer stasjonsvalg, ikke luftdistanse	Nettverkene som brukes i RTM er hovedsakelig basert på uttak av regionale veier og fylkesveier fra Nasjonal vegdatabank (NVDB). I tillegg inkluderes kommunale veier etter skjønsmessige vurderinger. Dette betyr at detaljnivået på vegnettet i modellen vil variere, og kan føre til at enkelte soner får et dårligere tilbud enn det som er reelt.
Grunnkretsene som er kodet inn i RTM er tilpasset datasettene som ellers benyttes i prosjektet	For togstasjoner som ikke er kodet inn må man gjøre manuelle vurderinger/manuell koding

Prosessen fungerer kun for stasjoner som er kodet inn i nettverket. Det er i tillegg nødvendig at togstasjonen er kodet inn med veiforbindelse. Kvaliteten på kodingen varierer fra region til region, og det har derfor vært nødvendig å supplere med litt manuell koding. Dette innebærer bl.a. at vi for noen stasjoner som vi mener eksisterer har kodet inn manglende veiforbindelse, samt at vi har gjort opprettinger i noen eksisterende veiforbindelser. Vi har også gjort en sammenligning opp mot NSBs stasjonsoversikt² for å få en oversikt over hvilke stasjoner det var utfordringer med. For en oversikt over disse stasjonene, se vedlegg.

3.2.2 Beregning av indeksverdier for togstasjoner/togstasjonsrelasjoner

Med opplegget beskrevet i forrige avsnitt får vi distanse langs bilvei fra hver grunnkrets til nærmeste togstasjon. Basert på det kan vi omregne sonedata (avsnitt 3.1.2) og pendlerdata (avsnitt 3.1.5) til togstasjonsnivå. Ved modellering skal disse dataene representere egenskaper ved togstasjoner, for eksempel antall arbeidsplasser i nærheten av stasjonen.

En utfordring er hvordan man definerer «nærhet» til stasjon. I det valgte opplegget bruker vi en intern vektning av inndata basert på distanse av grunnkrets til togstasjon. Tabell 3.5 viser vektene som brukes i beregningen for sonedata fra RTM.

Tabell 3.5: Vektene som brukes for indeksberegning av sonedata.

Distanse fra grunnkrets til stasjon	Brukt vekt.
≤ 2 km	1
> 5 og ≤10 km	0,6
> 10 og ≤20 km	0,4
> 20 og ≤50 km	0,2
>50	0

² <https://www.nsb.no/reisemal/stasjoner> pr. desember 2016

Vektene tolkes slik: Om tyngdepunktet til en grunnkrets ligger under 2 km fra nærmeste togstasjon vektes sonedata for denne grunnkretsen med full verdi for den gitte togstasjonen (for eksempel antall arbeidsplasser); om tyngdepunktet til grunnkretsen ligger 5-10 km fra togstasjonen så tas kun med 60 % av antallet for denne grunnkretsen (60 % av arbeidsplassene i grunnkretsen tilegnes togstasjonsindeksen), osv. Grunnkretser som ligger mer enn 50 km unna togstasjonen medregnes ikke i indeksen.

Mens sonedata fra RTM gjelder en grunnkrets (og etter omregning en togstasjon), kommer pendlerdata på grunnkretsrelasjonsnivå, dvs. par av bostedsgrunnkrets og arbeidsgrunnkrets).

Dermed vil «nærhet til togstasjon» både avhenge av distansen fra bostedsgrunnkrets til startstasjon (tilbringerdistanse) og sluttstasjon til arbeidsgrunnkrets ('frabringerdistanse') (eller omvendt ved retur-reise). For beregning av indeksdata for pendlerreiser tar vi – på forenklet vis- utgangspunkt i summen av tilbringer- og frabringerdistanse. Tabell 3.6 rapporterer vektene som har blitt brukt.

Tabell 3.6: Vektene som brukes for indeksberegning av pendlerdata.

Distanse fra grunnkrets til stasjon	Brukt vekt
≤ 2 km	1
> 5 og ≤ 10 km	0,8
> 10 og ≤ 20 km	0,6
> 20 og ≤ 50 km	0,4
> 50 og ≤ 100 km	0,2
> 100 km	0

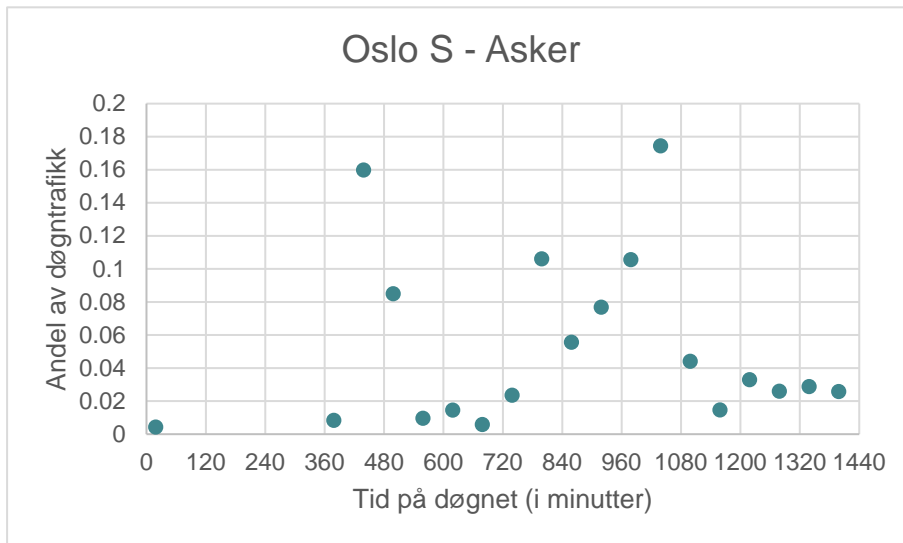
3.3 Opplegg for å konvertere snitt-tellinger til OD-matriser

Siden døgnfordelingene skal være på stasjon-til-stasjonsnivå er vi avhengig av å ha tidfestet data på stasjonsrelasjonsnivå.

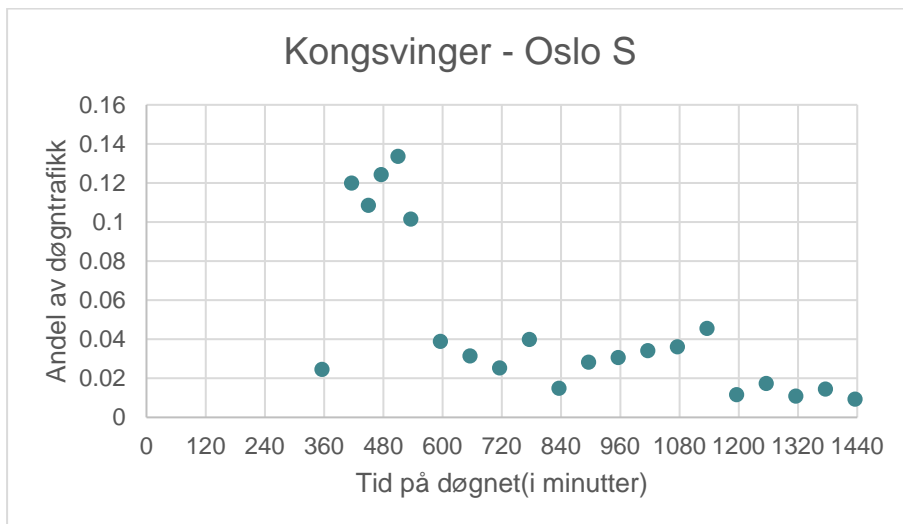
I dette prosjektet har vi hatt tilgang til snitt-tellinger og tilhørende rutetabell for ti strekninger (se avsnitt 3.1.4). Snitt-tellingene forteller hvor mange som sitter på toget mellom to stasjoner, men de forteller ikke hvilke stasjoner passasjerene reiser mellom (hvor de går på og av). Som nevnt i avsnitt 2.2 har vi derfor utarbeidet et opplegg for å konvertere data på snitt til data på OD-matrisenivå (origin – destination). Metoden kan oppsummeres i disse stegene:

1. Beregning av netto av- og påstigninger for hver stasjon og hver avgang
2. Estimering av antall avstigninger (og påstigninger) for hver stasjon og hver avgang. Til dette benyttes:
 - a. Snitt-tellinger
 - b. Andeler av-og påstigninger utarbeidet fra NSBs årsmatrise
 - c. Rutetabell (stoppmønster)
3. Etter å ha estimert antall påstigninger per stasjon, fordeler vi disse reisende basert på et typisk reisemønster (utarbeidet fra NSBs årsmatrise), samt tilgjengelige stasjoner (hvor toget stopper). Dette resulterer i en OD-matrise for påstigninger per avgang.

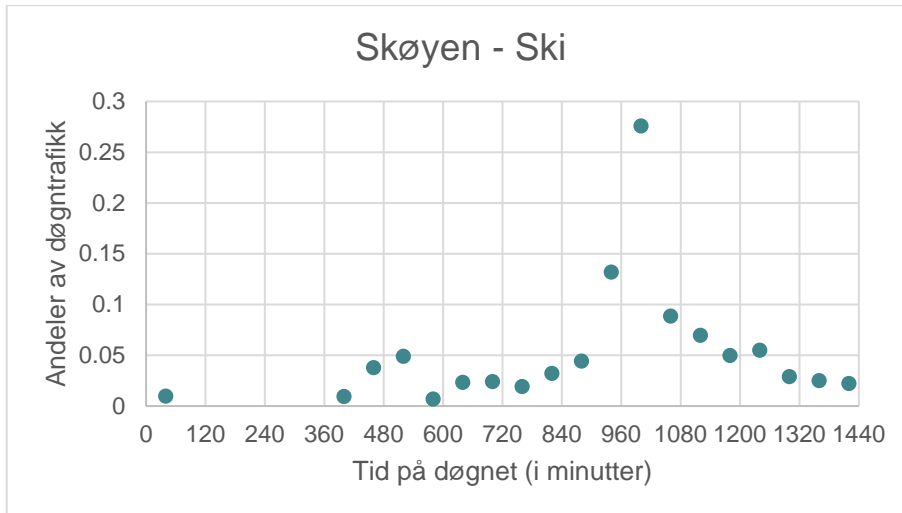
4. En OD-matrise for avstigninger (per avgang) utarbeides på samme måte som for påstigninger.
5. OD-matrisen for avstigninger og OD-matrisen for påstigninger skaleres slik at andelene summeres til 1.
6. De 2 OD-matrisene vektet sammen til 1 OD-matrise per avgang.
7. Ved å tidfeste ankomsttidene ved hjelp av rutetabellene kan vi lage døgnfordelingskurver for hver stasjon-stasjonsrelasjon. Se Figur 3.2 - Figur 3.4 for eksempler på slike døgnfordelingskurver, der den horisontale akse viser ankomsttid ved endestasjon i minutter målt fra midnatt (dvs. 480 er kl. 8 om morgenen, 960 kl. 16 om ettermiddagen).



Figur 3.2: Estimert døgnfordeling for reiser mellom Oslo S og Asker.

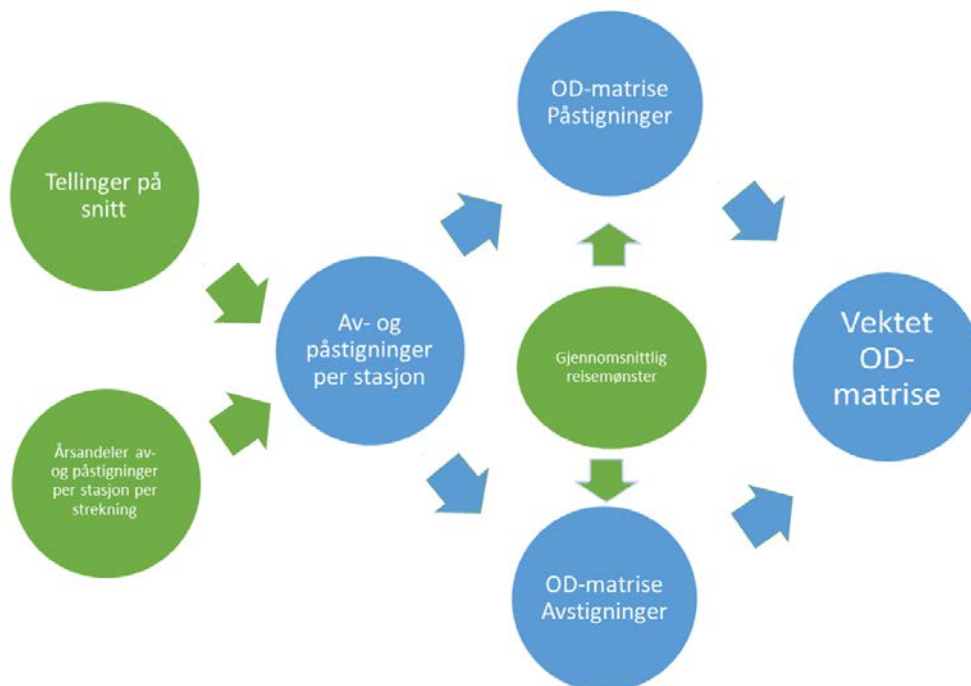


Figur 3.3: Estimert døgnfordeling for reiser mellom Kongsvinger og Oslo S.



Figur 3.4: Estimert døgfordeling på strekningen Skøyen – Ski.

Figur 3.5 viser dataflyten i denne konverteringen, hvor de grønne sirklene viser inndata (telldata). Telldata på snitt er passasjertellingene presentert i avsnitt 3.1.4, mens reisebøns og andeler av- og påstigninger er bearbejninger av NSBs årsmatrise.



Figur 3.5: Dataflyt snittnivå til stasjonsrelasjonsnivå (grønn: inndata, blå: interne beregninger).

3.3.1 Kommentarer til OD-matrisene

De vektete OD-matrisene må ansees å være veiledende. På grunn av tilgjengeligheten på data i dette prosjektet har vi benyttet et typisk reisebøns over året for å fordele trafikken langs toglinjen. Det er verdt å merke seg at et slikt reisebøns over et helt år, ikke nødvendigvis er representativt for hvordan trafikken fordeles seg over døgnet.

4 Etablering av reisehensiktsmodell

4.1 Definisjon av reisehensikter og fordeling i RVU

I Trenklin er etterspørselen segmentert i tre reisehensikter: arbeid, tjeneste og fritid/andre formål. Arbeidsreiser er reiser til/fra arbeid og inkluderer etter vårt skjønn også skolereiser. Tjenestereiser er reiser i arbeidstiden, hvor reisen typisk er betalt av arbeidsgiver. Fritid/andre formål fanger opp alle reiser som ikke er arbeids- eller tjenestereiser.

Vi har disaggregert disse tre formålene i undergrupper basert på inndelingen i RVU. Begrunnelsen for dette er at det antas at egenskaper ved togstasjoner har ulik effekt på (undergrupper av) reisehensikter. Antall fritidshus i nærheten av en togstasjon påvirker først og fremst andelen fritidsreiser, men ikke andelen handleturer osv. Estimerte andeler summeres enkelt opp til de 3 hovedreisehensiktene i Trenklin.

Ved å bruke en mest mulig disaggregert reisehensiktsgruppering ved modellering unngår vi en del av «aggregeringsfeilen» og kan anta ulike effekter for ulike undergrupper.

Tabell 4.1 viser antall observasjoner og andeler fordelt på ulike reisehensikter i RVU for henholdsvis «alle reiser» og for reiser med tog som hovedtransportmiddel.

Tabell 4.1: Andeler for ulike grupper reisehensikter i RVU.

Formål	Totalt (alle typer transportmidler)		Bare togreiser	
	Antall	Prosent	Antall	Prosent
Arbeid	32343	19,9	1264	56,4
Skole	5639	3,5	137	6,1
Tjeneste (reiser i arbeid)	5043	3,1	144	6,4
Handle/service	48165	29,6	189	8,4
Følge/omsorg	16750	10,3	43	1,9
Fritid	30564	18,8	210	9,4
Besøk	15886	9,8	156	7,0
Annet	8351	5,1	97	4,3
Totalt	162741	100,0	2240	100,0

Sammenlignet med «alle reiser» brukes tog i mye større grad til reiser til/fra arbeid og i mindre grad til handlereiser og reiser for følge/omsorg. Andelen fritidsreiser er også lavere på tog. Grunnen ligger trolig i at bil som transportmiddel er mer egnet for formål der man for eksempel transporterer varer eller er flere i følge.

For etablering av reisehensiktsmodellen har vi brukt kun togreiser til tross for et relativt lavt antall observasjoner. Fra et modellteknisk perspektiv er det mulig å inkludere alle typer reiser og tilføre dummy-variabler for togreiser i modellen. Denne tilnærming har dog den ulempen at man må anta at parameterne som beskriver sammenhengen mellom egenskaper ved stasjoner/stasjonsrelasjoner er lik for togreiser og reiser med andre transportmidler. Vi har vurdert denne antakelsen som problematisk i og med at reiser med bil kan være nokså ulike reiser med tog selv om reisehensikten er lik.

For estimering av modellen trengs nok observasjoner for hvert formål. Siden 43 observasjoner for «følge/omsorg» vurderes for lavt har vi slått sammen kategorien med «annet» til en ny kategori «andre reiser (inkl. følge og omsorg)».

Tabell 4.2 viser oppdeling av formål av togreiser i RVU oppdelt i virkedøgn og restdøgn.

Tabell 4.2: Andeler for togreiser i RVU oppsplittet i virkedøgn og restdøgn.

	Virkedøgn		Restdøgn	
	Antall	Prosent	Antall	Prosent
Arbeidsreiser	1259	57,5%	118	32,0%
Skolereiser	164	7,5%	1	0,3%
Tjenestereiser	153	7,0%	18	4,9%
Hand og Service	190	8,7%	24	6,5%
Fritidsreiser	180	8,2%	86	23,3%
Besøksreiser	118	5,4%	80	21,7%
Andre reiser (inkl. følge og omsorg)	127	5,8%	42	11,4%
Sum	2191	100%	369	100%

Det kan overraske at det er et høyt antall arbeidsreiser også for restdøgn, som inkluderer helger og hele juli måned. Skolereiser opptrer veldig sjelden på restdøgn. Andelen fritidsreiser og besøksreiser er naturlig nok større på restdøgn.

Reiseavstand (eller reisetid) antas å være en viktig forklaringsvariabel for reisehensikt. Vi har koblet reisetiden for togreiser (bestående av ombordtid og ventetid ved bytting) fra nasjonale tilbudsdata (avsnitt 3.1.3) til RVU observasjoner.

Tabell 4.3 viser hvordan reisehensikter fordeler seg innenfor ulike intervaller av reisetid.

Tabell 4.3: Fordeling av reisehensikter for ulike reisetidsgrupper for togreiser i RVU.

Reisetid tog(min)	Arbeid		Skole		Tjeneste		Handle/serv.		Fritid		Besøk		Annet	
	Ant.	Andel (%)	Ant.	Andel (%)	Ant.	Andel (%)	Ant.	Andel (%)	Ant.	Andel (%)	Ant.	Andel (%)	Ant.	Andel (%)
0-15	138	45.7	26	8.6	25	8.3	48	15.9	25	8.3	19	6.3	21	7.0
15-30	348	62.0	34	6.1	27	4.8	42	7.5	61	10.9	14	2.5	35	6.2
30-45	280	63.5	20	4.5	24	5.4	26	5.9	40	9.1	28	6.3	23	5.2
45-60	175	61.2	17	5.9	19	6.6	21	7.3	19	6.6	16	5.6	19	6.6
60-90	156	60.0	20	7.7	9	3.5	20	7.7	21	8.1	24	9.2	10	3.8
90-120	55	47.8	8	7.0	8	7.0	13	11.3	15	13.0	10	8.7	6	5.2
120-180	31	41.3	2	2.7	14	18.7	3	4.0	9	12.0	11	14.7	5	6.7
180-240	5	25.0	2	10.0	4	20.0	0	0.0	3	15.0	5	25.0	1	5.0
240-480	4	10.5	1	2.6	6	15.8	2	5.3	8	21.1	14	36.8	3	7.9
480-600	1	9.1	0	0.0	0	0.0	1	9.1	1	9.1	5	45.5	3	27.3
over 600	1	33.3	1	33.3	0	0.0	0	0.0	0	0.0	0	0.0	1	33.3

Ved å studere tabellen kan man se at sammenheng mellom andeler og distanse ikke er lineær for flere reisehensikter. Andelen tjenestereiser virker for eksempel å være høyest for reiser mellom 120-480 minutter.

En stor utfordring er at datagrunnlaget for lange reiser er veldig tynt. Det er kun 14 observasjoner over 480 minutter. Det er i seg selv altfor lite til å beregne pålitelige andeler for lange relasjoner. Dette må tas hensyn til ved spesifisering av modellen (avsnitt 4.3).

4.2 Statistisk modell

Reisehensiktfordelingen for stasjonspar s,e (der s er startstasjon, e er endestasjon) vil avhenge av antall ulike reiser som genereres nær stasjon s og attraheres nær e . De matematiske fordelingene for et diskret antall reisehensikter karakteriseres ved at alle andelene er mellom 0 og 100% og at andelene summeres opp til 100%. En type statistisk modell som er egnet for å beregne slike andeler basert på forklaringsvariabler, er logit-modeller. Disse er forholdvis enkle å sette opp og estimere.

Logitmodeller brukes ofte for å modellere valg av transportmidler, der man spesifiserer såkalte nyttefunksjoner for å beregne nytten til hver alternativ (tog, buss, bil, sykkel osv.). En parameter i en typisk lineær nyttefunksjon vil da representere den marginale nytten en egenskap har når man velger et alternativ.

I vår sammenheng kan det være noe misvisende å kalle funksjoner som driver andeler for reisehensikter for «nyttefunksjoner» fordi egenskaper ved stasjoner (antall arbeidsplasser, hoteller etc) ikke skaper «nytte» (i en økonomisk forstand) for beslutningstakeren. Vi bruker derfor heller uttrykket «genereringsfunksjon», altså en matematisk funksjon som forklarer hvilke reisehensikter som genereres mellom to togstasjoner.

Følgende matematiske funksjon brukes for å modellere reisehensikt:

$$P_{i,se} = \frac{\exp V_{i,se}}{\sum_j \exp V_{j,se}}$$

Der $P_{i,se}$ er sannsynligheten (andel) for reisehensikt i ved togstasjonspar se . $V_{i,se}$ er genereringsfunksjonen for reisehensikten i , mens $\sum_j \exp V_{j,se}$ summerer opp over alle reisehensikter (reisehensikt i , samt øvrige reisehensikter). Genereringsfunksjonene er spesifisert slik:

$$V_{i,se} = \beta_{i,0} + \beta_{i,restd.} + \sum_k \beta_{i,k} * X_{k,se}$$

Der $\beta_{i,0}$ er konstantledd, $\beta_{i,restd.}$ er dummy-variable for restdøgn, $X_{k,se}$ er forklaringsvariabler (indeksert med k) for togstasjonspar se og $\beta_{i,k}$ tilsvarende estimerte parametere gitt data.

4.3 Forklaringsvariabler i modellen

I dette avsnittet beskriver vi et utvalg av forklaringsvariabler ($X_{k,se}$ i forrige avsnitt) som inngår i modellen.

Det er tre ulike typer forklaringsvariabler i modellen

- 1) Dummy variabler for spesielle togstasjonsrelasjoner (1/0 variable)
- 2) Indeksverdier basert på sonedata (kontinuerlige i antall)
- 3) Reisetid mellom togstasjoner (kontinuerlig i minutter)

Disse variablene vil variere med togstasjonsrelasjon s,e (startstasjon, endestasjon).³

Dummy variabler for spesielle togstasjonsrelasjoner

Tre dummy variabler er etablert for å fange opp egenskaper ved togstasjoner, ved at togstasjonene defineres som enten a) lav/spredt befolkning rundt stasjonen, b) tilknyttet en flyplass, eller c) representerer storbysteder.

Tabell 4.4 angir kriterier som blir brukt for å etablere disse dummy variablene.

Tabell 4.4: Dummyvariabler for utvalgte togstasjonsrelasjoner i reisehensiktsmodellen.

Type	Kriterium for togstasjon	Navn i modellen
Spredt befolket («landlig»)	Befolkningsindeks rundt stasjon mindre enn 5,000	d_landlig
Primært for flyplass	Gardermoen, Rygge, Torp, Værnes	d_flyplass
Storbysteder	Oslo S, Bergen, Trondheim S	d_storbysteder

Indeksverdier basert på sonedata

Som beskrevet i avsnitt 3.2.2 har vi etablert indeksverdier for togstasjoner basert på sonedata (grunnkrets nivå) og avstand mellom grunnkretser og togstasjoner.

Hvilke indekser som inngår i hvilke genereringsfunksjoner er et resultat av uttesting. I flere runder har vi inkludert og ekskludert indeksverdier til vi har funnet et utvalg som er logisk og gir forventede effekter (riktig fortegn på parameteren $\beta_{i,k}$). Tabell 4.5 gir en oversikt over hvilken indeks som inngår i hver av genereringsfunksjonene.

³ I motsetning til typiske diskrete valgmodeller (f.eks. transportmiddelvalg), er disse valgmodellene konstante på tvers av alternativene (reisehensikt). Derfor må alltid en koeffisient til et alternativ normaliseres.

Tabell 4.5: Oversikt over indekser som inngår i reisebetsmodellene.

Indeks, sum av start- og endestasjon	Arbeid	Skole	Tjeneste	Fritid	Besøk	Handel og Service	Andre
Arbeidsplasser totalt (i tusen) (arb_totalt_korr_index_es_sum_1000)	X		X				
Arbeidsplasser salg av mat/overnatting (a33vh_index1_es_sum)				X			
Arbeidsplasser innenfor helsestudio-, massasje- og solstudiovirksomhet (a41tje_index1_es_sum)				X			
Arbeidsplasser i undervisning (a60und_index1_es_sum)		X					
Arbeidsplasser i detaljhandel med varer vi kjøper ofte (a31vh_index1_es_sum)						X	
Arbeidsplasser i detaljhandel med varer vi kjøper sjelden (a32vh_index1_es_sum)						X	
Arbeidsplasser i reparasjoner/verksteder, skomaker, etc (a34vh_index1_es_sum)						X	
Antall hoteller (numhots_index1_es_sum)							X
Antall hytter/fritidshus (hytfrit_index1_es_sum)				X			
Antall videregående skoler (vgskole_index1_es_sum)		X					
Antall plasser høyskole/universitet (uhskole_index1_es_sum)		X					
Befolkning (i tusen) (mk_all_index1_es_sum_1000)					X		

For å møte JBV's ønske om at modellen skal være retningsbalansert, (dvs. at $P_{i,se} \cong P_{i,es}$) har vi i den endelige modellspesifisering brukt summen av indeks for start- og endestasjon som inndata i modellen. Modellen skiller derfor ikke på om egenskapen er ved start- eller endestasjon.

Reisetid mellom togstasjoner

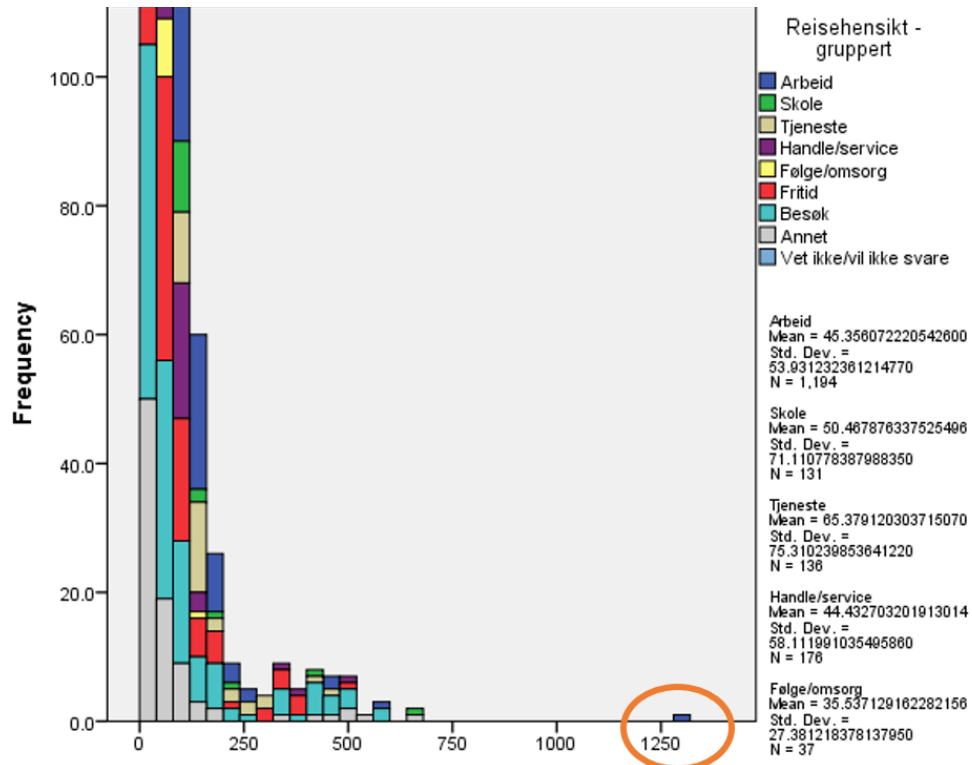
Reisetid definert i de nasjonale tilbudsmatrisene (avsnitt 3.1.3) inngår også i modellen. Reisetiden i modellen er summen av ombordtid på tog og ventetid ved et eventuelt togbytte (men ikke ventetid ved første togavgang). Fra et modelleringsperspektiv er denne variabelen mest krevende fordi sammenhengen antas å være ikke-lineær og fordi vi har et veldig tynt datagrunnlag for lange reisetider (se Tabell 4.3 i avsnitt 4.1).

Ulike spesifikasjoner har blitt testet for å fange opp ikke-lineariteten. Mye av uttestingen gikk på å finne meningsfulle knekkpunkter. Tynt datagrunnlag for lange reisetider gjorde det spesielt vanskelig å estimere meningsfulle knekkpunkter for lange reiser. Ved den endelige spesifiseringen har vi brukt en lineær effekt pluss en potenseffekt av reisetiden. De to variablene som inngår i genereringsfunksjonene er dermed reisetid og reisetid² («reisetid_sq»). Potensstermen er delt med 1000 for å unngå høye verdier i forklaringsvariabelen (dette er en ren normalisering som ikke påvirker resultatet).

4.4 Estimering

I estimeringen har vi tatt utgangspunkt i reiser i RVU hvor tog er oppgitt som reisemåte. Vi har måttet ekskludere observasjoner fra RVU hvor vi ikke har klart å koble til sonedata eller reisetid.

Videre har vi valgt å ekskludere én observasjon med en reisetid på rundt 1300 minutter (ca 22 timer), se markert observasjon i Figur 4.1. Reisetiden i seg selv kan være realistisk (med uheldig korrespondanse ved togbytte), men observasjonen er så avvikende i reisetid at den har stor innvirkning på estimatene. Ekskludering av denne uteliggeren førte til at modellen estimerte mer forventede sammenhenger mellom reisetid og andel arbeidsreiser.



Figur 4.1: Identifisering av uteligger.

Det endelige utvalget består av 2210 RVU-observasjoner.

Modellen er estimert i programvaren Biogeme (Bierlaire 2008) basert på maksimum likelihood metoden, dvs. at programmet finner et sett av parameterverdier som maksimerer sannsynligheten for observerte data. Det estimeres til sammen 58 parametere. Tabell 4.6 rapporterer de estimerte verdiene, samt tilhørende robuste t-verdier.

Ikke alle parameterne er signifikante (t-verdier på under 1,96 i absoluttverdi), men på grunn av tynt datagrunnlag har vi ikke valgt å ekskludere forklaringsvariabler så lenge parameterne har forventet fortegn.

Konstantleddene er alle lavere enn 0, som tilsvarer verdien vi valgte for normalisering av konstantledd til arbeidsreiser. Det er forventet siden arbeidsreiser har de største andelene i utvalget i RVU.

Parameterne til dummy variablene for restdøgn er negative, og store i absoluttverdi for skolereiser siden denne typen reiser er sjelden for restdøgn (som vist i Tabell 4.2, hvor det kun er én skolereise på restdøgn i RVU). Effekten av restdøgn på fritidsreiser, besøksreiser og andre reiser er som forventet positiv og signifikant. Effekten for tjenestereiser og handelsreiser er også positiv, men svakere i størrelse for effekt og signifikans.

Ved modellering av reisetid har vi normalisert parameteren til besøksreiser (fordi vi ved uttesting hadde behov for å modellere reisetidsfunksjoner for arbeidsreiser – den antatt viktigste reisehensikten – mer direkte). Det er ikke lett å tolke parameterverdien og vi viser til Figur 4.2 i neste avsnitt for en illustrasjon av hvordan reisetid påvirker sannsynligheten for ulike reisehensikter.

Dummy variabler for relasjoner til/fra «landlig», flyplass og storbystasjoner har ulike effekter sammenlignet med normalisert reisehensikt. Sannsynligheten for tjenestereiser er større for storby- og flyplass-stasjoner, og mindre sannsynlig for landlige stasjoner. Dette er som forventet. Fritidsreiser (og andre reiser) er mer sannsynlig for storbysenter, noe som kan henge sammen med kulturtilbudet i nærheten av disse stasjonene. En annen signifikant effekt er at fritidsreiser er mer sannsynlig fra/til flyplass-stasjoner.

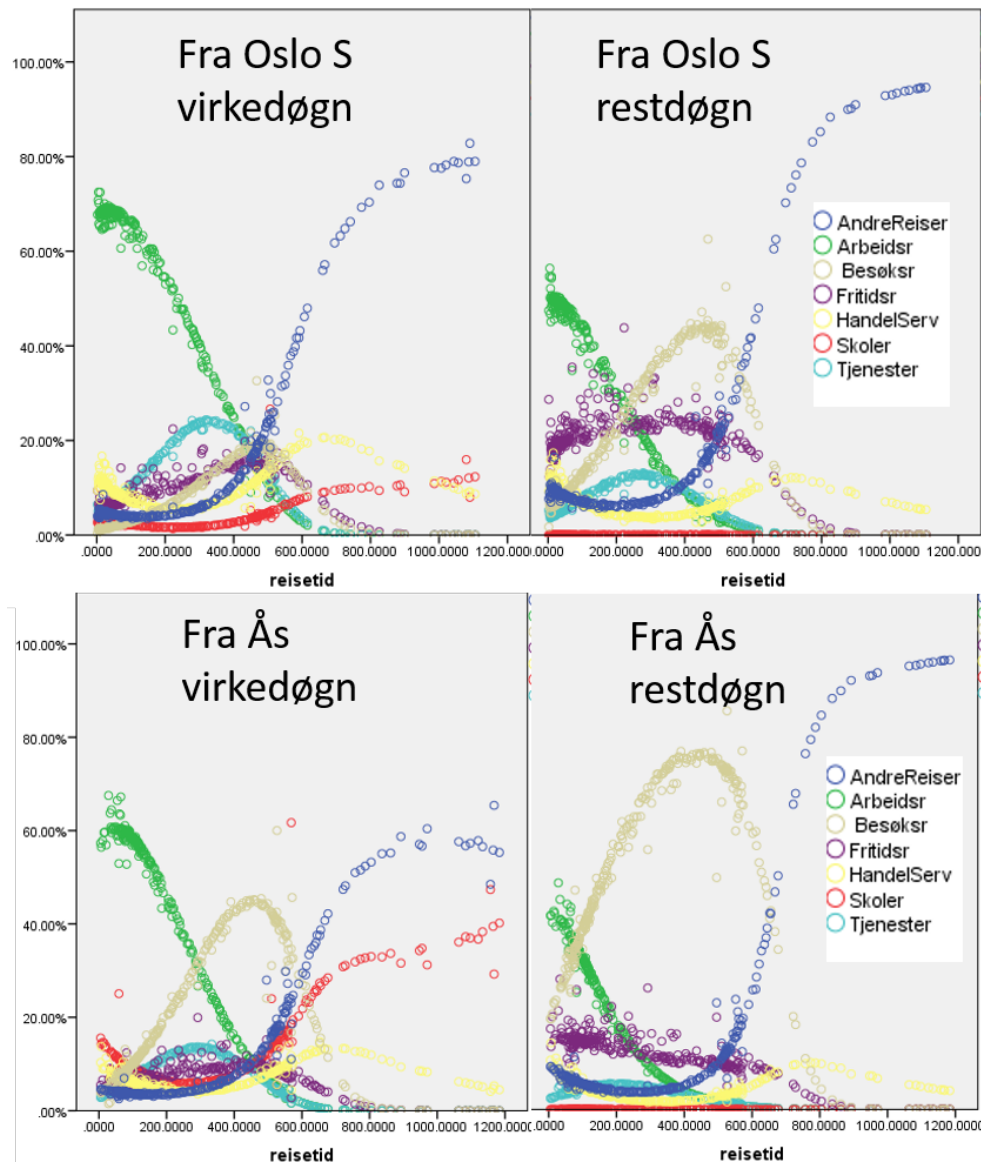
Alle parameterne for sonedata er positive. En positiv parameter betyr at en marginal økning i forklaringsvariabelen øker sannsynligheten for reisehensikteten. For eksempel har «arbeidsplasser totalt» en parameterverdi 0.0261. Dette betyr at flere arbeidsplasser rundt endestasjonen vil øke sannsynligheten for at en reise har arbeidsreise som hensikt. Som nevnt tidligere har vi i foregående testrunder ekskludert alle sonedata som ikke fikk fortegn som forventet.

Tabell 4.6: Estimeringsresultater for reisebensiktsmodell. (N=2210, antall parameter: 58, final LL: -2836.143; adjusted rho square: 0,295)

Parameter	Arbeidsreiser		Skolereiser		Tjenestereiser		Handel og Service		Fritidsreiser		Besøksreiser		Andre reiser	
	Verdi	t-verdi	verdi	t-verdi	verdi	t-verdi	verdi	t-verdi	verdi	t-verdi	verdi	t-verdi	verdi	t-verdi
konstantledd	0	normalisert	-1.58	-5.33	-3.1	-10.72	-1.76	-6.92	-2.65	-10.77	-2.79	-10.68	-2.54	-8.12
dummy for restdøgn	0	normalisert	-7.46	-50.26	0.145	0.49	0.343	1.34	1.47	7.79	1.93	9.01	1	4.07
Reisetid (minutter)	-0.0097	-2.83	-0.0182	-3.95	0.000446	0.1	-0.0172	-4.38	-0.00824	-2.21	0	normalisert	0.0234	3.17
Reisetid ² (1000 minutter)	0.000972	0.09	0.0256	2.58	-0.00756	-0.81	0.0236	3.23	0.00915	1.24	0	normalisert	-0.00167	-0.2
Dummy til/fra landlig	0	normalisert	0.0285	0.11	-0.653	-2.26	0.232	1.2	0.323	1.71	-0.457	-1.72	0.21	0.86
Dummy til/fra storbysent	0	normalisert	-0.739	-0.72	1.85	5.29	-0.451	-0.59	0.972	2.34	-0.949	-1.16	1.67	3.81
Dummy til/fra flyplass-stasjon	0	normalisert	0.147	0.51	0.468	2.32	-0.217	-0.65	0.257	1.2	0.803	3.06	0.493	1.71
Arbeidsplasser totalt (i tusen)	0.0261	7.04			0.0302	6.93								
Arbeidsplasser salg av mat/overnatting														
Arbeidsplasser innenfor helsestudio-, massasje- og solstudiovirksomhet									0.00273	2.58				
Arbeidsplasser i undervisning			0.000256	4.37										
Arbeidsplasser i detaljhandel med varer vi kjøper ofte							0.000243	0.5						
Arbeidsplasser i detaljhandel med varer vi kjøper sjelden							0.000333	2.48						
Arbeidsplasser i reparasjoner / verksteder, skomaker, etc.							0.00707	0.57						
Antall hoteller													0.0246	1.26
Antall hytter/fritidshus									0.000558	3.02				
Antall videregående skoler			0.0000551	0.5										
Antall plasser høyskole /universitet			0.000022	1.64										
Befolkning (i tusen)											0.0115	2.05		

4.5 Illustrering av resultater

Etter at parameterne er estimert kan vi bruke modellen for å beregne hensiktsfordelingene (for hhv virkedøgn og restdøgn) mellom alle togstasjonsrelasjoner i Norge (totalt 124 254 relasjoner). Figur 4.2 plotter andelene til reisehensiktsfordeling mot reisetid (i minutter) for utvalgte togstrekninger.



Figur 4.2: Reisehensiktsandeler etter reisetid. Hvert punkt representerer en endestasjon med henholdsvis Oslo S og Ås som startstasjon.

Andelen arbeidsreiser er klart høyest for korte relasjoner og avtar mer eller mindre lineært. For reiser over 700 km er andelen tilnærmet null. Det kan overraske at andelen arbeidsreiser er relativt høy (mellom 40% og 15 %) også for reiser mellom 200 og 400 km. Ved tolkning må det huskes at det er relative sammenhenger vi estimerer. Det totale antallet arbeidsreiser avtar mye sterkere enn det kan virke som her. Dette fordi antall

togreiser reduseres sterkt med økende distanse (se Tabell 4.3). Andelen tjenestereiser med tog øker først, og avtar igjen etter rundt 200-400 km (avhengig av relasjon og type døgn). Andelen til de andre reisehensikter er noe mer varierende, men i sum vil de andre reisehensiktene (dvs. utenom arbeids- og tjenestereisene) øke med distanse og vil være tilnærmet 100% for de lengste relasjonene (Oslo-Bodø for eksempel).

Sammenhengen mellom distanse og reisehensiktsandel virker relativt sterk og vil gjenkjennes for alle togstasjonsrelasjoner. Det virker som om sonedata og dummyvariabler forklarer relativt lite av variasjonen vi observerer i data. Tabell 4.7 rapporterer andeler for Trenklins tre reisehensiktsgrupper for stasjonspar i Oslo-Trondheim-korridoren, mens Tabell 4.8 gjør det samme for Oslo-Bergen-korridoren.

Tabell 4.7: Predikerte andeler for Trenklins hovedreisehensikter på utvalgte strekninger/ stasjonspar.

Predikerte andeler for virkedøgn i %			Til stasjon			
			Lillehammer	Lillestrøm	Oslo S	Trondheim S
Fra stasjon	Lillehammer	Arbeid		58	58	30
		Tjeneste		8	12	14
		Andre reiser		34	29	56
	Lillestrøm	Arbeid	58		71	21
		Tjeneste	8		5	12
		Andre reiser	34		24	67
	Oslo S	Arbeid	59	71		13
		Tjeneste	12	5		20
		Andre reiser	29	24		66
	Trondheim S	Arbeid	32	20	13	
		Tjeneste	14	12	20	
		Andre reiser	54	68	68	

Tabell 4.8: Predikerte andeler for Trenklins hovedreisehensikter på utvalgte strekninger/ stasjonspar.

Predikerte andeler for virkedøgn i %			Til stasjon			
			Bergen	Drammen	Geilo	Oslo S
Fra stasjon	Bergen	Arbeid		24	44	16
		Tjeneste		16	11	26
		Andre reiser		60	45	57
	Drammen	Arbeid	22		47	72
		Tjeneste	14		11	7
		Andre reiser	65		42	21
	Geilo	Arbeid	39	47		44
		Tjeneste	13	11		17
		Andre reiser	49	42		39
	Oslo S	Arbeid	14	72	45	
		Tjeneste	22	7	17	
		Andre reiser	64	21	39	

Noe overraskende er det en stor andel arbeidsreiser på relasjoner som Bergen-Geilo og Oslo S-Geilo. Det virker som modellen i liten grad fanger opp effekten av «ski-steder». Som omtalt senere i avsnitt 4.7 var det ikke nok observasjoner i RVU for å måle denne effekten direkte.

Tabell 4.7 og Tabell 4.8 viser at modellen predikerer tilnærmet retningsbalanserte andeler som etterspurt av JBV. Det er noen mindre avvik pga. at reisetiden (inkludert ventetid ved bytte av tog) som vi fikk tilsendt fra JBV (se avsnitt 3.1.3) viser noe variasjon med hensyn på retning. Modellen blir derfor ikke helt retningssymmetrisk (for eksempel er oppgitt reisetid mellom Oslo S og Bergen ulik oppgitt reisetid mellom Bergen og Oslo S).

4.6 Implementert modell

Modellen er implementert i Microsoft-Excel og er lagret under filnavnet *Reisehensiktsmodell_V1.xlsx* (med lagringsdato 09.12.2016 kl. 11:08). Filen er stor (267 MB) og kan derfor ikke vedlegges i vanlig epost.

Filen inneholder inndata, parameterverdier, beregningsligninger og resultater for reisehensikter mellom togstasjoner i Norge (og noen svenske stasjoner). For de stasjoner der vi ikke kunne koble sonedata har vi brukt en «reduert» modell der bare konstanter og reisetid inngår (estimert i Biogeme på samme måte som hovedmodellen). Det ligger også togstasjoner inne uten reisetider for 2014, for disse er det ikke beregnet andeler og de er filtrert bort i resultatet.

Filen har 6 ark:

- 1) **Resultatark:** viser beregnede markedsandeler for alle togstasjonsrelasjoner. Togstasjoner uten reisetid er filtrert ut som default, men kan inkluderes om ønsket (trykk på filterknappen i kolonne «Tilbudsdata_tiltj»)
- 2) **Stasjonsinfo:** viser for hvilke stasjoner det er manglende informasjon om sonedata og om reisetider
- 3) **Inndata:** inneholder verdier for alle forklaringsvariabler som inngår i modellen
- 4) **Parameterverdier:** viser parameterverdier for hovedmodell og redusert modell.
- 5) **Beregning_full:** inneholder alle beregninger for hovedmodellen
- 6) **Beregning_red:** inneholder alle beregninger for den reduserte modellen

Det er ulike muligheter for å bruke modellen:

- a) **Endring/oppdatering av inndata:** Når data i arket «inndata» endres vil fordelingene i resultatarket oppdateres.
- b) **Kalibrering av parameterne:** Gitt ny og bedre informasjon om de «ekte» reisehensiktsandeler i populasjonen kan man kalibrere konstantledd i ark «parameterverdier» slik at resultatene bedre tilsvarer forventet nivå.
- c) **Legge inn nye togstasjonsrelasjoner:** Her legger man inn nye linjer i resultat-, inndata- og beregningsark («beregning_full» om man har tilgang til sonedata, «beregning_red» om man ikke har det). I beregningsarket må man «trekke» ned cellene slik at beregningsligninger flytter seg nedover.

4.7 Kjente svakheter og forbedringsmuligheter

Ressursene i prosjektet tillater ikke ytterligere forbedringer av modellen og nåværende versjon bør anses som første (men ikke endelige) versjon av modellen.

Modellen produserer stort sett intuitive resultater og gjenspeiler – etter vårt inntrykk – reise mønsteret i RVU på en god måte. Som nevnt tidligere er modellen en generisk modell som naturlig nok bommer på enkelte relasjoner.

Hovedutfordringen i modellen er at datagrunnlaget for togreiser i RVU er tynt – spesielt på lange relasjoner (se diskusjon rundt Tabell 4.3), samt relasjoner til/fra hytte- og skisteder. Etter det vi vet har JBV tilgang til kundetilfredsdata som også inkluderer spørsmål om reisehensikt. Dessverre fikk JBV nei til å videresende disse dataene til TØI. Vi antar at modellen kan forbedres ved tilgang til et større datagrunnlag.

Modellen virker å predikere relativt høye andeler for arbeidsreiser og tjenestereiser for mellomlange relasjoner (200-400 km). Det vil føre til at den gjennomsnittlige tidsverdien i Trenklin trolig vil øke for mellomlange relasjoner sammenlignet med dagens andeler som ligger i modellen. Vi forslår at dette testes ut. Generelt anbefaler vi at modellen brukes med varsomhet og at det gjøres følsomhetsanalyser angående andeler av reisehensikter.

Etter ønske fra JBV har vi mot slutten av prosjektet innført (tilnærmet) retningsbalanse i modellen. Dette virker å ha ført til at variasjonen i reisehensiktsandel mellom stasjonspar, men også totalt sett, har blitt redusert. Dette kan være meningsfullt i en transportmodellsammenheng dersom det reduserer noe av den uønskete variasjonen (som kan skyldes tilfeldig variasjon i inndata). Uavhengig av dette mener vi at den endelige modellen virker å predikere for homogene andeler. For eksempel er estimert effekt av til/fra flyplass-stasjon (til tross for være signifikant) relativt liten. I foregående modellversjoner var effekten større.

Konkrete forbedringsmuligheter kan inndeles i kortsiktige (uten nytt datagrunnlag) og mer langsiktige. På kort sikt anbefaler vi å teste videre på modellspesifisering, med konkret mål om å bedre fange opp effekten av hytte-/ski-/fritidsstasjoner (Geilo, Finse osv.), og å forbedre avstandsfordelingen. På lengre sikt anbefaler vi å re-estimere og oppdatere hele modellen med nye data.

5 Etablering av døgnfordelinger

5.1 Definisjon av døgnfordeling

I prosjektutlysningen står det at hensikten med prosjektet er «(...) å framskaffe bedre informasjon om hvordan etterspørselen etter togreiser fordeler seg over døgnet for ulike reisehensikter og stasjonsrelasjoner (...).

Med *stasjonsrelasjon* menes et togstasjonspar (i, j) der i er startstasjon og j er endestasjon (det vi kalte s og e i forrige kapittel). Her er det avgjørende hvilken stasjon som er start- og endestasjon. Eksempelvis vil relasjonen $(Moss, Oslo S)$ ha andre egenskaper enn $(Oslo S, Moss)$ når det gjelder tidspunkt for størst etterspørsel etter togreiser. Det er flere som pendler fra Moss til Oslo S enn omvendt, derfor vil det være størst etterspørsel om morgenen for relasjonen $(Moss, Oslo S)$, mens etterspørselen for togreiser i motsatt retning vil være størst om ettermiddagen.

For *reisehensikter* skal det skilles mellom arbeidsreiser, forretningsreiser og øvrige reiser. I dette kapitlet betegnes reisehensikt med bokstaven H og vi omtaler de tre hensiktene som: arbeid, tjeneste og andre, det vil si at $H \in \{\text{arbeid, tjeneste, andre}\}$.

Da døgnfordelingene skal brukes i transportmodellen Trenklin, må de oppfylle kravene spesifisert under. En døgnfordeling skal

1. tolkes som fordeling over **ønsket ankomsttid** ved endestasjon j , gitt startstasjon i og reisehensikt H .
2. ha egenskapene til en **sannsynlighetsfordeling**, det vil si at alle verdier er positive og summerer seg til én over døgnet.
3. være en diskret fordeling definert for hvert **minutt i døgnet**.

Punktene over tilfredstilles ved å la den tilfeldige variabelen T være ønsket ankomsttid for en togreise fra i til j med hensikt H .

Da er døgnfordelingen $f_{i,j}^H(t)$, $t \in \{0, 1, 2, \dots, 1439\}$, definert ved at

$$P(t_1 \leq T \leq t_2) = \sum_{t_1 \leq t \leq t_2} f_{i,j}^H(t).$$

Det vil si at sannsynligheten for at ønsket ankomsttid er mellom tidspunktene t_1 og t_2 , finnes ved å summere døgnfordelingens verdier over samme intervall.

Hensikten med prosjektet er å etablere døgnfordelinger $f_{i,j}^H(t)$ for alle kombinasjoner av i , j og H ⁴.

Det er 339⁵ stasjoner i Norge. Dette gir 114 582 stasjonsrelasjoner, noe som betyr at det i alt skal det etableres 343 746 døgnfordelinger⁶.

⁴ Her følger vi notasjonen fra utlysningsteksten i stedet for notasjonen for en betinget fordeling: $f_{i,j}(t|h)$.

⁵ Dette tallet er noe høyere hvis man inkluderer stasjonene i Sverige mellom Halden og Göteborg.

⁶ Antallet døgnfordelinger som skal etableres er $3n(n-1)$ der n er antall togstasjoner (3 fordi vi har 3 reisehensikter).

Datagrunnlaget er ikke tilstrekkelig til å kunne gi differensiert informasjon om et så høyt antall fordelinger. Vi må derfor etablere fordelingene basert på de relasjonene vi har informasjon om, samt karakteristika ved den aktuelle stasjonsrelasjonen. Slike generaliseringer kan avvike betraktelig fra de faktiske fordelingene, eller fordelingene som kunne vært etablert med et bedre datagrunnlag. Dette kommer vi tilbake til i kapittel 5.9.

5.2 Statistisk modell for døgnfordelingene

En *finite mixture* fordeling er en fleksibel og parametrisk tilnærming til å estimere en sannsynlighetsfordeling når den statistiske populasjonen består av to eller flere subpopulasjoner. Når vi skal estimere døgnfordelingen over ønsket ankomsttid for arbeidsreiser fra stasjon i til stasjon j kan det være naturlig å tenke seg at det forenklet sett finnes to subpopulasjoner:

1. Individuer som bor ved stasjon i og arbeider ved stasjon j .
2. Individuer som bor ved stasjon j og arbeider ved stasjon i .

Reisende som tilhører den første gruppen vil typisk ankomme stasjon j om morgenen, mens individer i gruppe to vil ankomme stasjon j om ettermiddagen. Fordelingen over ønsket ankomsttid til stasjon j framkommer dermed som en kombinasjon av fordelingene for de to subpopulasjonene.

Matematisk defineres en *finite mixture* som

$$f(y) = \sum_{i=1}^K \lambda_i f_i(y), \quad \lambda_i \geq 0 \quad \forall i, \quad \sum_i \lambda_i = 1, \quad (1)$$

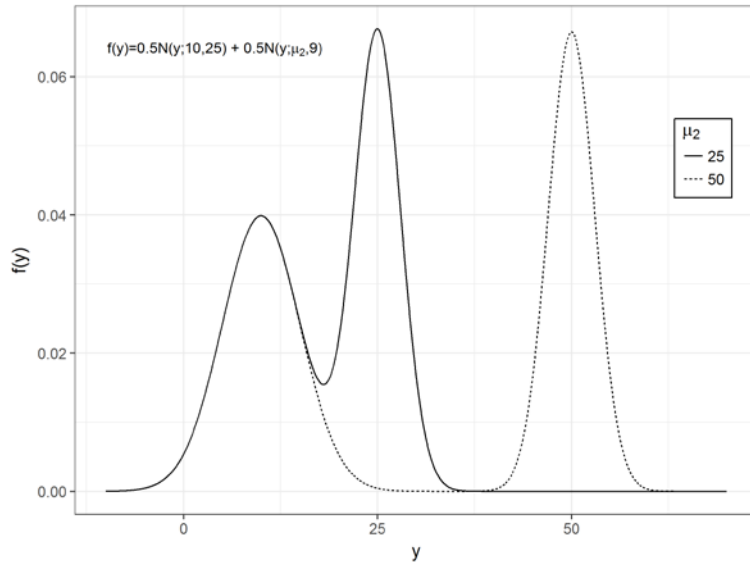
der f_1, f_2, \dots, f_K er et sett av sannsynlighetsfordelinger og $\lambda_1, \lambda_2, \dots, \lambda_K$ er et sett av vektorer. Hvis vi lar

$$f_i(y) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(y-\mu_i)^2}{2\sigma_i^2}}, \quad -\infty \leq y \leq \infty,$$

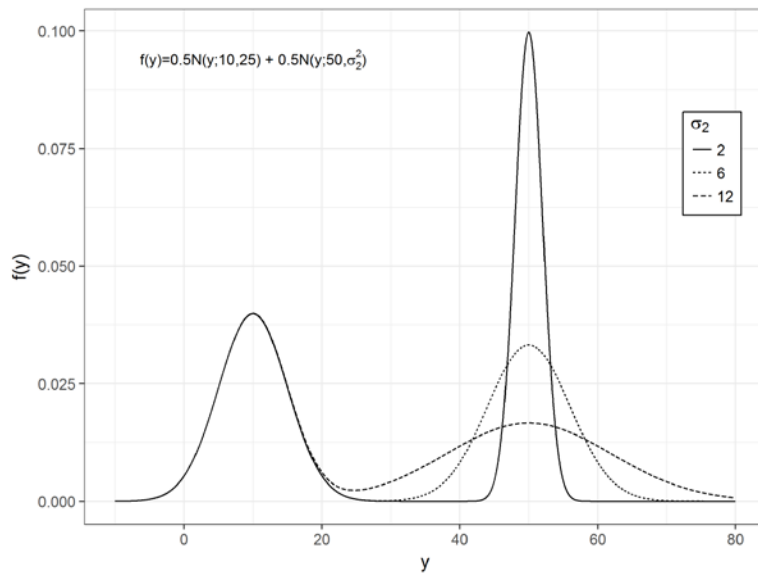
det vil si at f_i er en normalfordeling med forventning μ_i og varians σ_i^2 , får vi en *finite mixture of normals*. La $N(y; \mu, \sigma^2)$ betegne sannsynlighetstetthetsfunksjonen i (1). En mixture av to normalfordelinger kan da skrives som

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

Som for en vanlig normalfordeling er μ_1 og μ_2 lokasjonsparametere og σ_1 og σ_2 er skala-parametere. I Figur 5.1 og Figur 5.2 illustreres effekten av å variere henholdsvis μ_2 og σ_2 .



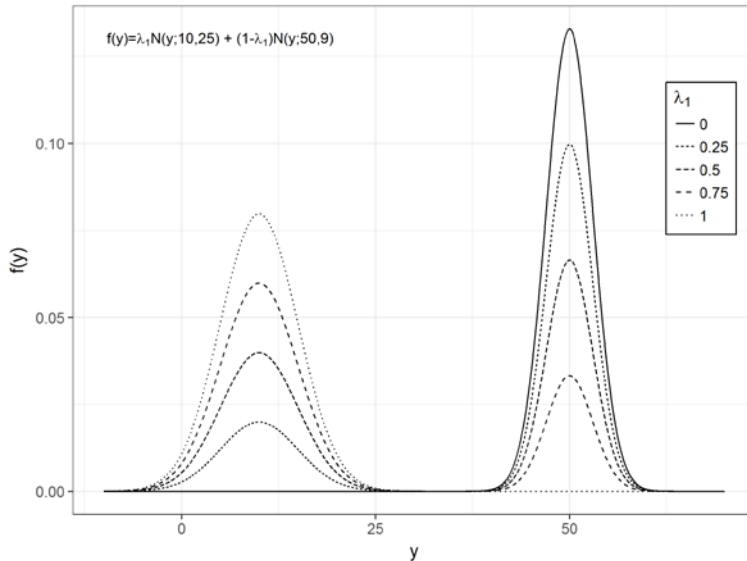
Figur 5.1: Illustrasjon av effekten av å variere μ_2 i en finite mixture of normals fordeling med to komponenter.



Figur 5.2: Illustrasjon av effekten av å variere σ_2 i en finite mixture of normals fordeling med to komponenter.

Vi ser at når μ_2 øker i verdi forskyves tyngdepunktet for den andre komponenten i fordelingen lengre mot høyre. Ved å øke eller minke σ_2 , øker eller minker spredningen om μ_2 .

I begge eksemplene over er $\lambda_1 = \lambda_2 = 0.5$. I Figur 5.3 illustrerer vi hva som skjer når λ_1 varierer mellom 0 og 1.



Figur 5.3: Illustrasjon av effekten av å variere λ_1 i en finite mixture of normals fordeling med to komponenter.

I de to ytterpunktene får vi en fordeling med enten bare komponent nummer to ($\lambda_1 = 0$) eller bare komponent nummer én ($\lambda_1 = 1$). I intervallet mellom 0 og 1, gir en høyere verdi av λ_1 en høyere vektning av komponent nummer én sammenlignet med komponent nummer to. Tolkningen av λ_1 og λ_2 er at de angir sannsynlighetene for at en observasjon tilhører de respektive underliggende subpopulasjonene.

Ved å variere antall komponenter K , samt lokasjons/skala-parametrene og vektene kan man få fram sannsynlighetstettheter med svært forskjellige egenskaper.

Trenklin modellen krever en diskret fordeling, mens mixture of normals fordelingen er definert for alle reelle tall. Dette kan enkelt løses ved å finne verdien av tetthetsfunksjonen for et sett av punkter (her: minutt i døgnet) og så re-skalere slik at summen blir én.

I utlysningsteksten står det at døgnfordelingen skal kunne karakteriseres ved parametre og inngangsvariabler slik at $f_{i,j}^H(t) := f_{i,j}^H(t; \boldsymbol{\theta}, \mathbf{x})$, der $\boldsymbol{\theta}$ er en vektor av parametre som innvirker på funksjonsformen gjennom \mathbf{x} , en vektor av variabler som karakteriserer stasjonsrelasjonen.

Modellen beskrevet over tar ikke høyde for at vi ønsker å la ulike egenskaper ved stasjonsrelasjonen styre formen på døgnfordelingen. En løsning på dette er å bruke en modelltype som kalles *mixtures of linear regressions*. Vi følger definisjonen i Benaglia mfl. (2009) og antar at vi har observasjoner $\mathbf{y}_1, \dots, \mathbf{y}_n$ med tilhørende sett av forklaringsvariabler $\mathbf{x}_1, \dots, \mathbf{x}_n$, med $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ for $i = 1, \dots, n$. La $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)^T$ og \mathbf{X} være $n \times p$ matrisen som består av vektorene med forklaringsvariablene. Anta at hver observasjon $(\mathbf{y}_i, \mathbf{x}_i)$ tilhører en av K klasser. Betinget på medlemskap i komponent k , er sammenhengen mellom \mathbf{y}_i og \mathbf{x}_i den vanlige lineære regresjonsmodellen

$$\mathbf{y}_i = \mathbf{x}_i^T \boldsymbol{\beta}_k + \epsilon_i,$$

der $\epsilon_i \sim N(0, \sigma_k^2)$ og $\boldsymbol{\beta}_k$ og σ_k^2 er henholdsvis den p -dimensjonale vektoren med regresjonskoeffisienter og feilvariansen for komponent nummer k . Den betingede fordelingen for $\mathbf{y}_i | \mathbf{x}_i$ er da

$$f(\mathbf{y}_i | \mathbf{x}_i) = \sum_{k=1}^K \lambda_k N(\mathbf{y}_i; \mathbf{x}_i^T \boldsymbol{\beta}_k, \sigma_k^2).$$

Her er forventningsverdien for komponent k gitt ved $\mathbf{x}_i^T \boldsymbol{\beta}_k$ og vi kan la egenskapene ved stasjonsrelasjonen inngå her. Det er også mulig å la vektene, også kalt *mixing proportions*, avhenge av forklaringsvariablene, det vil si å la $\lambda_k = \lambda_k(\mathbf{x}_i)$ og anta enten en parametrisk eller ikke-parametrisk form for sammenheng. Vi henviser til Benaglia mfl. (2009) for flere detaljer.

For å kunne benytte de to modellklassene beskrevet over til å estimere sannsynlighetsfordelingen $f_Y(\mathbf{y})$ til den tilfeldige variabelen Y , trenger man et tilfeldig utvalg $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$ der alle observasjonene antas å komme fra den samme underliggende sannsynlighetsfordelingen.

For å kunne estimere fordeling over ønsket ankomsttid med disse statistiske metodene trenger man observasjoner over ønsket ankomsttid for et tilfeldig utvalg av reisende. Da vi ikke har hatt tilgang til slike data direkte, har vi i stedet tilpasset data vi har hatt tilgjengelig. Vi har benyttet den nasjonale reisevaneundersøkelsen (RVU) og passasjertellinger fra NSB. De nødvendige tilpasningene som er gjort er beskrevet i kapittel 5.3.

I løpet av prosjektets gang har vi også vært innom andre tilnærminger til å modellere døgnfordelingene. Under beskriver vi disse kort, og forklarer hvorfor de ble forkastet.

Alternativ I

La $N_j(t)$ være antall passasjerer som går av på stasjon j ved tidspunkt t . Videre anta at $N_j(t)$ er Poissonfordelt med intensitetsfunksjon $\mu_j(t)$. Intensitetsfunksjonen kan modelleres som en lineær kombinasjon av ulike forklaringsvariabler og andre effekter. I dette tilfellet er det naturlig med en fleksibel døgn effekt og en romlig effekt av stasjonens lokasjon som kan ta høyde for at stasjoner som ligger nær hverandre ofte ligner mer enn stasjoner langt fra hverandre. En modell som dette tilhører klassen av latente Gaussiske modeller, og kan tilpasses data for eksempel med programvaren R-INLA (Rue mfl. 2009). Ved å la døgn effekten være en såkalt *syklisk random walk effekt* får man kontinuitet ved midnatt, noe som kan være viktig i etterspørselsmodellering.

Ved å bruke automatiske registreringer av antall avstigende passasjerer på ulike stasjoner ved ulike tidspunkt har man de nødvendige observasjonene for en slik modelltilnærming. Det er en fleksibel metode for å estimere sammenhengen mellom stasjonsegenskaper og reiseetterspørselens fordeling over døgnet. Problemet med metoden er at den ikke tar høyde for hvor de reisende kommer fra, samt at passasjertellinger mangler informasjon om reisehensikt. Vi gikk derfor ikke videre med denne modellen i etableringen av de relasjonsspesifikke døgnfordelingene.

Alternativ II

Hvis man tenker på ankomsttid T og reisehensikt H som to tilfeldige variabler, er sannsynligheten for ankomsttid gitt reisehensikt gitt ved

$$P(T|H) = \frac{P(T \cap H)}{P(H)} = \frac{P(H|T)P(T)}{P(H)}.$$

For en bestemt stasjonsrelasjon vil man derfor ha at

$$f_{i,j}(t|h) \propto f_{i,j}(h|t) \times f_{i,j}(t).$$

Det vil si at døgnfordelingen for ankomsttid gitt reisehensikt er proporsjonal med produktet av hensiktsfordelingen gitt reisetid og en ankomsttidsfordeling. Vi tenkte å estimere hensiktsfordelingen basert på RVU og ankomsttidsfordelingen basert på passasjetellinger fra NSB. Problemet med denne metoden var at det ikke er tilstrekkelig med observasjoner i RVU til å splitte opp hensiktsfordelingen for ulike relasjoner. Da hvilke stasjoner som er start- og endestasjoner, sammen med pendlerstrømmens retning, i stor grad styrer døgnfordelingen ved endestasjonen gav det ikke mening å bruke én hensiktsfordeling gitt tidspunkt for alle relasjoner og metoden ble derfor lagt til side.

5.3 Modelltilnærming og datagrunnlag

Vi skal modellere hvordan etterspørselen etter togreiser fordeler seg over døgnet for ulike reisehensikter og stasjonsrelasjoner. Hvilke data som er tilgjengelig styrer i stor grad hvilke modelleringstilnærminger som er mulige, og i dette prosjektet har vi hatt tilgang til to datakilder som sier noe om etterspørselsfordelingen over døgnet:

- Den nasjonale reisevaneundersøkelsen (RVU 2013–14)
- Passasjetellinger fra NSB

Vår konklusjon er at innen rammen av prosjektet, og med de data vi har hatt tilgjengelig, er tilnærmingen med mixture fordelinger den best egnede metoden.

I de neste avsnittene beskrives de to datakildene nærmere, og vi gjør rede for nødvendige tilpasninger som er gjort for å kunne benytte datasettene til å etablere døgnfordelinger.

5.3.1 RVU 2013–14

Den nasjonale reisevaneundersøkelsen (RVU) er en rik kilde til informasjon om reiser som foretas i Norge. Spørreundersøkelsen er beskrevet nærmere i Hjorthol mfl. (2014). I RVU for 2013–14 har i alt 54 925 respondenter svart på spørsmål om reisene de foretok på en bestemt registreringsdag, eller dagen før intervjuet fant sted.

Reisekjeder splittes opp etter formål, slik at reisekjeden

Hjemme → Barnehage → Jobb

gir de to reisene

Hjemme → Barnehage, formål: følge/omsorg

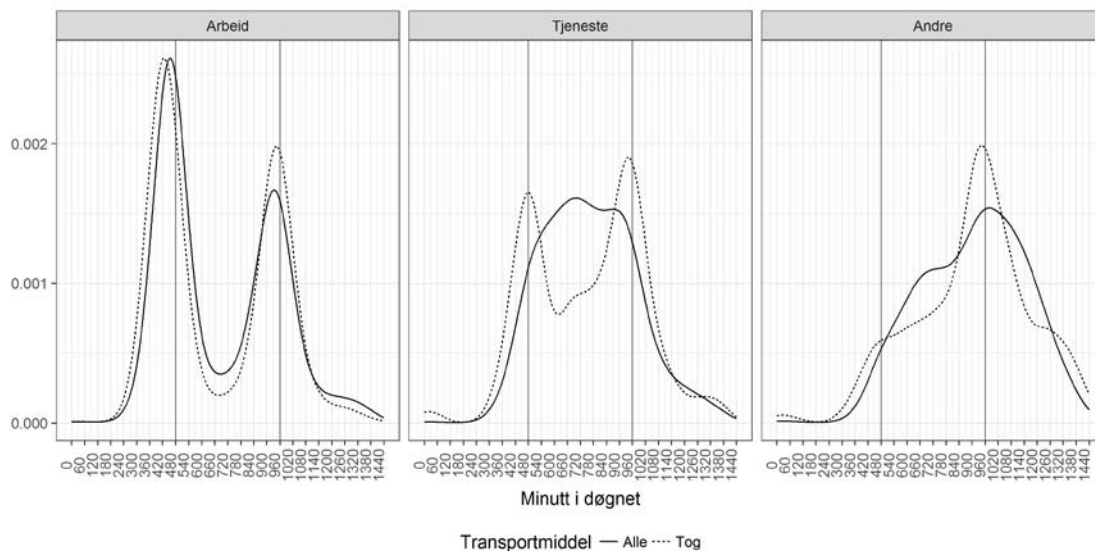
Barnehage → Jobb, formål: arbeid

Det er i alt 199 583 reiser som er registrert. Tilhørende hver reise er det, i tillegg til formålet med reisen, også angitt når reisen startet, hvor lenge reisen varte, hvor lang reisen var, hvilket transportmiddel som ble benyttet, samt en rekke andre variabler som for eksempel måned og ukedag. Reisen er stedfestet ved grunnkrets for start- og endepunkt.

Formålet med reisen registreres som en av åtte kategorier: arbeid, skole, tjeneste, handel/service, følge/omsorg, fritid, besøk eller annet. I dette prosjektet er arbeids- og skolereiser slått sammen til kategorien *arbeid*, mens reiser med formål handel/service, følge/omsorg, fritid, besøk eller annet inngår i kategorien *andre reiser*.

Da respondentene har oppgitt avreisetidspunkt for sine ulike reiser er det mulig å estimere sannsynlighetsfordelinger for avreisetidspunkt for reiser med ulike transportmidler og reisehensikter.

I Figur 5.4 sammenlignes fordelingene for reiser der tog var hovedtransportmiddel med fordelingene basert på alle reiser uansett transportmiddel⁷ (inkludert togreiser).



Figur 5.4: Sannsynlighetsfordelinger over avreisetidspunkt estimert på bakgrunn av reiser i den nasjonale reisevaneundersøkelsen 2013–14 for ulike reisehensikter. Fordelingene er estimert ikke-parametriske med binned kernel density estimation (med funksjonen `kde` fra R-pakken `KernSmooth` (Wand, 2015)) med båndbredde på 60 minutter. Figurene sammenligner fordelingene basert på alle reiser, uansett transportmiddel (beltrukken linje), med reiser der tog er hovedtransportmiddel (stiplet linje). Antall observasjoner for alle transportmidler er: arbeidsreiser ($N = 39\,395$), tjenestereiser ($N = 6\,279$) og andre reiser ($N = 99\,940$), og for togreiser gjelder: arbeidsreiser ($N = 1\,423$), tjenestereiser ($N = 155$) og andre reiser ($N = 615$). Merk at fordelingene gjelder for virkedager, det vil si at reiser på lørdager, søndager, eller i juli er ekskludert. Tidspunktene 08:00 (480) og 16:00 (960) er markert med vertikale linjer.

Figur 5.4 viser at hovedvekten av arbeidsreiser har avreisetidspunkt omkring kl. 7:30 eller kl. 16, altså ved normalarbeidsdagens start eller slutt. At det er færre arbeidsreiser som starter om ettermiddagen skyldes at reiser til butikk, barnehage, treningssenter eller andre fritidsaktiviteter havner i kategorien andre reiser. Legg merke til at fordelingen estimert basert på kun togreiser, har en morgentopp som er litt tidligere, og en ettermiddagstopp som er litt seinere, enn fordelingen basert på reiser med alle transportmidler.

Fordelingen for tjenestereiser med tog har to topper, en om morgenen og en om ettermiddagen, mens når alle transportmidler inkluderes får fordelingen en annen form, med hovedvekten av tjenestereiser innen normal arbeidstid. Fordelingen for tjenestereiser med tog baserer seg på kun 155 observasjoner, så usikkerheten er stor. Men det er kanskje rimelig at reiser med tog i tjenesteøyemed primært er reiser som foretas om morgenen eller ettermiddagen, mens kortere reiser til/fra møter, eller oppdrag for håndverkere, i større grad spres utover arbeidsdagen, og at det da gjerne benyttes andre transportmidler enn tog. I kategorien andre reiser inngår reiser med formål handel/service, følge/omsorg, fritid, besøk eller annet. Det er en gradvis økning i denne type reiser utover dagen, og toppene kommer kl. 16 for begge fordelingene. Den største andelen andre reiser starter når normalarbeidsdagen slutter, dette er naturlig da kun virkedagsturer for ukedagene mandag–

⁷ Transportmidler som inngår i RVU er: til fots, sykkel, el-sykkel, t-bane, trikk/bybane, buss/rutebil/ekspresbuss i rute, tog, bil (som fører), bil (som passasjer), drosje/taxi, rutefly, rutebåt, moped, motorsykkel, turbuss/chartret buss, annen båt/fritidsbåt/småbåt, ferge (som bilfører/passasjer), ferge (som fotgjenger/busspassasjer), charterfly, traktor, snøscooter eller annet.

fredag er inkludert, men det er også en god del reiser som starter tidligere på dagen og dette forklarer den tyngre halen mot venstre.

Fordi fordelingene for togreiser har andre egenskaper enn fordelingene basert på alle reiser, og fordi vi skal knytte informasjon om togstasjonsrelasjonen til observasjonene, velger vi å kun benytte togreiser i den videre estimeringen av døgnfordelinger. Da forkaster vi en stor andel reiser og sitter igjen med relativt lite data. Samtidig skiller tog seg fra andre transportmidler på flere områder, især er det mindre fleksibilitet med tanke på avreisetidspunkt enn for eksempel for bil eller sykkel. I vedlegg 2 vises en figur som sammenligner døgnfordelinger over avreisetidspunkt for ulike transportmidler.

Tilpassing av RVU til prosjektets formål

For å kunne bruke svarene i RVU til å estimere døgnfordelinger må det gjøres noen tilpasninger. Disse beskrives i de to neste avsnittene.

Virkedøgn og inndelingen av døgnet

Virkedøgn defineres som ukedager (mandag – fredag) i alle måneder med unntak av feriemåneden juli.

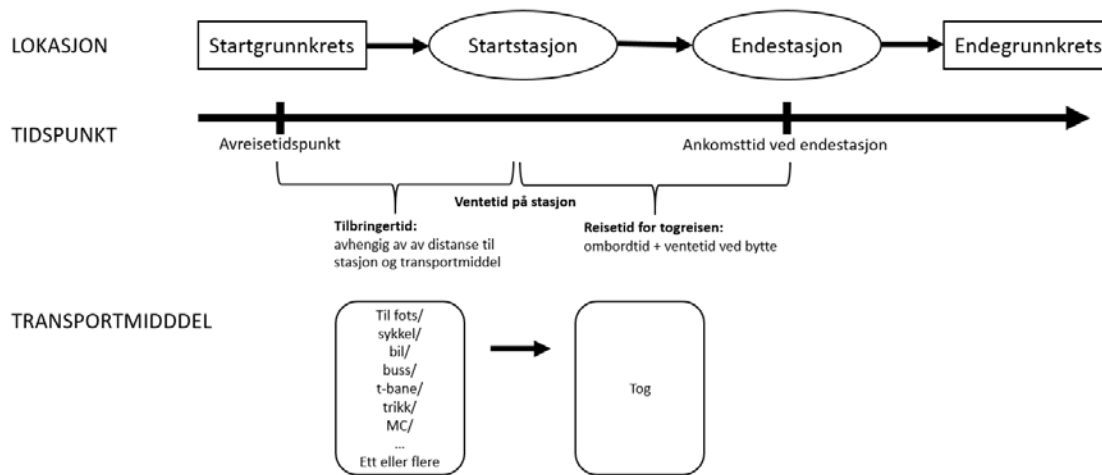
I RVU regnes et reisedøgn fra kl. 04:00 til 03:59, mens i Trenklin modellen starter døgnet ved midnatt, det vil si at døgnet starter kl. 00:00 og slutter kl. 23:59.

Døgnfordelingene som estimeres skal være for virkedøgn og dermed skal lørdag og søndag utelates. For å unngå at reiser mellom kl. 00:00 og 03:59 fredager regnes som virkedøgn og at reiser mellom kl. 00:00 og 03:59 søndager regnes som restdøgn, beregnes en justert ukedagsvariabel. I tillegg er det ukedag ved ankomsttidspunktet som er relevant for døgnfordelingene, så for reiser som strekker seg inn i et nytt døgn korrigeres det både for ankomsttid og ukedag for ankomst. For eksempel vil en reise som starter fredag kl. 22:30 og varer i 2 timer, få ankomstdag lørdag og ankomsttid kl. 00:30, selv om dette i RVU sammenheng regnes som fredag.

Fra avreisetidspunkt til ankomsttid ved endestasjon

I RVU er en reise stedfestet med grunnkrets for reisens start- og endepunkt, og tidfestet med tidspunkt for avreise fra startstedet. I tillegg har respondenten angitt hvor lang tid reisen tok. For en typisk arbeidsreise betyr det at vi vet grunnkrets for bo- og arbeidssted, tidspunkt for når respondenten startet reisen, samt når vedkommende var framme ved sluttstedet. Dette er ikke direkte den informasjonen vi trenger for å estimere døgnfordelingene. Da trenger vi å vite hvilke stasjoner som var start- og endestasjon for togreisen, og når respondenten ankom endestasjonen.

Figur 5.5 viser en skjematisk oversikt over en reises forløp og den informasjonen som ligger i RVU.



$$\text{Ankomsttid ved endestasjon} = \text{Avreisetidspunkt} + \text{Tilbringertid} + \text{Ventetid} + \text{Reisetid}$$

Figur 5.5: Oversiktsfigur over stedsfesting og tidfesting av reiser i RVU og den implisitte informasjonen om ankomsttid ved endestasjon som blir beregnet.

Figuren over viser at ankomsttid ved endestasjonen kan beregnes ved å legge tilbringertid, ventetid og reisetid til avreisetidspunktet. Fra NSBs nasjonale tilbudsmatriser vet vi reisetid, og ventetid ved bytte, for togreiser mellom alle stasjonspar. For å beregne ankomsttid ved endestasjon er det nødvendig med noen antakelser:

1. Anta at de to togstasjonene som er benyttet er stasjonene med kortest avstand fra start- og endegrunnkrets (tyngdepunktet i grunnkretsen).
2. Distansen (i vegnettet) mellom grunnkretstyngdepunkt og startstasjon, samt kjeden av transportmidler benyttet før togreisen kan brukes til å beregne tilbringertid.

Koblingen mellom grunnkretser og togstasjoner er beskrevet i kapittel 3.2.

I RVU har respondenten oppgitt alle transportmidler som ble benyttet på reisen i kronologisk rekkefølge, dette gir en transportmiddelkjede og to eksempler fra datasettet er:

Til fots → Sykkel → Tog

Til fots → Bil/buss → Tog → Bil/buss

Vi benytter informasjonen om transportmidlene som er brukt før togreisen startet til å beregne tid fra startsted til togstasjon.

Metodikken for å beregne tilbringertid beskrives under. Før denne beregningen ble 28 av 2560 reiser fjernet med begrunnelse som følger:

- Seks reiser der ingen av transportmidlene i kjeden er tog. Det er mulig at hovedtransportmiddel er korrigert til tog og at det er derfor reisene er blitt med i datasettet, dette ser vi bort fra og ekskluderer disse reisene.
- Åtte reiser der et av transportmidlene benyttet er fly. Dette vanskeliggjør å anslå tilbringertid og reisetid da togstasjonene med kortest avstand til grunnkretsene trolig ikke er de som er benyttet.
- Tretten reiser der et av transportmidlene benyttet er ferge. På samme måte som for fly er det vanskelig å slå fast både reisetid ombord på tog og tilbringertid i dette tilfellet.
- To reiser der det i alt er benyttet fire transportmidler før togreisen. Ikke nødvendigvis stasjonen som ligger nærmest som er benyttet som startstasjon.

La d være avstanden mellom grunnkretstygndepunktet og nærmeste togstasjon, og la v_{tr} betegne hastigheten til et bestemt transportmiddel. Vi benytter følgende verdier:

Transportmiddel	Til fots	Sykkel	Bil/buss/MC	T-bane/trikk	Annet
Hastighet v_{tr}	5 km/t	15 km/t	40 km/t	20 km/t	15 km/t

Antall minutter på et transportmiddel beregnes med formelen

$$tid_{tr} = \frac{d \text{ km}}{v_{tr} \text{ km/t}} \times \frac{60 \text{ min}}{t} \quad (2)$$

Vi benytter distanse, hastighet og kjeden av transportmidler til å beregne tilbringertiden som følger:

1. Tog første transportmiddel:
 - a. Hvis $d \leq 1.7$ km, anta at respondent går denne distansen.
 - b. Hvis $d > 1.7$ km, anta at respondenten har kjørt bil/buss/etc. Her antar vi dermed at gangtid aldri er mer enn 20 minutter.
2. Tog andre transportmiddel:
 - a. Beregn tilbringertid med formelen i (2).
 - b. Hvis første transportmiddel er til fots og $d > 1.7$ km, anta i stedet at bil/buss/MC eller sykkel er benyttet og sett hastigheten til 30 km/t som et gjennomsnittsanslag.
3. Tog tredje transportmiddel:
 - a. Beregn tilbringertid med formelen i (2) ved å anta hastigheten til det raskeste av de to første transportmidlene og legg til fem minutter i byttetid.
4. Tog fjerde transportmiddel:
 - a. Beregn tilbringertid med formelen i (2) ved å anta hastigheten til det raskeste av de tre første transportmidlene og legg til ti minutter i byttetid.

Det er mange feilkilder knyttet til denne beregningen: Antakelsen om at respondentene starter i grunnkretstygndepunktet, feilregistreringer i RVU og alle de implisitte antakelsene som er beskrevet i de fire trinnene over, som i mange tilfeller kan være direkte feil. I tabellen under vises en oversikt over hvor mange reiser som havner i de ulike kategoriene.

Kategori	1a	1b	2a	2b	3	4	Uten distanse	Totalt
Antall reiser	462	133	1075	157	502	48	155	2532

Flest reiser er i kategori 2a: ett transportmiddel benyttet før tog og ingen antakelser basert på avstand. Hvis vi undersøker transportmiddelfordelingen blant disse 1075 reisene har 67% (722) gått, 25% (273) reist med bil, buss eller MC, 4% syklet (47) og 3% (33) tatt trikk eller t-bane.

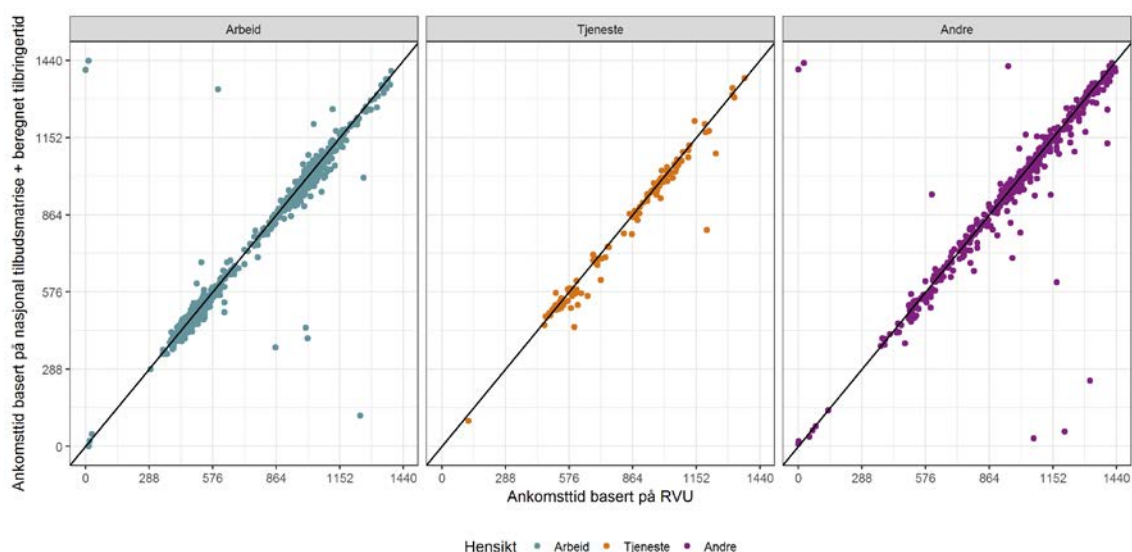
Vår metode for å beregne tilbringertid passer nok best for reiser der respondenten har gått til nærmeste togstasjon, og de utgjør en stor undergruppe i kategori 2a, men disse reisene utgjør kun 29% av alle reisene i datasettet og vi har ingen måte å sjekke antakelsene for andre typer reiser da vi ikke har informasjon om byttepunktene.

Usikkerheten i tilbringertidsestimatet er stor og vi vil derfor sammenligne resultater når vi benytter fire ulike variabler for ankomstid:

Tabell 5.1: Oversikt over ankomsttidsvariabler. Ombordtid og ventetid ved bytte kommer fra de nasjonale tilbudsmatrisene. Merk at avreisetidspunkt kommer fra RVU for alle variablene.

Variabelnavn	Beregning
Ankomsttid RVU	Avreisetidspunkt + korrigeret reisetid fra RVU (variabel: <i>korrtid</i>)
Ankomsttid NSB	Avreisetidspunkt + ombordtid + ventetid ved bytte
Ankomsttid NSB + 15 min	Avreisetidspunkt + ombordtid + ventetid ved bytte + tilbringertid på 15 minutter for alle reiser
Ankomsttid NSB + beregnet	Avreisetidspunkt + ombordtid + ventetid ved bytte + beregnet tilbringertid

I Figur 5.6 sammenlignes ankomsttiden basert på RVU med ankomsttiden basert på de nasjonale tilbudsmatrisene med beregnet tilbringertid.



Figur 5.6: Sammenligning av ankomsttider der reisetiden i RVU er benyttet (x-aksen) og reisetiden fra NSB med beregnet tilbringertid (y-aksen).

Figuren over viser at det for noen reiser er ganske stor forskjell mellom RVU ankomsttiden og den beregnede ankomsttiden basert på ombordtid og tilbringertid, men i all hovedsak er det godt samsvar. Ankomsttiden basert på ombordtid er nærmere det tidspunktet vi ønsker enn den basert på reisetiden i RVU, fordi variabelen fra RVU også inkluderer reisetiden mellom endestasjon og destinasjon, men det er usikkerhet knyttet til beregningen av tilbringertid og tilordningen av togstasjoner. Vi velger å benytte alle de fire variablene i Tabell 5.1 for å se hvilke utslag det gir i parameterestimaterne. I vedlegg 3 finnes en sammenligning mellom de andre ankomsttidsvariablene.

5.3.2 Passasjertellinger – telledata

Vi har mottatt passasjertellinger for en rekke av NSB sine toglinjer. For å kunne benytte disse dataene måtte de transformeres til OD-format. Dette er beskrevet i kapittel 3.3.

5.3.3 Data som beskriver relasjonen

Vi benytter to sett av inndata om stasjonsrelasjonene: relativ pendlerindeks og reisetid.

Relativ pendlerindeks

For stasjonsrelasjon (i, j) defineres den relative pendlerindeksen ρ_{ij} som forholdet mellom pendlerstrømmen fra i til j og pendlerstrømmen fra j til i . Pendlerstrømmen er representert ved pendlerindeksen beskrevet i avsnitt 3.1.5, og er en omregning av SSBs pendlerstatistikk på grunnkrets nivå til å gjelde stasjonsrelasjoner. Vi har benyttet pendlerindeksen for alle arbeidskategorier.

Den relative pendlerindeksen er gitt ved uttrykket

$$\rho_{ij} = \frac{\text{pendlerindeks}_{ij}}{\text{pendlerindeks}_{ji}}$$

Hvis det er like mange som bor ved stasjon i og pendler til stasjon j som omvendt, vil $\rho_{ij} = 1$. Hvis det derimot er flere som pendler fra i til j enn motsatt retning vil $\rho_{ij} > 1$, og tilsvarende, hvis pendlerstrømmen i hovedsak går fra j til i vil $\rho_{ij} < 1$. Merk også at $\rho_{ji} = 1/\rho_{ij}$.

Tabell 5.2 viser den relative pendlerindeksen for utvalgte togstasjonsrelasjoner.

Tabell 5.2: Noen eksempler på verdien av den relative pendlerindeksen for utvalgte stasjonspar.

Stasjon i	Stasjon j	Relativ pendlerindeks ρ_{ij}
Asker	Oslo S	3,764
Oslo S	Asker	0,266
Halden	Oslo S	10,356
Oslo S	Halden	0,097
Larvik	Sandefjord	0,924
Sandefjord	Larvik	1,083

Relasjonen Asker – Oslo S har en relativ pendlerindeks på 3,764, noe som indikerer at pendlerstrømmen er mer enn tre ganger så stor fra Asker til Oslo S enn i motsatt retning. For relasjonen Halden – Oslo S er den relative pendlerindeksen over ti, altså går pendlerstrømmen i all hovedsak i retning Oslo. Mellom Larvik og Sandefjord er derimot verdien av ρ_{ij} mye nærmere 1, noe som betyr at det er omtrent like mange som pendler fra Larvik til Sandefjord som fra Sandefjord til Larvik.

Den relative pendlerindeksen sier ingenting om hvor mange som pendler mellom togstasjonene og er en variabel som kun skal benyttes til å forklare om hovedtyngden av arbeidsreisene har ankomst til endestasjonen om morgenen eller ettermiddagen.

Reisetid

Reisetiden mellom to stasjoner vil virke inn på hvilket tidspunkt etterspørselen etter togreiser er størst. Eksempelvis vil arbeidsreiser fra arbeid til bosted få ankomsttid som styres av reisetiden. To reisende som begge starter ved Oslo S kl. 16:30, men der den ene skal til Lillestrøm og den andre til Eidsvoll, vil få ulik ankomsttid fordi det tar lengre tid å reise til Eidsvoll enn til Lillestrøm. Om morgenen er dette mindre relevant da det er avreisetidspunktet som må tilpasses for å ankomme arbeid til ønsket tid. Tilsvarende vil gjelde også for tjenestereiser og andre reiser: Ankomsttiden til en aktivitet (forretningsmøte, kino, besøk, etc.) vil være styrt av aktivitetens oppstarttid og det er *avreisetidspunktet* som må legges tidligere hvis reisetiden er lang. Ankomsttidspunktet hjem igjen vil derimot forskyves til seinere på dagen for lange reiser og reisetid ansees derfor som en viktig variabel for å beskrive dognfordelingene.

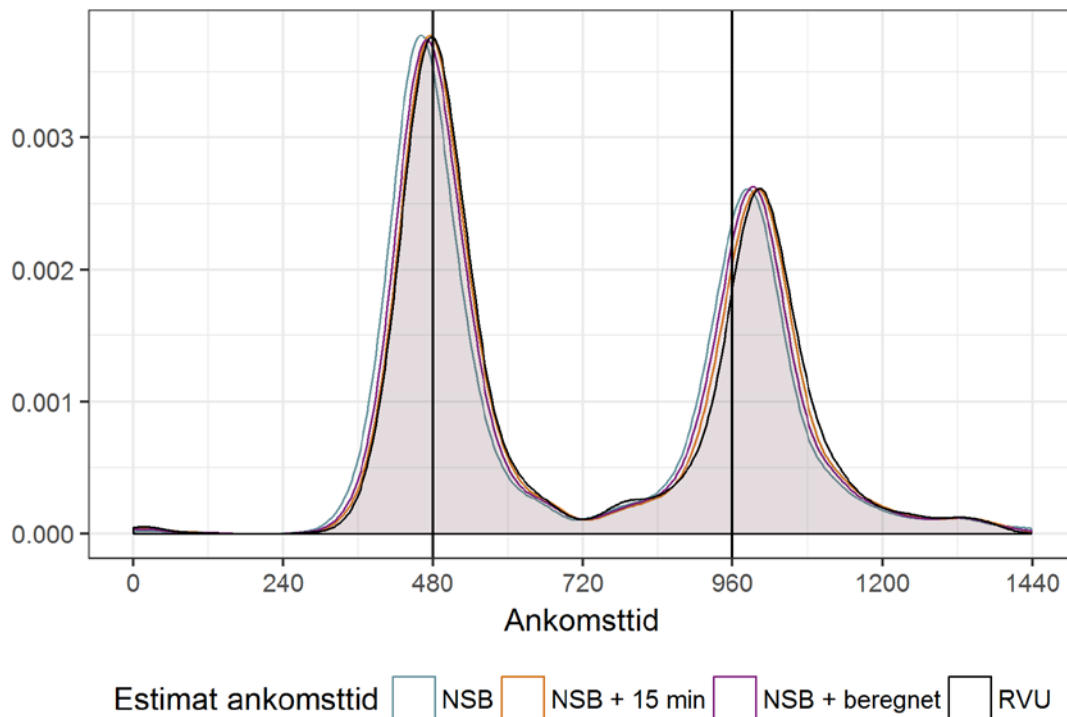
5.4 Fordelinger basert på RVU

Vi har tilpasset *mixture fordelinger* som beskrevet i kapittel 5.2 til ankomsttidene basert på RVU. Parameterne er estimert med EM-algoritmen⁸ og til dette har vi benyttet statistikkprogramvaren R (R Core Team, 2016) med pakken `mixtools`⁹ (Benaglia mfl. 2009).

5.4.1 Arbeidsreiser

Det er 1415 arbeidsreiser i datasettet. På grunn av manglende stedsfesting finnes ikke ankomsttider basert på tilbudsmatrisene for 204 av disse reisene.

For å estimere døgnfordelingene benytter vi fire ulike ankomsttidsvariabler: én der både avreisetidspunkt og reisetid kommer fra RVU, og tre andre der avreisetidspunktet kommer fra RVU, men der reisetiden kommer fra NSB sine nasjonale tilbudsmatriser og tilbringertiden er henholdsvis null for alle reiser, 15 minutter for alle reiser eller beregnet på bakgrunn av distanse og transportmiddelbruk (se avsnitt 5.3.1). Figur 5.7 viser døgnfordelingen over ankomsttid for de fire ulike ankomsttidsvariablene.



Figur 5.7: Døgnfordelinger over ankomsttid for arbeidsreiser. Fordelingene er estimert ikke-parametriske (kernel density estimation, med $bw = 30$). Figuren viser fordelinger for de fire ulike tidsvariablene. For RVU ankomsttiden er $N = 1415$, for de resterende variablene er $N = 1211$. Kl. 08:00 og kl. 16:00 er markert med vertikale linjer.

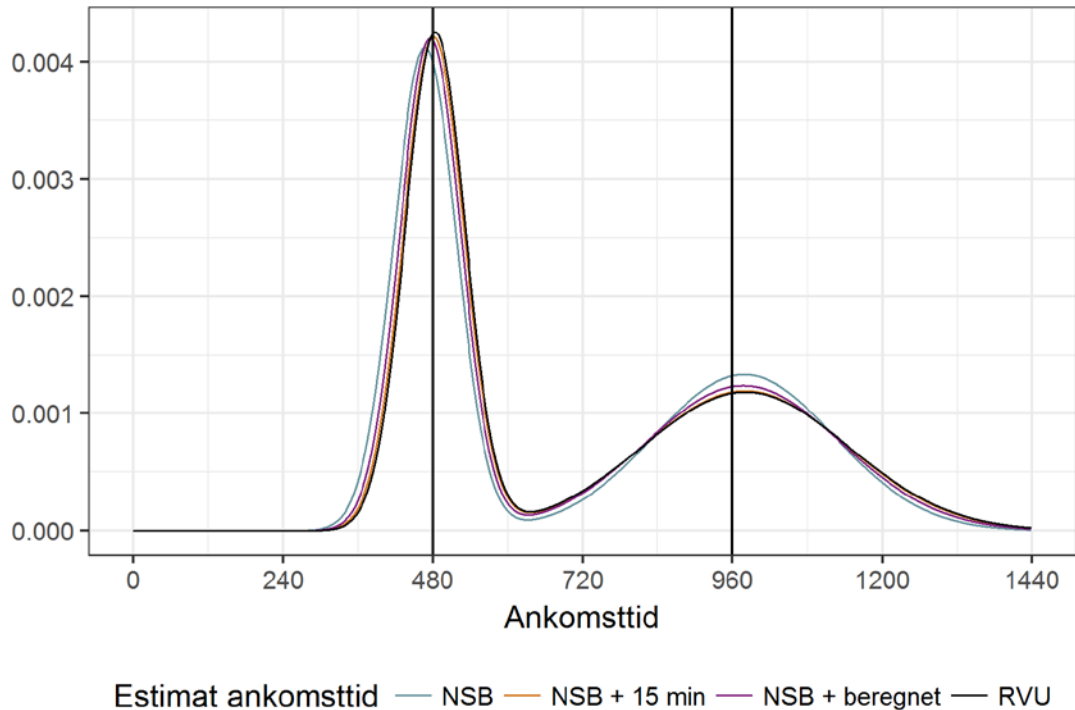
Fordelingene i figuren over er estimert ikke-parametriske og vi ser to toppe: en rett før kl. 8 og en litt etter kl. 16. Dette er som forventet. Det er en liten forskyvning mellom fordelingene basert på de ulike variablene. Når ankomsttiden kun er beregnet fra reisetiden ombord på toget kommer toppene tidligere enn for de andre ankomsttidene, dette er

⁸ Expectation-maximization algoritmen er en iterativ metode for maximum likelihood estimering, se for eksempel Dempster mfl. (1977).

⁹ Helt konkret er funksjonene `normalmixEM()` og `regmixEM()` benyttet.

naturlig da det ikke er tatt hensyn til tilbringertid. På samme måte kommer toppene seinere når reisetiden angitt i RVU er benyttet, da denne tidsvariabelen er tid til destinasjon og ikke endestasjon.

Vi ønsker parametriske modeller for døgnfordelingene og tilpasser derfor mixtures of normals fordeling med to komponenter til data. Ser vi på Figur 5.7 virker dette som et naturlig valg: Vi har en bi-modal fordeling og begge komponentene er relativt symmetriske. I Figur 5.8 viser vi hvordan mixture fordelingene for arbeidsreisene ser ut og i Tabell 5.3 viser vi parameterestimaten.



Figur 5.8: Døgnfordelinger over ankomsttid for arbeidsreiser. Fordelingene er spesifisert som mixtures med to normalfordelte komponenter og estimert med EM-algoritmen. Figuren viser fordelinger for de fire ulike tidsvariablene. For RVU ankomsttiden er $N = 1415$, for de resterende variablene er $N = 1211$. Kl. 08:00 og kl. 16:00 er markert med vertikale linjer.

Vi ser at komponenten om morgenen er relativt lik den i Figur 5.7, men at ettermiddagskomponenten har større spredning. Dette skyldes nok at arbeidsreiser midt på dagen og om kvelden fanges opp av denne komponenten og at dette øker variansen. Fordelingene varierer ikke mye mellom de ulike ankomsttidene. Dette kommer også tydelig fram av parameterestimaten i Tabell 5.3.

Tabell 5.3: Parameterestimaten for døgnfordelingene i Figur 5.8 (arbeidsreiser).

Parameter	RVU	NSB	NSB + 15 min	NSB + beregnet
λ_1	0,51	0,52	0,51	0,51
λ_2	0,49	0,48	0,49	0,49
μ_1	483 (08:03)	467 (07:47)	481 (08:01)	475 (07:55)
μ_2	981 (16:21)	979 (16:19)	980 (16:20)	978 (16:18)
σ_1	48	50	49	49
σ_2	165	144	164	157

Vi ser at vektingen mellom morgen- og ettermiddagskomponenten er relativt lik, med $\lambda_1 = 0,51$ og med $\lambda_2 = 0,49$. Da reiser fra jobb og til butikk/trening/hente i barnehagen vil falle i kategorien *andre reiser*, er det naturlig at vekten for morgenkomponenten er litt større enn ettermiddagskomponenten. Med en annen definisjon av reisehensikter vil man kunne forvente mer symmetri.

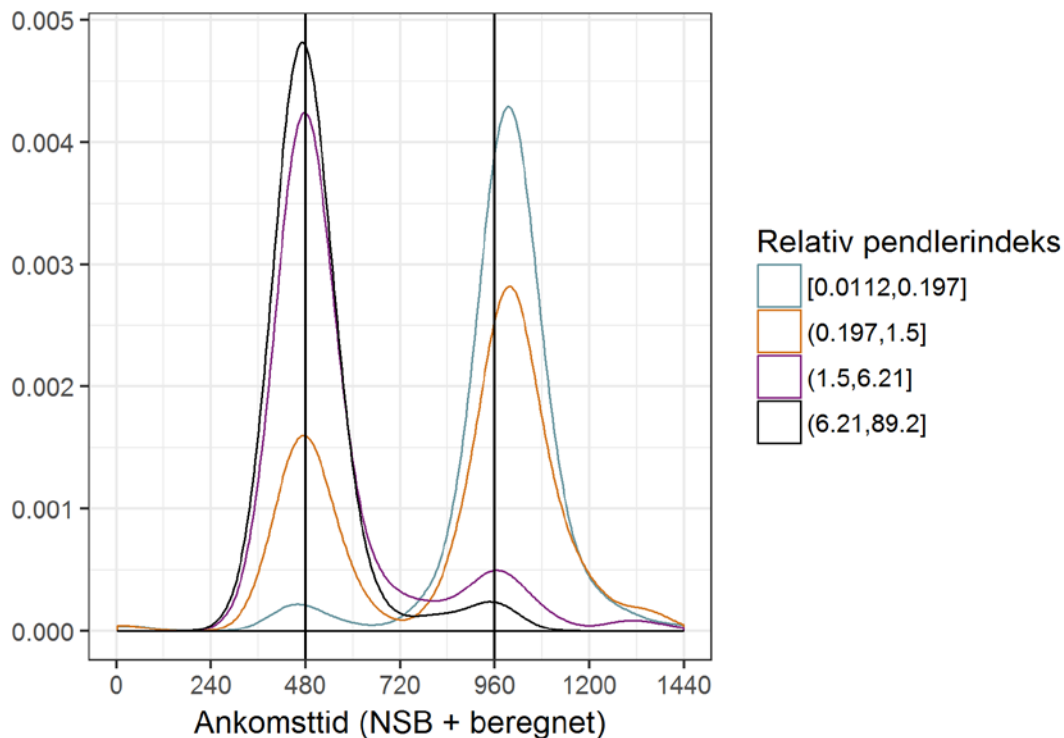
Fordelingene estimert her tar ikke hensyn til om folk reiser til eller fra arbeid, eller egenskaper ved relasjonen. I neste avsnitt tar vi inn forklaringsvariabler for pendlerstrøm og reisetid for å finne relasjonsspesifikke fordelinger.

Fordelinger med relasjonsdata

Relativ pendlerindeks

For stedfestede togreiser i RVU har vi antatt¹⁰ hvilke stasjonsrelasjoner som er benyttet, og kan derfor knytte den relative pendlerindeksen ρ_{ij} som beskrevet i avsnitt 5.3.3 til de observerte ankomsttidene for å forklare formen på døgnfordelingene.

Hvis vi deler opp datasettet i fire grupper etter verdien av ρ_{ij} og finner døgnfordelinger i disse gruppene ved hjelp av kernel density estimering, får vi fordelingene i Figur 5.9.



Figur 5.9: Døgnfordelinger for arbeidsreiser der datasettet er splittet opp i fire basert på verdien av den relative pendlerindeksen.

For høye verdier av den relative pendlerindeksen er all vekt på morgenkomponenten. Som beskrevet tidligere betyr høye verdier av ρ_{ij} at det er flere arbeidsreiser fra stasjon i til j enn motsatt. Relasjonen (Halden, Oslo S) har en indeksverdi på omtrent ti og havner i den høyeste kategorien her (svart kurve). Fordelingen stemmer med intuisjonen om at etterspørselen etter arbeidsreiser fra Halden til Oslo S er størst om morgenen og at ønsket ankomsttid ved Oslo S er rundt kl. 8. Indeksverdien for relasjonen (Oslo S, Halden) er

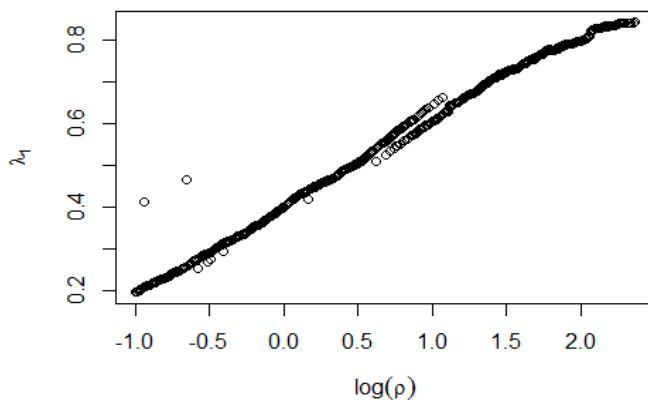
¹⁰ Omkodingen fra grunnkrets til stasjon, se avsnitt 3.2.

omtrent 1/10 og denne døgnfordelingen (grønn kurve) har hovedvekt om ettermiddagen. Dette stemmer godt overens med at alle som pendler til Oslo S fra Halden skal tilbake til Halden etter arbeidstid.

I en mixture fordeling er det parameterne λ_1 og λ_2 som styrer vektningen mellom de to komponentene (jf. Figur 5.3) og vi ønsker derfor å finne en empirisk sammenheng mellom vektene og den relative pendlerindeksen. Det finnes metoder for å estimere vektene som en funksjon av forklaringsvariabler, men i pakken mixtools gjøres dette ikke-parametrisk og da vi ønsker å finne en funksjonssammenheng har vi valgt vår egen tilnærming:

Datasettet sorteres i stigende rekkefølge basert på verdien av ρ_{ij} . Deretter estimeres parameterverdiene i en løkke som i steg én benytter observasjon 1 til b , i steg to benytter observasjon 2 til $b + 1$, osv. helt til estimeringsdatasettet består av observasjon $N - b$ til N . I tillegg til å estimere mixture parameterne finner vi også gjennomsnitt og kvartiler av alle ρ_{ij} verdiene i estimeringsdatasettet.

Når vi plottes λ_1 mot logaritmen av den relative pendlerindeksen viser det seg å være en lineær sammenheng¹¹ (se Figur 5.10).



Figur 5.10: Sammenhengen mellom den relative pendlerindeksen og vekten for morgenkomponenten λ_1 .

Vi setter derfor opp følgende regresjonsligning

$$\lambda_1 = \gamma_0 + \gamma_1 \ln \rho,$$

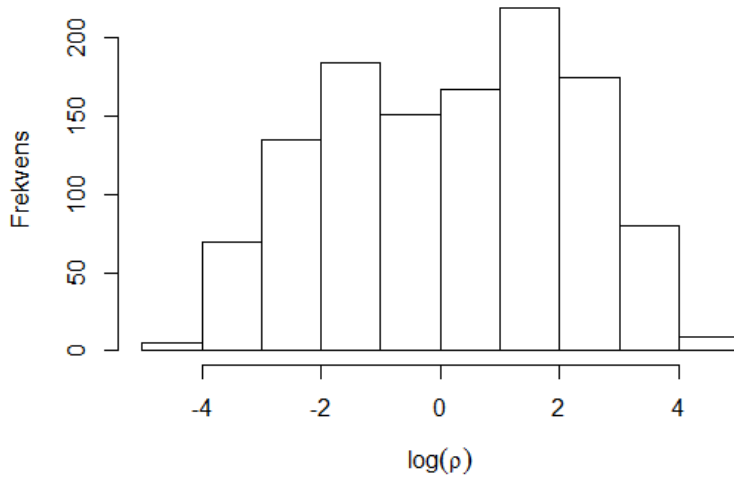
og estimerer γ_0 og γ_1 fra datasettet som består av par av mixture vekt for morgenkomponenten (λ_1) og relativ pendlerindeks (ρ) for de $N - b + 1$ estimeringsdatasettene.

Vi får at $\gamma_0 = 0,4$ og $\gamma_1 = 0,2$. Disse verdiene er stabile uavhengig av ankomststidsvariabel. Vi velger derfor å benytte variabelen med beregnet tilbringertid.

Tolkningen av γ_0 og γ_1 , er at γ_0 er vekten for morgenkomponenten når den relative pendlerindeksen er 1, altså når like mange pendler i begge retninger. Man kunne tenke seg at denne verdien burde være 0,5, for å ha symmetri. I stedet får vi empirisk 0,4. At vi ikke har symmetri skyldes trolig en kombinasjon av effekten av at det er flest arbeidsreiser om morgenen, da hjemreiser også kan havne i kategorien andre reiser (henting, handling, trening etc.), og den geografiske spredningen av respondentene i RVU.

¹¹ Her har vi brukt gjennomsnittsverdien av den relative pendlerindeksen i estimeringsdatasettet. Vi får tilsvarende resultat med medianen.

Figur 5.11 viser et histogram over \ln relativ pendlerindeks og det er tydelig at det er en overvekt av reiser med relativ pendlerindeks som er større enn én. Satt på spissen observerer vi arbeidsreisen fra Halden til Oslo S, men ikke hjem igjen. I tillegg er det en asymmetri rundt 0. Begge disse faktorene bidrar til asymmetrien i den empiriske sammenhengen.



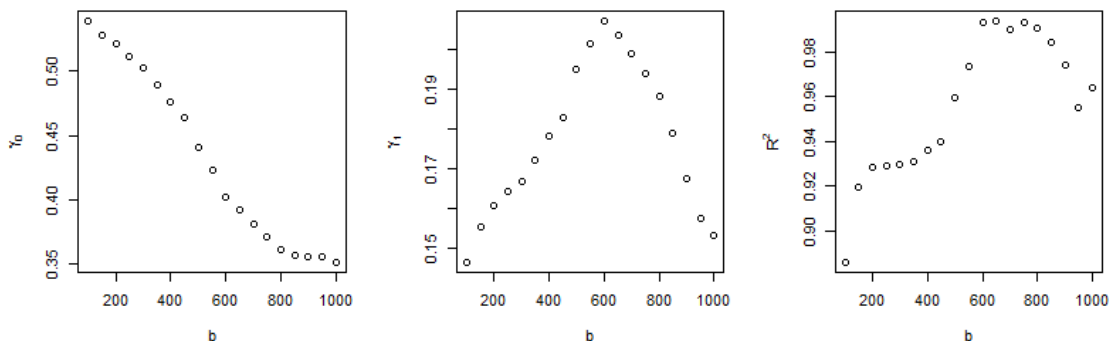
Figur 5.11: Histogram over den relative pendlerindeksen for arbeidsreiser i RVU 2013-14.

Tolkningen av γ_1 er at når den relative pendlerindeksen er større enn 1 vil vekten av morgenkomponenten øke, mens den vil minke når ρ er mindre enn 1.

Ved veldig små pendlerindekser er det en fare for at λ_1 blir negativ. Dette skjer hvis $\rho < e^{-\gamma_0/\gamma_1} = 0.135$. På samme måte risikerer vi en vekt som er større enn 1 ved høye verdier av ρ . Dette skjer hvis $\rho > e^{1-\gamma_0/\gamma_1} = 20,08$. Vi definerer derfor mixture vektene ved

$$\lambda_1 = \min\{1, \max\{0, \gamma_0 + \gamma_1 \ln \rho\}\}, \quad \lambda_2 = 1 - \lambda_1.$$

I beregningene over brukte vi $b = 600$. For å undersøke hvor følsomme parameterne er for størrelsen på estimeringsdatasettet, varierer vi b fra 50 til 1000. Resultatene av denne sensitivitetsanalysen er vist i Figur 5.12.



Figur 5.12: Sensitivitetsanalyse for størrelsen på estimeringsdatasettet.

Parameterverdiene varierer i andre desimal med størrelsen på estimeringsdatasettet. Sammenhengen som er utledet empirisk fra RVU bør derfor kalibreres mot observerte telledata, og fordelingene bør kontrolleres for utvalgte relasjoner.

Vi har i tillegg sjekket om estimatene forandrer seg med bruk av kun fulltidsjobber som pendlerindeks, men dette endrer ikke verdiene.

Reisetid

Det er naturlig å tenke seg at tidspunktet for ettermiddagstoppen er seinere for stasjonsrelasjoner med lang reisetid. Vi tilpasser derfor en mixture of linear regressions modell med reisetid som forklaringsvariabel, slik at

$$\mu_k = \alpha_k + \beta_k \times \text{reisetid}, k=1,2.$$

Vi får parameterestimater som vises i Tabell 5.4.

Tabell 5.4: Parameterestimater for en mixture of linear regression modell for ankomsttid for arbeidsreiser med reisetid som forklaringsvariabel.

Parameter	RVU	NSB	NSB + 15 min	NSB + beregnet
λ_1	0,54	0,52	0,54	0,52
λ_2	0,46	0,48	0,46	0,48
α_1	482	452	484	480
α_2	996	978	993	971
β_1	0,09	0,39	0,00	-0,1
β_2	0,41	0,06	0,38	0,24
σ_1	73	50	73	52
σ_2	99	142	100	149

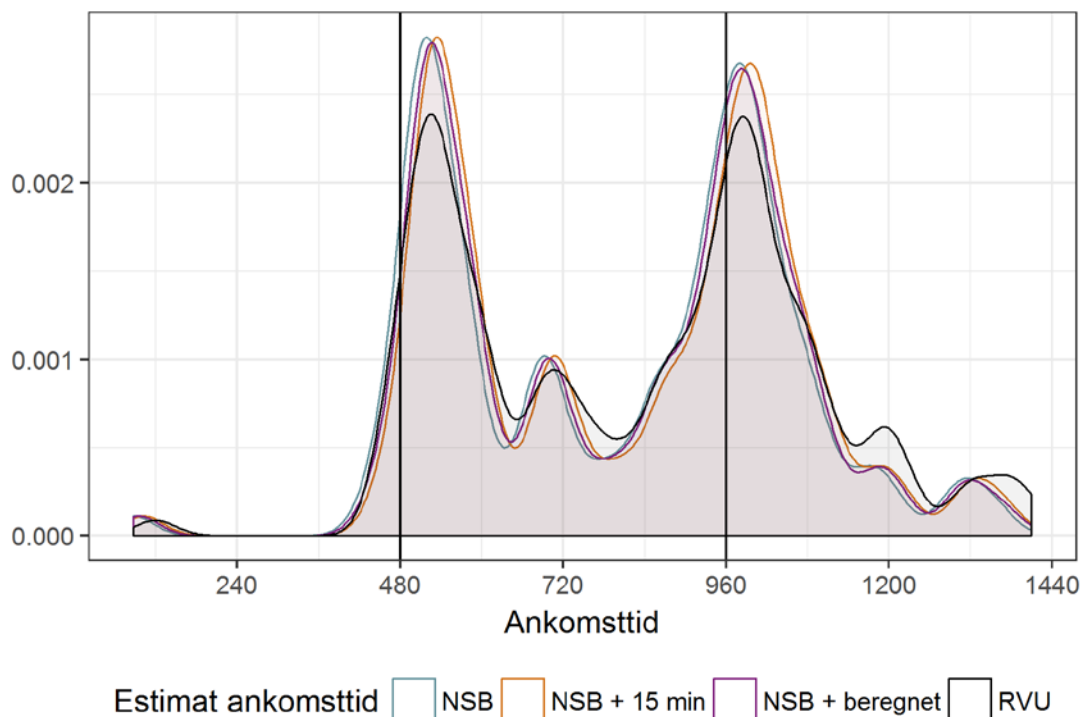
Tabellen viser at estimatene varierer mellom de ulike ankomstidsvariablene, og det er nødvendig med kalibrering mot telledata for å kunne konkludere med en endelig døgnfordelingsmodell for arbeidsreiser.

Ved å bruke relasjonsspesifikke variabler for relativ pendlerindeks og reisetid får vi fordelinger tilpasset de ulike relasjonene. For å unngå urealistiske parameterverdier vil det være nødvendig å sette noen restriksjoner på eksempelvis hvor seint ettermiddagstoppen kan komme hvis reisetiden er veldig lang.

Merk at det i alle tilfeller er estimert større standardavvik om ettermiddagen enn om morgenen. Dette var i mindre grad tilfellet når fordelingen ble estimert ikke-parametrisk.

5.4.2 Tjenestereiser

RVU datasettet inneholder svært få observasjoner for tjenestereiser med tog, kun 151 reiser totalt og 120 reiser hvis reisetidene basert på tilbudsmatrisen benyttes. Dette gir et svært dårlig estimeringsgrunnlag. I Figur 5.13 vises empiriske døgnfordelinger estimert ikke-parametrisk.



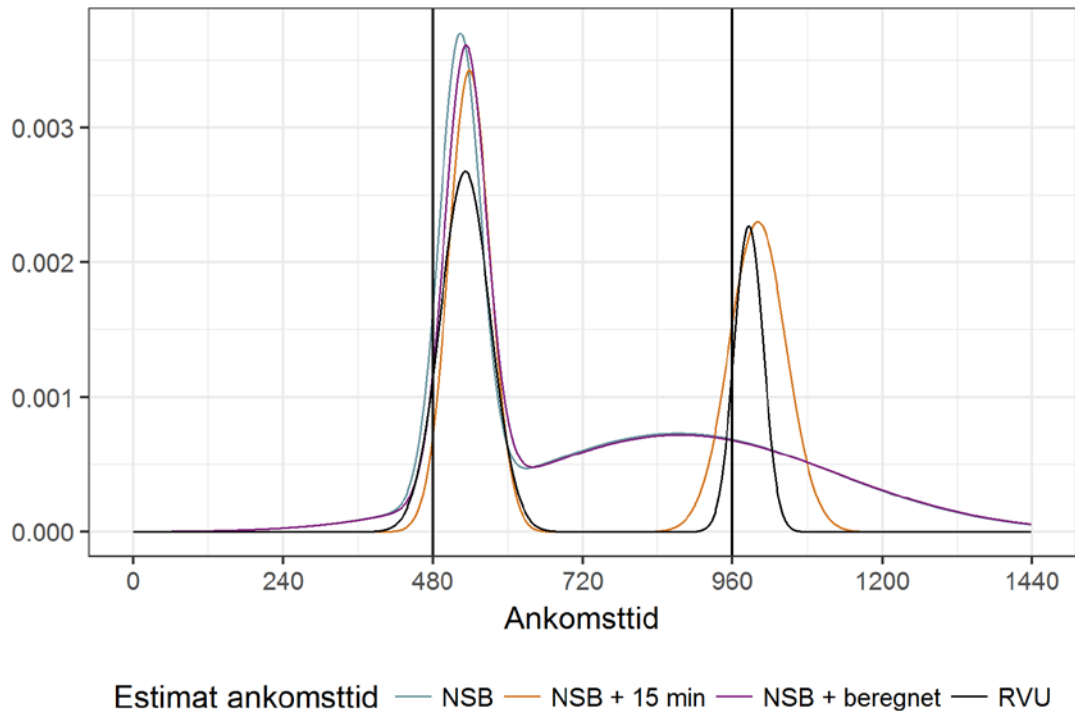
Figur 5.13: Døgnfordelinger over ankomsttid for tjenestereiser. Fordelingene er estimert ikke-parametrisk (kernel density estimation, med $bw = 30$). Figuren viser fordelinger for de fire ulike tidsvariablene. For RVU ankomsttiden er $N = 151$, for de resterende variablene er $N = 120$. Kl. 08:00 og kl. 16:00 er markert med vertikale linjer.

Fordelingene for tjenestereiser skiller seg fra arbeidsreisene på to punkter: morgentoppen kommer etter kl. 8 og det er en liten topp midt på dagen. På bakgrunn av dette velger vi å tilpasse en mixture fordeling med tre komponenter.

Mixture fordelingene vises i Figur 5.14 og parameterverdiene er vist i Tabell 5.5.

Tabell 5.5: Parameterestimatene for døgnfordelingene i Figur 5.14 (arbeidsreiser).

Parameter	RVU	NSB	NSB + 15 min	NSB + beregnet
λ_1	0,27	0,45	0,28	0,28
λ_2	0,13	0,28	0,26	0,45
λ_3	0,60	0,26	0,45	0,27
μ_1	532	873	538	532
μ_2	986	523	1000	876
μ_3	929	985	888	992
σ_1	40	248	33	34
σ_2	23	33	45	250
σ_3	235	45	248	48



Figur 5.14: Døgnfordelinger over ankomsttid for tjenestereiser. Fordelingene er spesifisert som mixtures med tre normalfordelte komponenter og estimert med EM-algoritmen. Figuren viser fordelinger for de fire ulike tidsvariablene. For RVU ankomsttiden er $N = 151$, for de resterende variablene er $N = 120$. Kl. 08:00 og kl. 16:00 er markert med vertikale linjer.

Mixture fordelingene fanger ikke opp den lille toppen omkring kl. 12 som vises på de ikke-parametriske fordelingene. Det er svært lite empirisk grunnlag til å si noe om tjenestereiser med tog, og når vi forsøker å ta inn reisetid som forklaringsvariabel blir det problemer med konvergens av EM-algoritmen. Hvis vi reduserer antall komponenter til to, får vi parameterestimaterne i Tabell 5.6.

Tabell 5.6: Parameterestimater for en mixture of linear regression modell for ankomsttid for tjenestereiser med reisetid som forklaringsvariabel.

Parameter	RVU	NSB	NSB + 15 min	NSB + beregnet
λ_1	0,28	0,43	0,76	0,30
λ_2	0,72	0,57	0,24	0,70
α_1	528	564	747	540
α_2	905	962	986	890
β_1	0,19	-0,27	0,10	-0,13
β_2	0,56	0,48	0,62	0,71
σ_1	38	99	252	34
σ_2	200	127	33	194

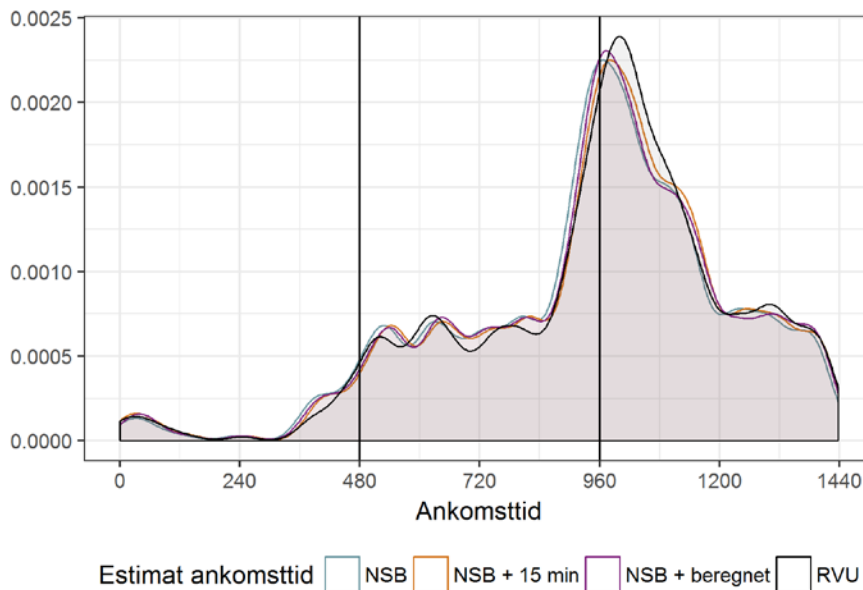
Disse estimatene er mer sprikende enn estimatene for arbeidsreisefordelingen og indikerer instabilitet. Da estimeringsgrunnlaget er så tynt som det er for tjenestereiser, kan en mulig løsning være å benytte arbeidsreisefordelingen og legge til en tredje komponent med forventningsverdi kl. 12 og stor varians for å få opp etterspørselen midt på dagen. Dette

blir en vurdering som må gjøres på bakgrunn av etablert kunnskap eller innhenting av ytterligere data.

5.4.3 Andre reiser

I kategorien andre reiser inngår reiser med formål handel/service, følge/omsorg, fritid, besøk eller annet. I RVU-datasettet er det 593 reiser i denne kategorien. Ankomsttider basert på tilbudsmatrisene finnes ikke for 131 av disse reisene.

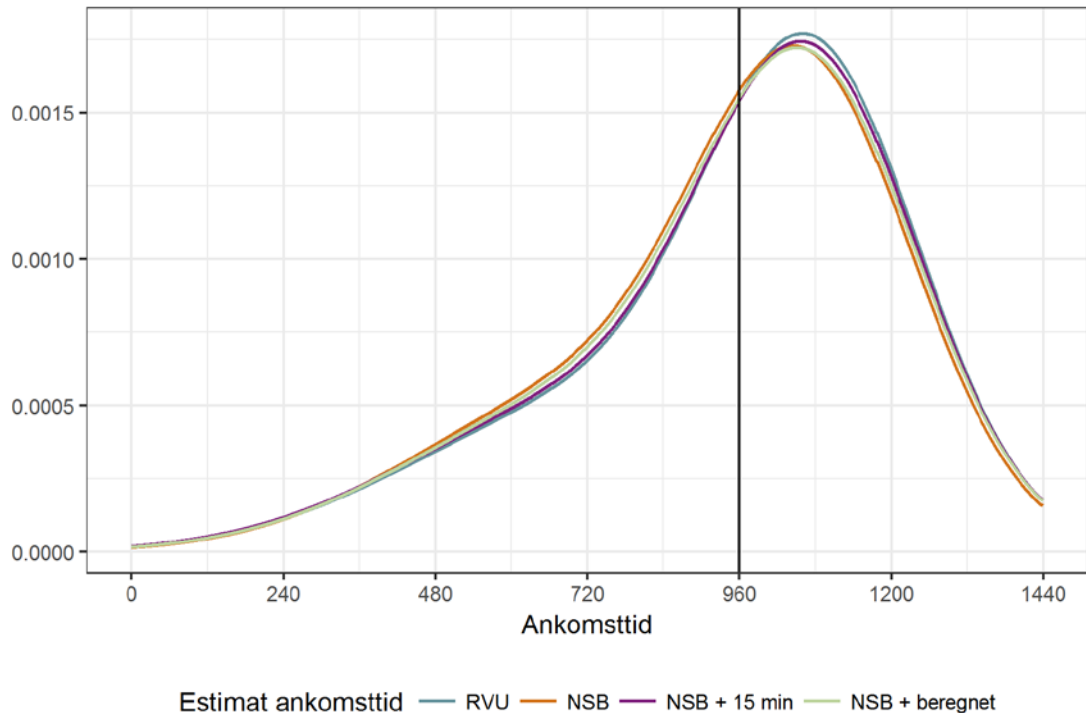
I Figur 5.15 viser vi hvordan døgnfordeling over ankomsttid ved endestasjon ser ut for andre reisehensikter for de fire ulike ankomsttidvariablene.



Figur 5.15: Empirisk døgnfordeling av andre reiser (kernel density estimate, $bw = 30$) for de ulike variablene for ankomsttid. For ankomsttiden basert på RVU er $N = 593$, for de øvrige er $N = 462$.

Fordelingene ser like ut, med unntak av at fordelingen basert på reisetiden angitt i RVU er forskjøvet noe mot høyre. Dette er rimelig med tanke på at denne ankomsttiden gjelder for reisens endepunkt og ikke endestasjon. Fordelingen ser ut til å bestå av to fordelinger, en relativt flat fordeling mellom kl. 8 og midnatt og en spiss fordeling med toppunkt kl. 16.

Basert på denne observasjonen tilpasses en mixture of normals med to komponenter. De estimerte fordelingene er i Figur 5.16 og parameterverdiene er i Tabell 5.7.



Figur 5.16: Døgnfordelinger over ankomsttid for andre/øvrige reisebensikter. Fordelingene er spesifisert som mixtures med to normalfordelte komponenter og estimert med EM-algoritmen. Figuren viser fordelinger for de fire ulike tidsvariablene. For RVU ankomsttiden er $N = 593$, for de resterende variablene er $N = 462$.

Tabell 5.7: Parameterestimer for døgnfordelingene i Figur 5.16 (andre reiser).

Parameter	RVU	NSB	NSB + 15 min	NSB + beregnet
λ_1	0,35	0,36	0,36	0,36
λ_2	0,65	0,64	0,64	0,64
μ_1	724	712	724	723
μ_2	1079	1067	1077	1073
σ_1	282	268	284	277
σ_2	168	172	170	173

Døgnfordelingene som er estimert som mixtures gjensker ikke helt de ikke-parametriske fordelingene. Etterspørselen blir høyere midt på dagen enn den er i Figur 5.15, men fordelingene viser et toppunkt like etter kl. 16 og fallende etterspørsel utover kvelden.

Med en annen definisjon av reisebensikt, slik at reiser fra jobb og inntil butikk/barnehage/etc. ville klassifiseres som en arbeidsreise, ville døgnfordelingen trolig ha en mindre markant topp omkring kl. 16.

Hvis vi tar inn reisetid som forklaringsvariabel for μ_1 og μ_2 får vi fordelinger med parameterne i Tabell 5.8.

Tabell 5.8: Parameterestimer for en mixture of linear regression modell for ankomsttid for andre reiser med reisetid som forklaringsvariabel.

Parameter	RVU	NSB	NSB + 15 min	NSB + beregnet
λ_1	0,35	0,32	0,35	0,34
λ_2	0,65	0,68	0,65	0,66
α_1	734	717	754	743
α_2	1068	1061	1073	1067
β_1	-0,04	-0,73	-0,74	-0,76
β_2	0,19	0,07	0,09	0,12
σ_1	281	238	267	254
σ_2	173	172	170	173

Her ser vi at ved lengre reisetid vil forventningsverdien for komponenten med forventningsverdi rundt kvart på seks forskyves til seinere om kvelden og forventningsverdien for komponenten med forventningsverdi midt på dagen forskyves til tidligere på dagen.

Vi undersøker også hva som skjer ved å bruke både reisetid og andelen handel/service-reiser fra reisehensiktsmodellen som forklaringsvariabler. Resultatene er i Tabell 5.9.

Tabell 5.9: Parameterestimer for en mixture of linear regression modell for ankomsttid for andre reiser med reisetid og andelen handelsreiser/andelen andre reiser som forklaringsvariabel.

Parameter	RVU	NSB	NSB + 15 min	NSB + beregnet
λ_1	0,98	0,97	0,22	0,37
λ_2	0,02	0,03	0,78	0,63
α_1	1057	1064	916	795
α_2	1653	1624	1046	1191
$\beta_1^{\text{reisetid}}$	-0,12	-0,51	1,07	-0,83
$\beta_2^{\text{reisetid}}$	-0,54	-0,56	-0,54	-0,15
$\beta_1^{\text{handelsandel}}$	-270	-284	147	-67
$\beta_2^{\text{handelsandel}}$	-698	-660	-230	-284
σ_1	268	263	63	261
σ_2	1	7	303	167

Det er stor variasjon i parameterestimaterne og tydelige tegn på instabiliteter. Selv om det er ønskelig å beskrive fordelingene ved hjelp av andre egenskaper enn reisetid, velger vi å kun benytte reisetid som forklaringsvariabel for døgnfordelingene også for andre reisehensikter. Med et større datasett vil det være naturlig å teste variabler som beskriver stasjonsomlandene for å undersøke om dette innvirker på døgnfordelingens form.

5.4.4 Oppsummering

Dette delkapittelet har presentert en lang rekke modeller uten å trekke noen endelig konklusjon. Før døgnfordelingene spesifiseres vil vi presentere resultater basert på telldata fra NSB.

Mixture fordelingene for *arbeidsreiser* og *andre reiser* virker rimelige, men det er stor usikkerhet knyttet til tjenestereisene. Kjente svakheter ved modellene presentert over er knyttet til

- usikkerheten i tilordningen av togstasjoner til grunnkretser,
- usikkerheten i beregningen av ankomsttid ved endestasjon,
- få observasjoner for tjenestereiser og andre reiser med tog,
- fordelingene er (før omregning) kontinuerlige,
- fordelingene har en diskontinuitet ved midnatt.

5.5 Mixture parametere fra telldata

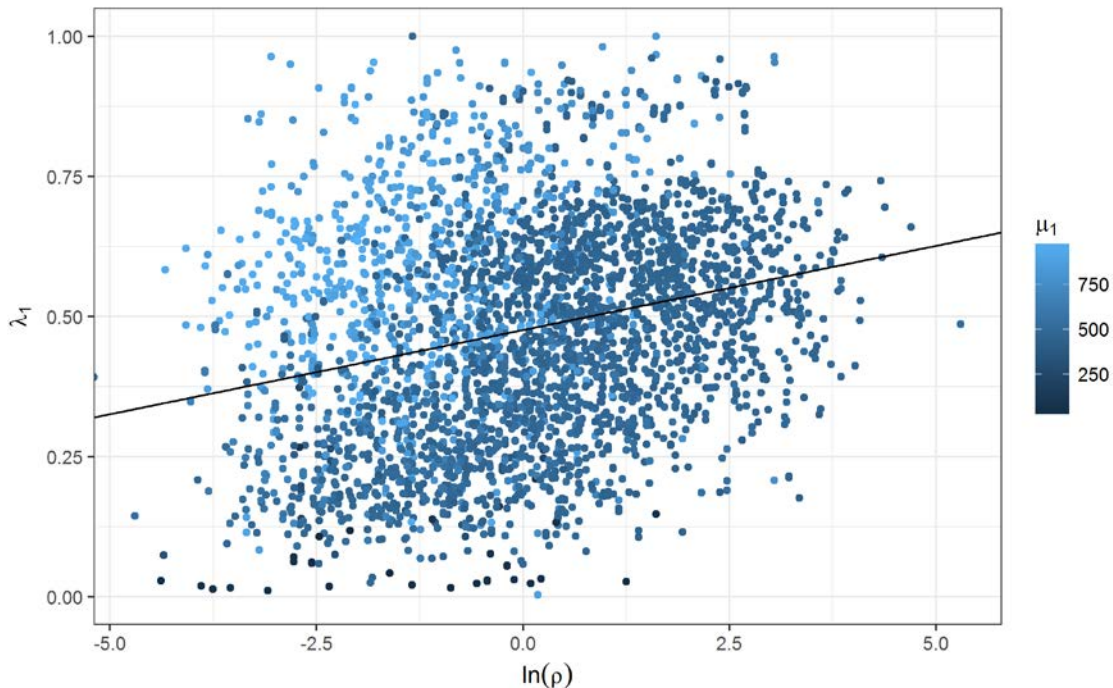
Telldataene er beskrevet i kapittel 3.1.4 og 3.3. Når telldataene er omgjort til relasjonsnivå har de form som en diskret sannsynlighetsfordeling der andelen reisende angir sannsynligheten for den gitte ankomsttiden. Fordelen med disse dataene er at vi her har tilgang til faktisk ankomsttid ifølge rutetabellen, men det er knyttet en del usikkerhet til representativiteten av tellingene og omgjøringen av tellinger fra snitt til OD-format. I tillegg inneholder ikke disse dataene informasjon om reisehensikt, slik at det er summen av alle typer reiser som observeres med passasjertellinger og det må derfor forventes avvik mellom de to datakildene. Det er også mye variasjon i telldataene, især for korte relasjoner. Vi ønsker likevel å sammenligne mixture parametere vi estimerer fra telldata med estimatene vi fant basert på RVU.

Vi tilpasser mixtures of normals med to komponenter ved først å samle observasjoner fra «telle-fordelingen» for en gitt relasjon, det vil si at vi trekker minutt i døgnet med sannsynlighet lik den korresponderende andelen i telle-fordelingen. Vi trekker 10 000 ankomsttider fra telle-fordelingen for en relasjon og tilpasser deretter en mixture til dette datasettet. Da får vi, for hver stasjonsrelasjon vi har telldata for, et sett av parametere

$$\theta = (\lambda_1, \lambda_2, \mu_1, \mu_2, \sigma_1, \sigma_2).$$

Ved å koble reisetid og relativ pendlerindeks for stasjonsrelasjonene til parameter-datasettet, kan vi undersøke sammenhengen mellom disse variablene og mixture parameterne.

Først undersøker vi om også telldata støtter en empirisk sammenheng mellom relativ pendlerindeks og komponent-vektene. I Figur 5.17 har vi plottet $\ln \rho$ mot λ_1 .



Figur 5.17: Sammenhengen mellom relativ pendlerindeks og vekten for komponent nummer én.

Det er ingen tydelig sammenheng, men ved å fargelegge punktene basert på verdien til μ_1 , kommer det fram at det er relasjoner der forventningsverdien for første komponent er høyere enn 720 (etter kl. 12:00) og lavere enn 240 (kl. 04:00) som avviker mest fra en lineær sammenheng.

Hvis vi tilpasser en lineær funksjon til punktene i figuren over finner vi sammenhengen:

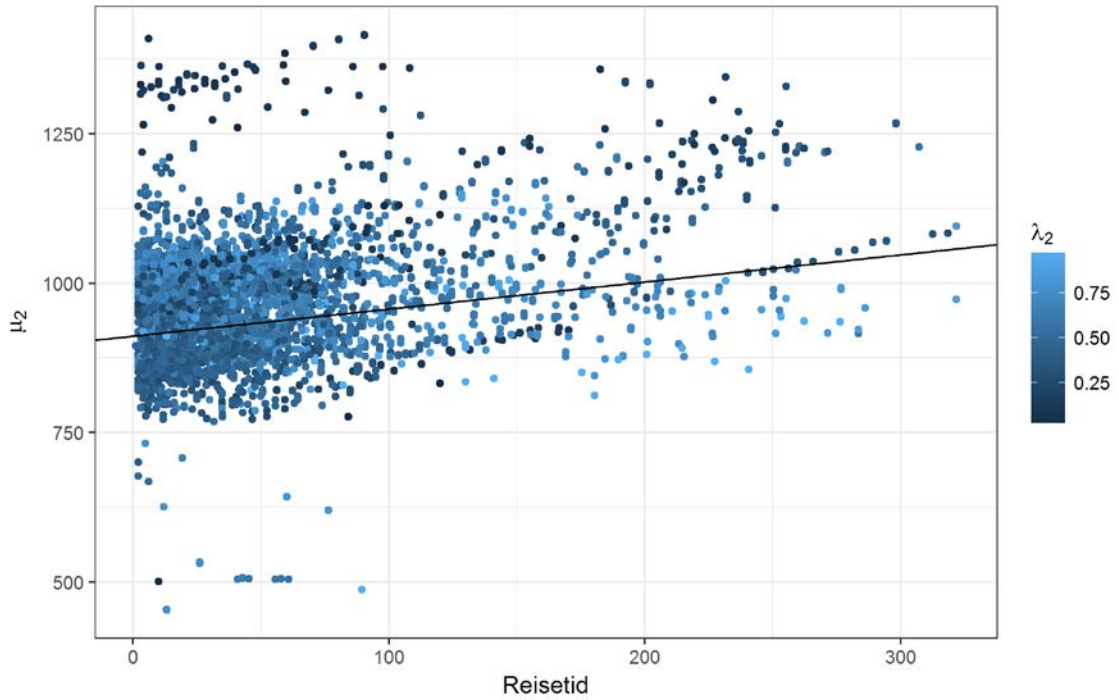
$$\lambda_1 = 0,46 + 0,03 \times \ln \rho.$$

Dette er linjen tegnet inn på figuren ($R^2=0,064$). Hvis kun relasjoner der μ_1 er mellom kl. 06 og 12, blir $R^2=0,22$ og parameterne 0, 43 og 0,05.

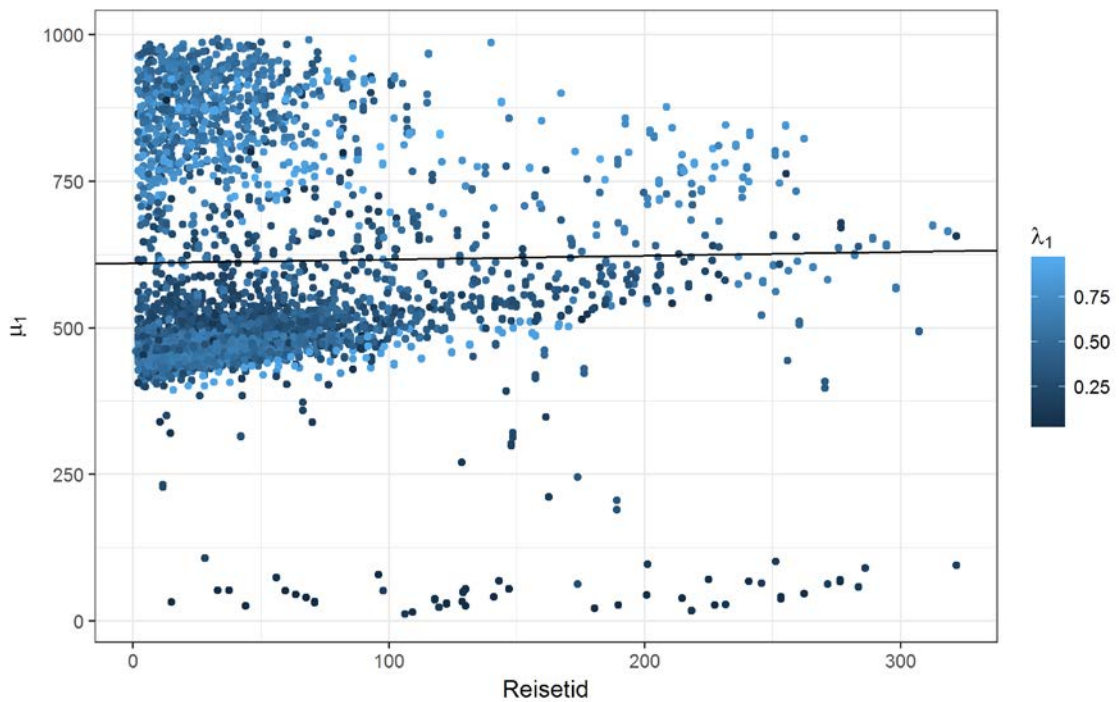
Til sammenligning var verdiene basert på RVU $\gamma_0 = 0,4$ og $\gamma_1 = 0,2$.

Tatt i betraktning støy fra tellinger, konvertering fra snitt til OD, usikkerheten i estimeringen av mixtures fra telldata, samt at data ikke er splittet opp etter hensikt er det ikke å forvente å finne en sterk sammenheng mellom vektene og den relative pendlerindeksen basert på telldata. Det er likevel interessant å observere at også basert på telldata estimeres konstantleddet til å være mindre enn 0,5.

Når det gjelder parameterne μ_1 og μ_2 lot vi disse avhenge av reisetiden mellom stasjonsrelasjonene. I Figur 5.18 og Figur 5.19 viser vi disse sammenhengene basert på telldata.



Figur 5.18: Sammenhengen mellom reisetid og forventningsverdien for komponent nummer to.



Figur 5.19: Sammenhengen mellom reisetid og forventningsverdien for komponent nummer én.

Det er mye støy i disse parameterverdiene, men vi kan likevel tilpasse lineære regresjonsmodeller og får følgende sammenhenger:

$$\begin{aligned}\mu_1 &= 577 - 0,03 \times \text{reisetid} \\ \mu_2 &= 945 + 0,55 \times \text{reisetid}\end{aligned}$$

For μ_2 er reisetiden en signifikant forklaringsvariabel, det er den ikke for μ_1 . Dette stemmer med antakelsen om at om morgenen er ankomsttiden styrt av når man starter på jobb, mens om ettermiddagen vil ankomsttiden styres av lengden på togreisen.

Tolkningen av sammenhengen mellom μ_2 og reisetiden er at for hvert minutt tillegg i reisetid blir ankomsttiden 0,55 minutter seinere. Minutt 945 er kl. 15:45 og for en stasjonsrelasjon med en reisetid på en halvtime får ettermiddagskomponenten forventningsverdi kl. 16:02.

I Tabell 5.10 oppsummeres lokasjons- og skalaparameterne estimert fra telldata. Disse verdiene kan fungere som en hjelp til å sette begrensninger på parameterverdiene.

Tabell 5.10: Oppsummering av parameterestimatene basert på telldata.

	Minimum	Nedre kvartil	Median	Gjennomsnitt	Øvre kvartil	Maksimum
μ_1	12	468	513	614	792	1414
μ_2	22	890	955	936	1018	1415
σ_1	5	51	72	135	231	579
σ_2	4	89	185	172	242	495

5.6 Oppsummering av døgnfordelinger

Det er utfordrende å skulle trekke noen endelig konklusjon om døgnfordelingene. Parameterverdiene vi finner spriker i enkelte tilfeller ganske mye, samtidig støtter både RVU-observasjonene og telldata en intuisjon om hvordan fordelingene bør se ut. Oppsummert finner vi empirisk støtte for at:

- Det er større spredning i ankomsttid om ettermiddagen enn om morgenen.
- Lengre reisetid mellom stasjonene gjør at ettermiddagstoppen kommer på et senere tidspunkt.
- Reisetid har lite å si for ankomsttiden om morgenen.
- Den relative pendlerindeksen fungerer som et mål på om komponent én eller to skal få størst vekt.

Det utfordrende med å ha flere kilder med data er at det er vanskelig å si hvilke som bør ilegges mest vekt. I vårt tilfelle har vi RVU, der det på den ene siden er dokumentert når respondenter starter en reise med ulike reisehensikter, mens det på den andre siden er knyttet større usikkerhet til når de faktisk ankom sin endestasjon. Fordelen med RVU er at vi har både tidfesting, stedsfesting og reisehensikt. For telldataene har vi ingen informasjon om reisehensikt, og informasjonen om relasjonsfordelingen er estimert. Men telldataene er tidfestet ved hjelp av rutetabeller med minuttpresisjon, en fordel sammenlignet med de relativt grove antakelsene gjort i beregningen av tilbringertid.

Vi kan ikke etablere døgnfordelinger med høyere presisjonsnivå enn datagrunnlaget tilsier, og med forbehold om dette, samt den metodiske usikkerheten knyttet til modelleringsvalgene som er gjort, konkluderer vi med følgende døgnfordelinger:

For stasjonsrelasjon (i, j) og reisehensikt H , har vi at

$$f_{i,j}^H(t) = \frac{\lambda_1^H N(t; \mu_1^H, \sigma_1^H) + \lambda_2^H N(t; \mu_1^H, \sigma_1^H)}{C_{i,j}^H}, t \in \{0, 1, 2, \dots, 1439\}, \quad (3)$$

der $\lambda_2^H = 1 - \lambda_1^H$ og normaliseringskonstanten er gitt ved

$$C_{i,j}^H = \sum_{t=0}^{1439} \lambda_1^H N(t; \mu_1^H, \sigma_1^H) + \lambda_2^H N(t; \mu_1^H, \sigma_1^H).$$

For enklere notasjon har vi ikke tatt med at mixture-vektene og lokasjonsparameterne er relasjonsavhengige. Vi oppsummerer fordelingene for de ulike hensiktene under.

I valget av parameterverdier er det sett på parameterestimaterne med og uten reisetid som forklaringsvariabel basert på RVU, og estimatene basert på telldata, og gjort et kompromiss mellom disse, samt en avrunding til hele timer, eksempelvis kl. 16:00 i stedet for 15:57. Denne vurderingen bør etterprøves ved å validere fordelinger for viktige relasjoner med enten passasjertellinger eller etablert kunnskap.

Arbeidsreiser

For arbeidsreiser er hovedtrekket at mixture-vektene bestemmes av den relative pendlerindeksen og at lokasjonsparameteren for ettermiddagskomponenten er avhengig av reisetiden. Vi velger å la morgenkomponenten være konstant. Vi har at

$$\lambda_1^{arbeid}(i, j) = \min \{ 1, \max \{ 0, \gamma_0 + \gamma_1 \ln \rho_{ij} \} \}.$$

og at $\mu_2^{arbeid} = \alpha^{arbeid} + \beta^{arbeid} \times T_{ij}$, der T_{ij} er reisetiden fra stasjon i til stasjon j .

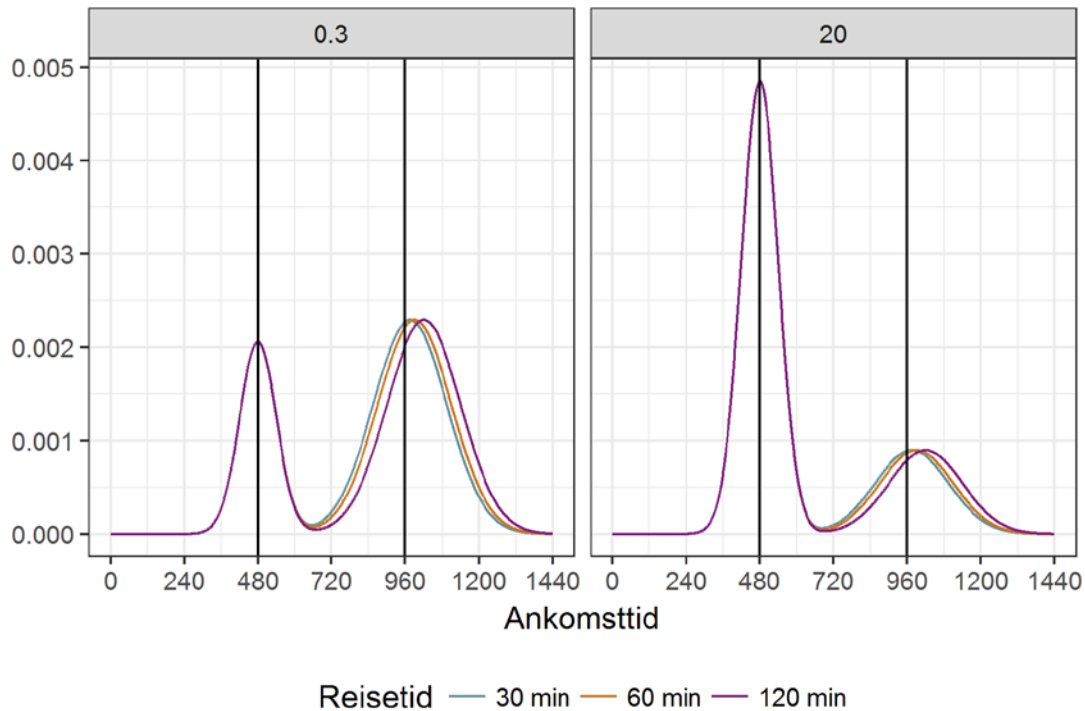
Parameterne vi konkluderer med vises i Tabell 5.11.

Tabell 5.11: Parameterverdier for døgnfordelingsmodellen for arbeidsreiser.

Parameter	Verdi
γ_0	0,43
γ_1	0,1
μ_1^{arbeid}	480 (kl. 08:00)
α^{arbeid}	960 (kl. 16:00)
β^{arbeid}	0,5
σ_1^{arbeid}	60
σ_2^{arbeid}	120

Vi legger inn en begrensning slik at lokasjonsparameteren for ettermiddagskomponenten ikke kan bli større enn 1080 (kl. 18:00).

Figur 5.20 illustrerer døgnfordelinger for ulike verdier av reisetid og relativ pendlerindeks.



Figur 5.20: Den etablerte døgnfordelingsmodellen for arbeidsreiser for tre forskjellige reisetider og to verdier av den relative pendlerindeksen: 0,3 i venstre panel og 20 i høyre panel.

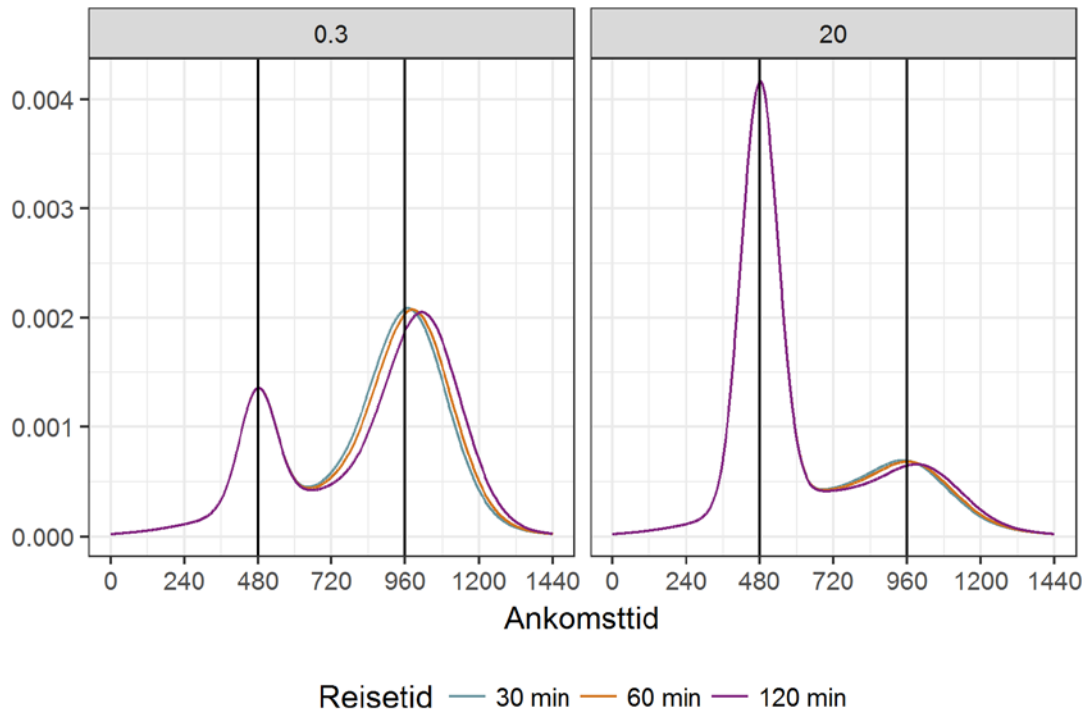
Tjenestereiser

På grunn av svært lite datagrunnlag for tjenestereiser med tog foreslår vi å benytte døgnfordelingene for arbeidsreiser med en liten justering slik at andelen reiser midt på dagen blir høyere.

Dette kan oppnås ved å innføre en tredje komponent med lokasjonsparameter kl. 12. Vi setter $\mu_0^{tjeneste} = 720$ og $\sigma_0^{tjeneste} = 300$. De fire andre lokasjon- og skalaparameterne beholdes som i Tabell 5.11, men vektene må justeres for å ta høyde for en tredje komponent. Vi setter $\lambda_0^{tjeneste} = 0,3$.

Dette gir døgnfordelinger som i Figur 5.21. Disse er svært like som arbeidsreisefordelingene, men med større etterspørsel midt på dagen.

Vi antar her at tjenestereiser følger samme mønster som arbeidsreiser og det er usikkert hvorvidt denne løsningen er realistisk. Antakelsen baserer seg på formen på den empiriske fordelingen i Figur 5.4, og betyr at de samme faktorene som genererer og attraherer arbeidsreiser virker inn på samme måte for tjenestereiser. I enkelte tilfeller kan dette kanskje stemme, men trolig er det en for grov tilnærming. Men med så lite datagrunnlag som finnes om tjenestereiser med tog, er dette vårt beste forslag. Vi vil anbefale innhenting av ytterligere data for å etablere bedre kunnskap om etterspørselen etter tjenestereiser med tog. Vi kommer tilbake til dette i kapittel 5.9.



Figur 5.21: Den etablerte døgnfordelingsmodellen for tjenestereiser for tre forskjellige reisetider og to verdier av den relative pendlerindeksen: 0,3 i venstre panel og 20 i høyre panel.

Andre reiser

For de andre reisehensiktene (besøk, handel, etc.), er ikke pendlerindeksen viktig da dette er en type reiser som i stor grad foretas etter arbeidstid uavhengig av relasjon. Mixturevektene er estimert til henholdsvis 0,35 og 0,65.

På samme måte som for de andre hensiktene velger vi å la lokasjonsparameteren for morgenkomponenten være konstant, mens lokasjonsparameteren for ettermiddagskomponenten er gitt ved:

$$\mu_2^{andre} = \alpha^{andre} + \beta^{andre} \times T_{ij},$$

der T_{ij} er reisetiden fra stasjon i til stasjon j .

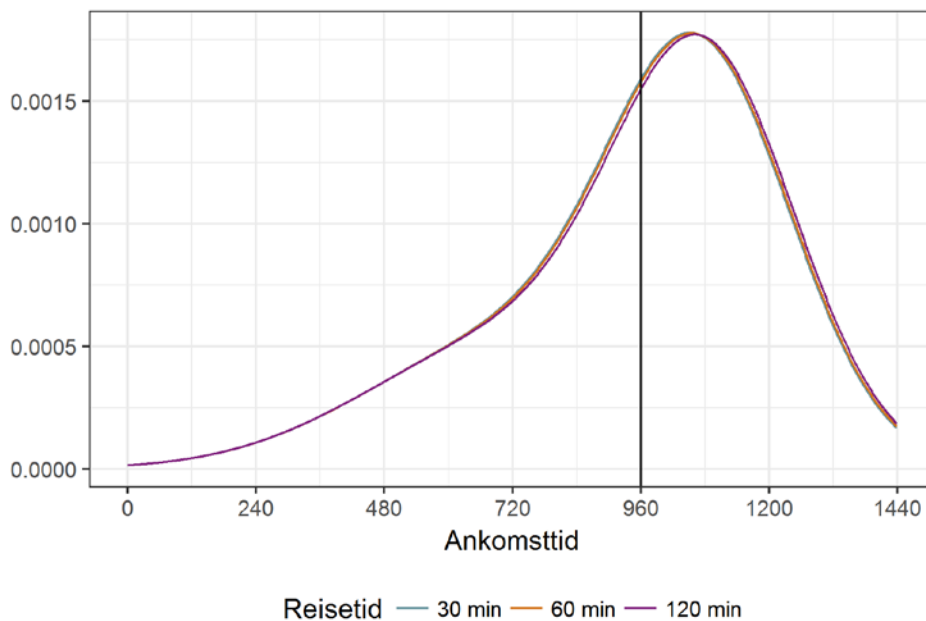
En oppsummering av parameterverdiene finnes i Tabell 5.12.

Tabell 5.12: Parameterverdier for døgnfordelingsmodellen for øvrige reiser.

Parameter	Verdi
λ_1	0,35
λ_2	0,65
μ_1^{andre}	720 (kl. 12:00)
α^{andre}	1070 (kl. 17:50)
β^{andre}	0,1
σ_1^{andre}	270
σ_2^{andre}	170

Vi legger inn en begrensning slik at lokasjonsparameteren for ettermiddagskomponenten ikke kan bli større enn 1140 (kl. 19:00).

Figur 5.22 viser den etablerte døgnfordelingsmodellen for tre forskjellige reisetider. Vi ser at reisetid ikke gir særlig synlige utslag i modellen.



Figur 5.22: Den etablerte døgnfordelingsmodellen for øvrige reiser for tre forskjellige reisetider.

Som nevnt tidligere har vi forsøkt å implisitt inkludere egenskaper ved stasjonsomlandene ved å benytte andeler fra reisehensiktsmodellen som forklaringsvariabler. Dette gav resultater som varierte mye med hvilken ankomstidsvariabel som ble brukt, i tillegg var estimatene av en slik størrelsesorden at det var kun snakk om små forskyvinger i lokasjonsparameterne. Vi valgte derfor en enkel modell med reisetid som eneste forklaringsvariabel, og som man ser av figuren over gir heller ikke dette noen særlig endring i døgnfordelingen for andre reisehensikter.

5.7 Validering av de endelige døgnfordelingene mot transformerte passasjertellinger

For å undersøke hvor godt vi treffer med døgnfordelingene sammenligner vi marginalfordelingen for ankomsttid med telldata for utvalgte stasjonsrelasjoner.

Fordelingen over ankomsttid er beregnet som

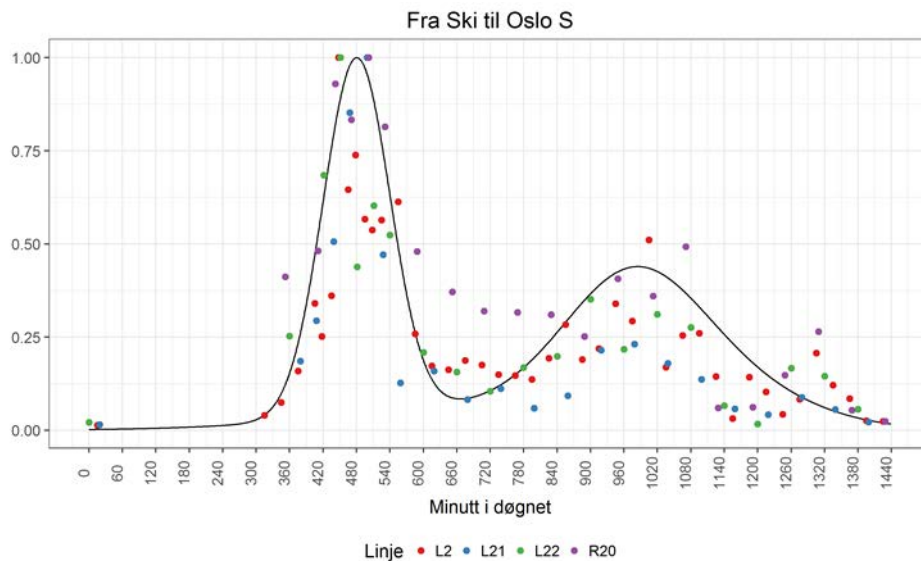
$$f_{i,j}(t) = \sum_h f_{i,j}(t|h)g_{i,j}(h),$$

der $g_{i,j}(h)$ er hensiktsfordelingen og h er *arbeid, tjeneste* eller *andre reisehensikter*.

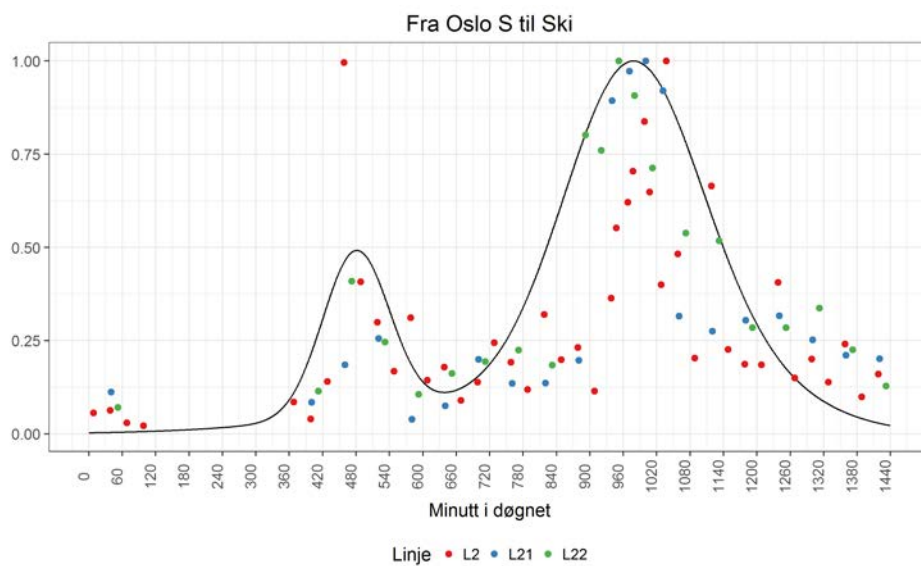
Etter innspill fra Jernbaneverket har vi undersøkt hvor godt vi treffer for relasjonene:

- Oslo S – Lillestrøm
- Oslo S – Ski
- Oslo S – Drammen
- Oslo S – Fredrikstad
- Oslo S – Hamar
- Oslo S – Tønsberg

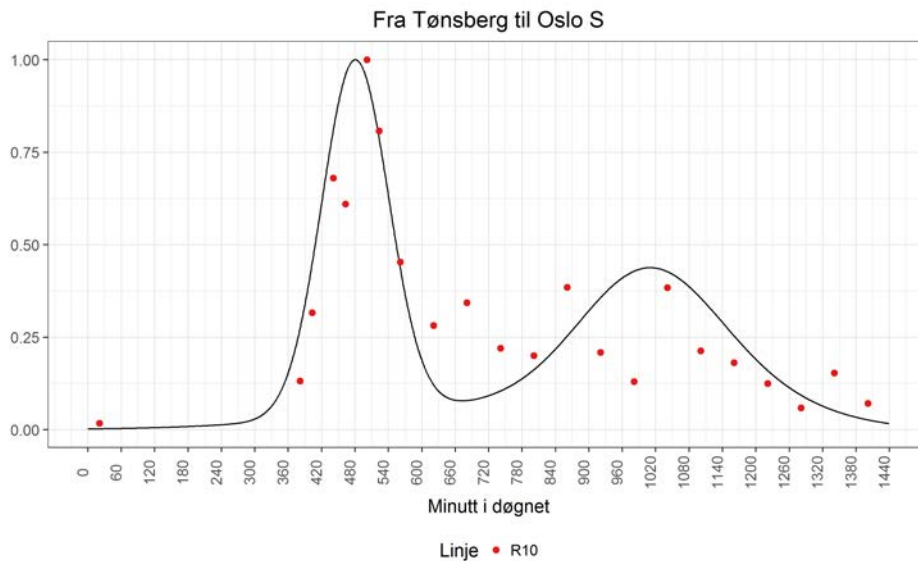
Vi har undersøkt begge retninger og alle figurer finnes i vedlegg 4. Under viser vi en sammenligning for utvalgte relasjoner. Den heltrukne linjen er den estimerte fordelingen, mens punktene angir ankomsttidsfordelingen for aktuelle toglinjer.



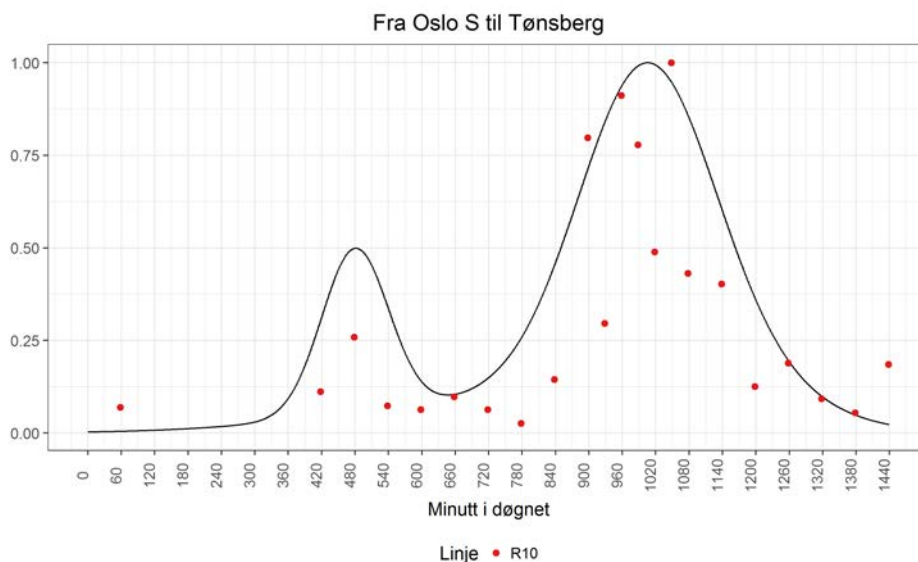
Figur 5.23: Validering av døgnfordeling (alle bensikter) mot transformerte passasjertellinger for relasjonen Ski – Oslo S. Fordelingen og passasjertellingene er normert ved å dele på respektive maksimumsverdier for visualiseringens skyld.



Figur 5.24: Validering av døgnfordeling (alle bensikter) mot transformerte passasjertellinger for relasjonen Oslo S – Ski. Fordelingen og passasjertellingene er normert ved å dele på respektive maksimumsverdier for visualiseringens skyld.



Figur 5.25: Validering av døgnfordeling (alle hensikter) mot transformerte passasjertellinger for relasjonen Tønsberg – Oslo S. Fordelingen og passasjertellingene er normert ved å dele på respektive maksimumsverdier for visualiseringens skyld.



Figur 5.26: Validering av døgnfordeling (alle hensikter) mot transformerte passasjertellinger for relasjonen Oslo S – Tønsberg. Fordelingen og passasjertellingene er normert ved å dele på respektive maksimumsverdier for visualiseringens skyld.

Som forventet er det ikke perfekt samsvar, men gitt

1. støy/tilfeldigheter i de opprinnelige telldata (manuelle tellinger)
2. usikkerhet rundt vår transformering fra snitt til OD
3. usikkerhet i reisehensiktsmodellen
4. tynt datagrunnlag i RVU for estimering av døgnfordelinger

treffer disse fordelingene ganske godt.

Det er verdt å merke seg at den relative pendlerindeksen riktig fanger opp om hovedvekten av etterspørselen er om morgenen eller ettermiddagen. I tillegg er det godt samsvar mellom tidspunkt for når toppene kommer for døgnfordelingene og telldata.

For relasjonen Oslo S – Ski viser telldata at det er like mange som ankommer med linje L2 om morgenen som om ettermiddagen. Basert på den relative pendlerindeksen karakteriseres punktet om morgenen som en avvikende observasjon, noe som stemmer med vår kunnskap om strekningen. Observasjonen illustrerer hvordan tilfeldigheter og variasjon i passasjertellinger kan gi store utslag. Effekter som dette forklarer mye av støyen som kunne observeres i figurene i kapittel 5.5.

For enkelte relasjoner ser det ut som etterspørselen midt på dagen og om kvelden estimeres som for lav. Dette kan skyldes at det er flest arbeidsreiser i RVU slik at andelen tjenestereiser og øvrige reiser estimeres som lavere enn de faktisk er.

5.8 Implementering av modell

Tilpassing og uttesting av døgnfordelingene er gjort med programvaren R. De endelige modellene som presentert i kapittel 5.6 er implementert i Microsoft Excel.

Til oppdragsgiver leveres:

1. CSV-filer med døgnfordelinger for alle relasjoner.
2. Excel-fil med inndata til døgnfordelingene: relativ pendlerindeks og reisetid for relasjoner der dette er tilgjengelig.
3. Excel-filer med beregning av døgnfordelingene (en fil for hver hensikt).

Modell-filene følger en mal med egne ark for parameterverdier, restriksjoner, inndata, beregninger – og et eget ark for den endelige fordelingen. Ved å endre på inndata (reisetid og relativ pendlerindeks) vil det genereres en ny fordeling.

5.9 Svakheter ved døgnfordelingene og behov for nye data

Ambisjonen i prosjektet var å måle og statistisk forklare døgnfordeling over ankomsttid ved endestasjon som funksjon av både reisehensikt og variabler som beskriver relasjonen og stasjonsomlandene. Dette krever mye data og vi har – med det tilgjengelige datagrunnlaget – klart å finne en sammenheng mellom døgnfordelingene for arbeidsreiser og pendlerstatistikk og reisetid. Andre forklaringsvariabler (sonedata og andeler fra hensiktsmodellen) har blitt testet for de andre døgnfordelingene, men datagrunnlaget har vist seg å være for tynt til å kunne identifisere flere stabile effekter. Som konsekvens framstår de estimerte fordelinger nokså homogene over stasjonsrelasjoner, spesielt for øvrige reiser der pendlerindeksen ikke inngår i modellen. Fordelingene vil i virkeligheten trolig variere mer enn det modellen predikerer. Samtidig tror vi at ønsket ankomsttid faktisk kan antas å være relativt stabil etter at man har kontrollert for reisehensikt og reiselengde, og det er ikke åpenbart hvorfor en gitt stasjonsrelasjon skulle ha en signifikant annen døgnfordeling enn en annen stasjonsrelasjon. I dette avsnittet oppsummerer vi utfordringer og mulige feilkilder fra datagrunnlaget, og diskuterer behov for nye data.

Ved oppstarten av prosjektet forventet vi å kunne bruke passasjertellinger som primær datakilde for døgnfordelingene. Passasjertellingene for de utvalgte togstrekningene kom på snitt-format, dvs. antall passasjerer ombord mellom to stasjoner. Tidsangivelsen for hvordan etterspørselen fordeler seg over døgnet er precis med disse dataene, men ulempen er at vi ikke vet hvilke togstasjoner passasjerene kommer fra og reiser til. For å kunne si noe om ankomsttid ved stasjon B, gitt at man reiste fra stasjon A, har vi først estimert av- og påstigninger ved hver stasjon og deretter kombinert dette med årsmatriser over antall

reisende mellom stasjonspar. Dette gir estimerte passasjertellinger på OD-format, men telldata mangler informasjon om reisehensikter. Tanken var å beregne tidsdifferensierte hensiktsfordelinger basert på RVU, det vil si andelen arbeidsreiser, tjenestereiser og øvrige reiser for hver time i døgnet, og bruke dette til å splitte opp telldata etter hensikter. Dette viste seg å ikke være en farbar vei fordi datagrunnlaget for togreiser i RVU var så tynt at det ikke lot seg gjøre å estimere tidsdifferensierte hensiktsfordelinger for ulike relasjonsegenskaper. Da det ikke gav mening å bruke samme oppsplittingsfaktor for alle relasjoner, valgte vi å bruke RVU som vår viktigste kilde til informasjon om ønsket ankomsttid gitt reisehensikt.

Det har vært nødvendig å tilpasse RVU til prosjektets formål. Dette medfører at vi med våre antakelser har innført feilkilder det er viktig å være klar over. For eksempel vil tilordningen av nærmeste togstasjon til grunnkretser medføre at vi for eksempel ikke fanger opp reiser der en person som bor i Ås velger å kjøre bil til Ski stasjon og ta toget derfra fordi det gir flere togavganger å velge mellom inn til Oslo enn fra Ås stasjon. I tillegg er ankomsttid ved endestasjon funnet ved hjelp av NSBs tilbudsmatriser og beregning av tilbringertid til startstasjonen. På grunn av usikkerheten i datagrunnlaget bruker vi avrundede parameterestimer i de endelige modellene.

Prosjektet har illustrert behovet for bedre data om togreiser og togreisende. For tjenestereiser er datagrunnlaget i RVU svært tynt. Vi har hentet ut reiser der tog er hovedtransportmiddel, men for mange tjenestereiser kan man tenke seg at toget fungerer som tilbringertransportmiddel til en flyplass og at hovedtransportmiddelet er fly. Da vi kun vet start- og endegrunnkrets for hele reisen kan vi ikke hente ut den nødvendige informasjonen om selve togreisen. Det hadde vært ønskelig med reisevanedata der hovedfokus var togreiser. Vi vet at det finnes kundetilfredsundersøkelser blant togreisende der reisehensikt rapporteres, men dette er data vi ikke har fått tilgang til i dette prosjektet. I tillegg finnes årsmatriser blant annet basert på NSBs billettstatistikk. Hvis informasjon om billettkjøp var tilgjengelig på et mer disaggregert nivå, kunne billettstatistikk (for enkeltreiser) være en viktig datakilde for å tidfeste togreiser på OD format.

Forbedringspotensialet for døgnfordelingsmodellene knytter seg til videre uttesting av forklaringsvariabler for stasjonsomlandene som kan fange opp variasjon mellom relasjoner, samt innhenting av nye og bedre datakilder.

6 Referanser

Benaglia, T., Chauveau, D., Hunter, D. R. & Young, D. S. (2009). mixtools: An R Package for Analyzing Finite Mixture Models. *Journal of Statistical Software*, 32(6), 1-29. <http://www.jstatsoft.org/v32/i06/>

Bierlaire, M. (2008). An introduction to BIOGEME Version 1.8: <http://biogeme.epfl.ch/v18/tutorialv18.pdf>.

Dempster, A. P., Laird, N.M. & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society: Series B*, 39(1):1-38.

Flügel, S. & Hulleberg, N. (2016). Trenklin 2 – Gjennomgang av modellen og drøfting av anvendelsesområde. *TØI rapport 1534/2016*. Transportøkonomisk institutt, Oslo.

Hjorthol, R., Engebretsen, Ø. & Uteng, T. P. (2014). Den nasjonale reisevaneundersøkelsen 2013/14 – nøkkelrapport. *TØI rapport 1383/2014*. Transportøkonomisk institutt, Oslo.

Manski, C., & Lerman, S. (1977). The Estimation of Choice Probabilities from Choice Based Samples. *Econometrica*, 45(8), 1977-1988. doi:10.2307/1914121

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

Ranheim (2016) Trenklin versjon 2.8: Inkrementell etterspørselsmodell for jernbanetraffikk, Utkast per 8.02.2016

Rekdal J., Odd I. Larsen, Arne Løkketangen og Tom N. Hamre (2013) TraMod_By Del 1: Etablering av nytt modellsystem RAPPORT 1203, Møreforkning.

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society: Series B*, 71(2):319-392.

Wand, M. (2015). KernSmooth: Functions for Kernel Smoothing Supporting Wand & Jones (1995). R package version 2.23-15. <https://CRAN.R-project.org/package=KernSmooth>

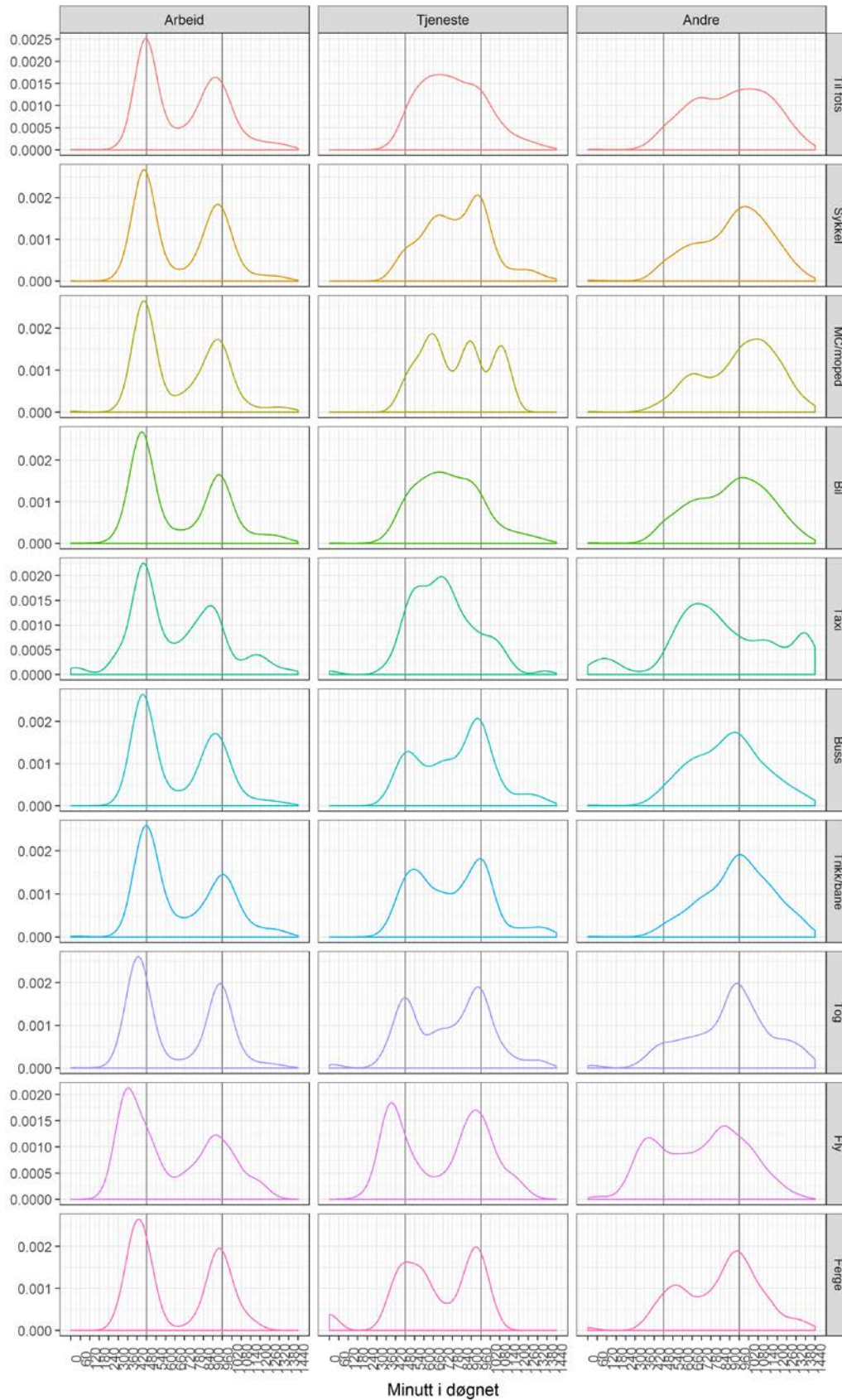
Vedlegg

Vedlegg 1: Stasjoner som ikke er kodet inn i RTM og/eller ikke finner nærmeste grunnkrets

Stasjon	Mangler	Kommentar
Berekvam stasjon	X	Flåmsbanen
Bjørnfjell stasjon	X	Ofofbanen
Blomheller stasjon	X	Flåmsbanen
Ed stasjon	X	Sverige
Flekkefjord stasjon	X	Ikke kodet inn i RTM, sidespor til Sørlandsbanen
Flåm stasjon	X	Flåmsbanen
Göteborg Central	X	Sverige
Hallingskeid stasjon	X	Ikke kodet inn i RTM, stasjonen har ikke bilveiforbindelse
Håreina stasjon	X	Flåmsbanen
Karlstad C	X	Sverige
Katterat stasjon	X	Ofofbanen
Kjosfossen stasjon	X	Flåmbanen
Kongsvoll stasjon	X	Kodet inn, men lang veitilgang
Kopperå stasjon	X	Kodet inn, men ikke tilgang til vei
Kvitfjell stasjon	X	Hunderfossen/Kvitfjell
Lunden stasjon	X	Flåmbanen
Movatn stasjon	X	Kodet inn, men ikke tilgang til vei
Narvik stasjon	X	Ofofbanen
Reinunga stasjon	X	Flåmsbanen
Rombak stasjon	X	Ofofbanen
Straumsnes stasjon	X	Ofofbanen
Svenningdal stasjon	X	Nedlagt 2014
Søsterbekk stasjon	X	Ofofbanen
Trollhättan stasjon	X	Sverige
Tverlandet	X	Ny 2015
Vatnahalsen stasjon	X	Flåmsbanen
Øxnered stasjon	X	Sverige

Vedlegg 2: Døgnfordelinger over avreisetidspunkt med ulike transportmidler

For å illustrere hvordan døgnfordelingene ser ut for ulike transportmidler og reisehensikter vises fordelinger over avreisetidspunkt i Figur V2 1. Merk at disse fordelingene er estimert på bakgrunn av varierende antall observasjoner, og usikkerheten er derfor ikke den samme for hver fordeling. I Tabell V2 1 vises en oversikt over antall observasjoner for kombinasjoner av transportmiddel og reisehensikt. Det er verdt å merke seg at det er få observasjoner for tjenestereiser. Det er ikke gjort noen form for korreksjon for utvalgsskjevhet, derfor avviker transportmiddelfordelingen i Tabell V2 1 fra den i Hjorthol mfl. (2014). Tallene er kun tatt med for å illustrere hvor mange reiser som er registrert i de ulike kategoriene.



Figur V2 1: Døgfordelinger over avreisetidspunkt for ulike transportmidler og reisebensi- basert på RVU 2013–14.

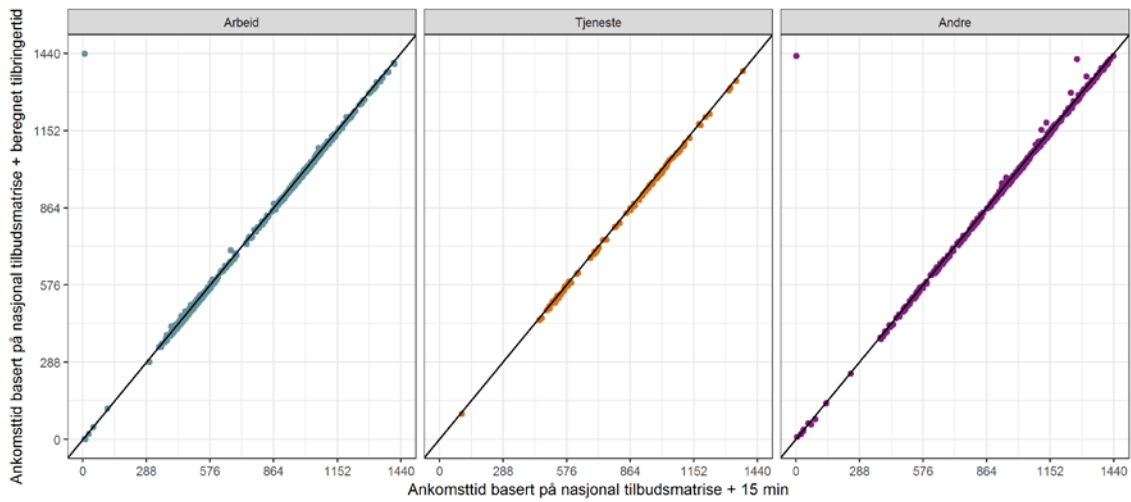
Tabell V2 1: Antall arbeidsreiser, tjenestereiser og andre reiser med ulike transportmidler. Datakilden er RVU 2013–14. Reiser lørdag, søndag eller i juli måned er ekskludert og ulike transportmiddelkategorier er slått sammen (se Tabell 3). Merk at det ikke er korrigert for utvalgsskjevbet, slik at andelene presentert i denne tabellen representerer ikke den nasjonale transportmiddelfordelingen. RVU består av et nasjonalt utvalg og regionale tilleggsutvalg, det må derfor korrigeres for ulik uttrekks sannsynlighet i ulike geografiske områder for å kunne gi estimater på populasjonsnivå. Her er tallene kun med for å illustrere det varierende antallet observasjoner.

	Arbeid	Tjeneste	Andre
Bil	22 854 (58,2%)	4 517 (72,4%)	66 670 (67,0%)
Til fots	5 503 (14,0%)	729 (11,7%)	23 059 (23,2%)
Buss	4 613 (11,7%)	264 (4,2%)	3 291 (3,3%)
Sykkel	3 259 (8,3%)	148 (2,4%)	3 735 (3,8%)
Tog	1 423 (3,6%)	155 (2,5%)	615 (0,6%)
Trikk/bane	911 (2,3%)	98 (1,6%)	954 (1,0%)
MC/moped	284 (0,7%)	16 (0,3%)	349 (0,4%)
Ferge	219 (0,6%)	35 (0,6%)	213 (0,2%)
Fly	135 (0,3%)	183 (2,9%)	287 (0,3%)
Taxi	85 (0,2%)	96 (1,5%)	326 (0,3%)
Totalt	39 286	6 241	99 499

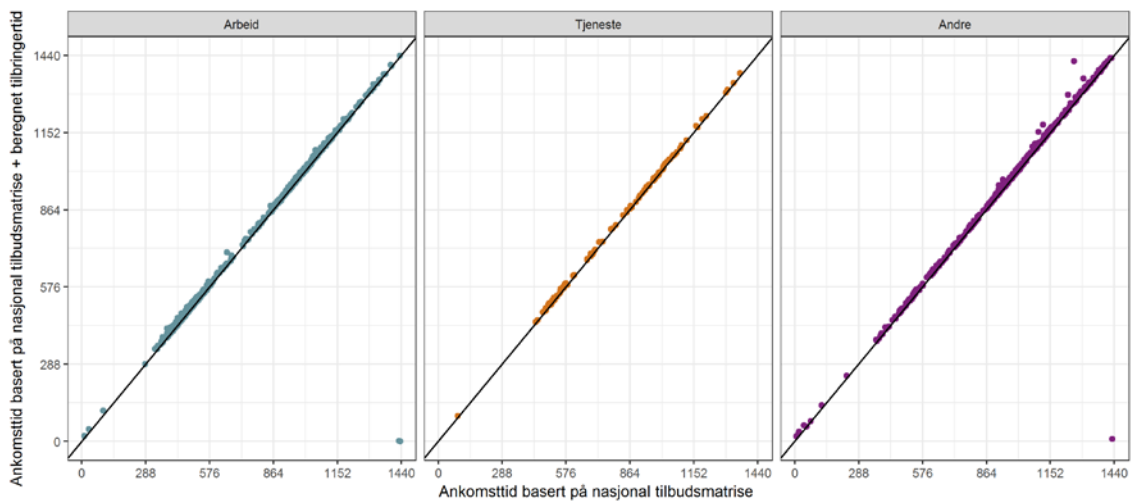
Tabell V2 2: Oversikt over sammenslåtte transportmiddelkategorier. Kategoriene annen båt/fritidsbåt/småbåt, traktor, snøscooter, og annet er ekskludert her.

Transportmiddel	Transportmiddel i RVU
Til fots	
Sykkel	Sykkel, el-sykkel
MC/moped	Moped, motorsykkel
Bil	Bil (fører), bil (passasjer)
Taxi	
Buss	Buss/rutebil/ekspresbuss i rute, turbuss/chartret buss
Trikk/bane	Trikk/bybane, t-bane
Tog	
Fly	Rutefly, charterfly
Ferge	Ferge (som bilfører/passasjer), ferge (som fotgjenger/busspassasjer), rutebåt

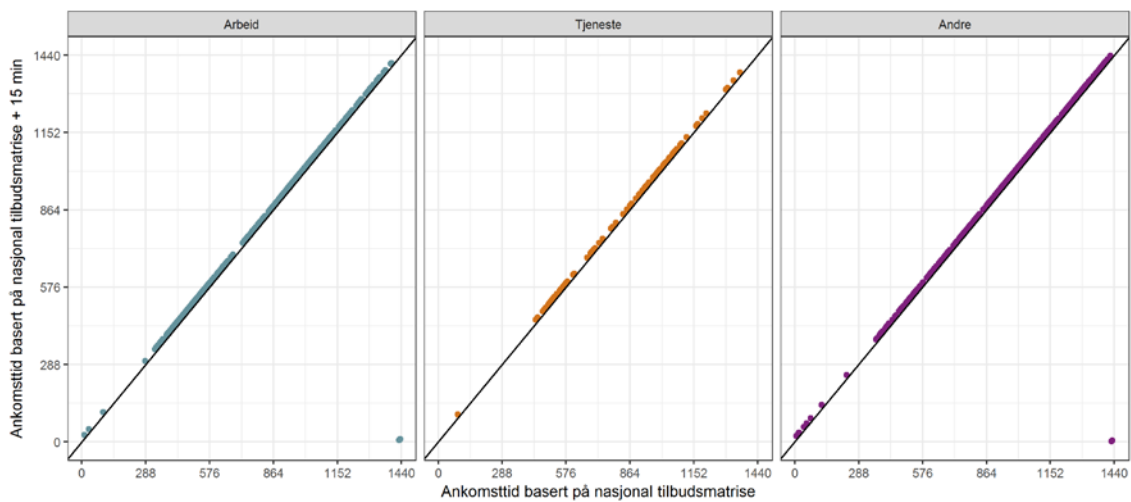
Vedlegg 3: Sammenligning av ankomsttidsvariabler



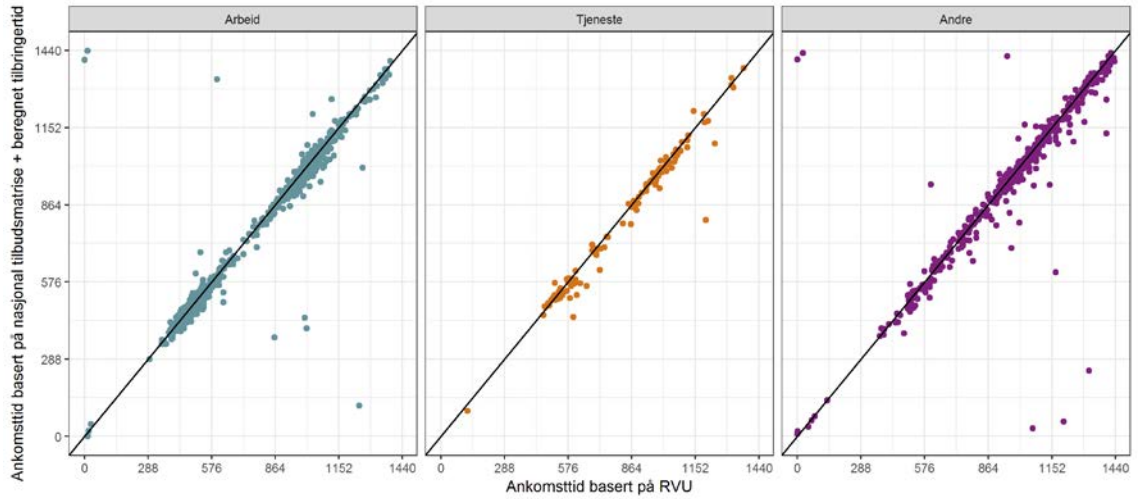
Hensikt ● Arbeid ● Tjeneste ● Andre



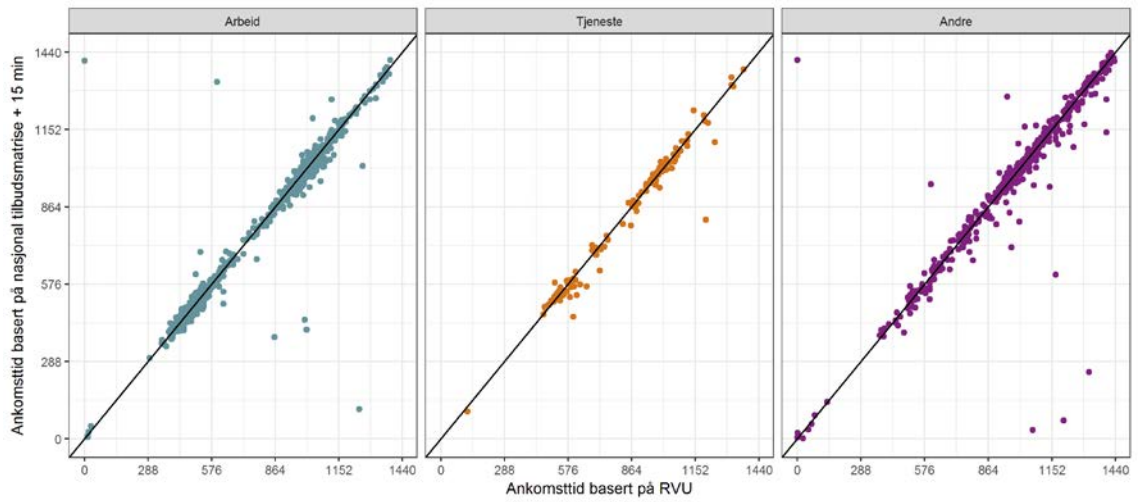
Hensikt ● Arbeid ● Tjeneste ● Andre



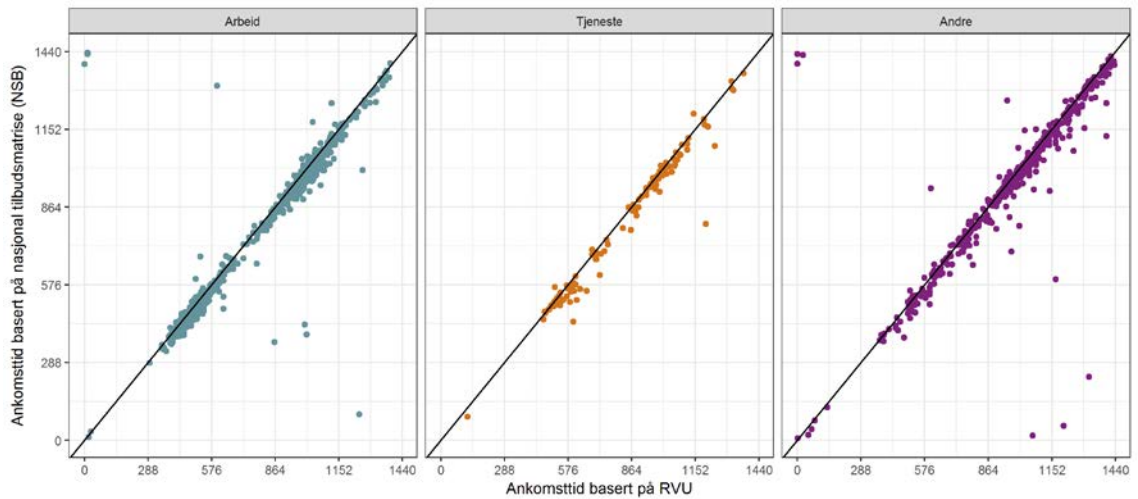
Hensikt ● Arbeid ● Tjeneste ● Andre



Hensikt ● Arbeid ● Tjeneste ● Andre



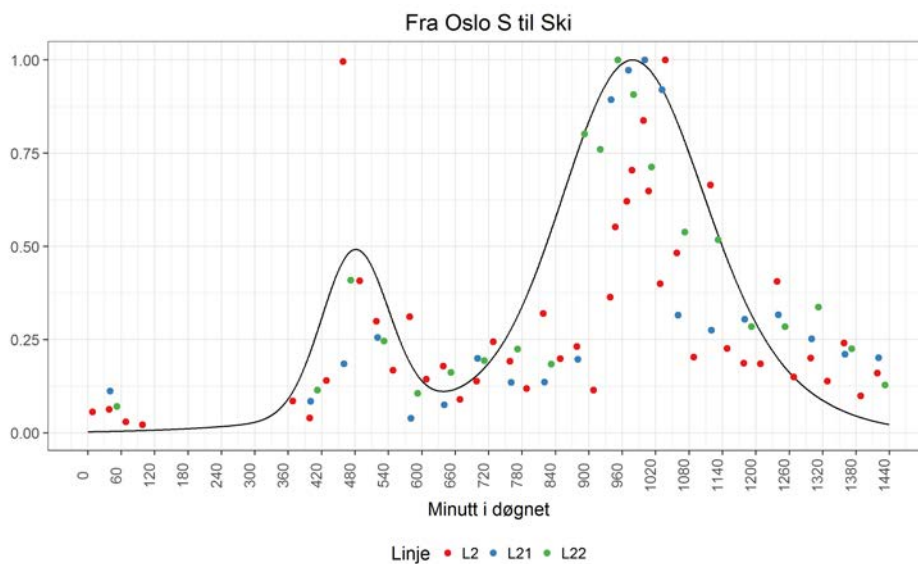
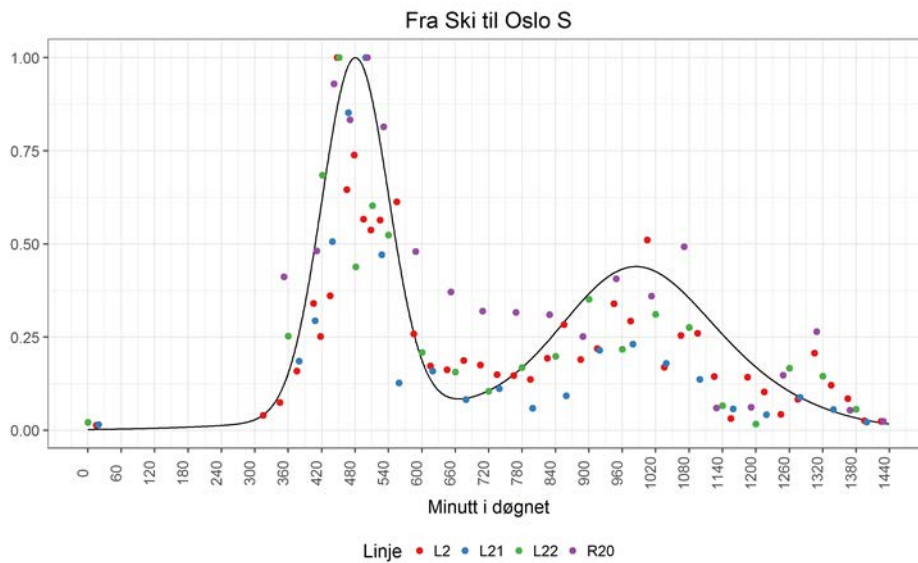
Hensikt ● Arbeid ● Tjeneste ● Andre

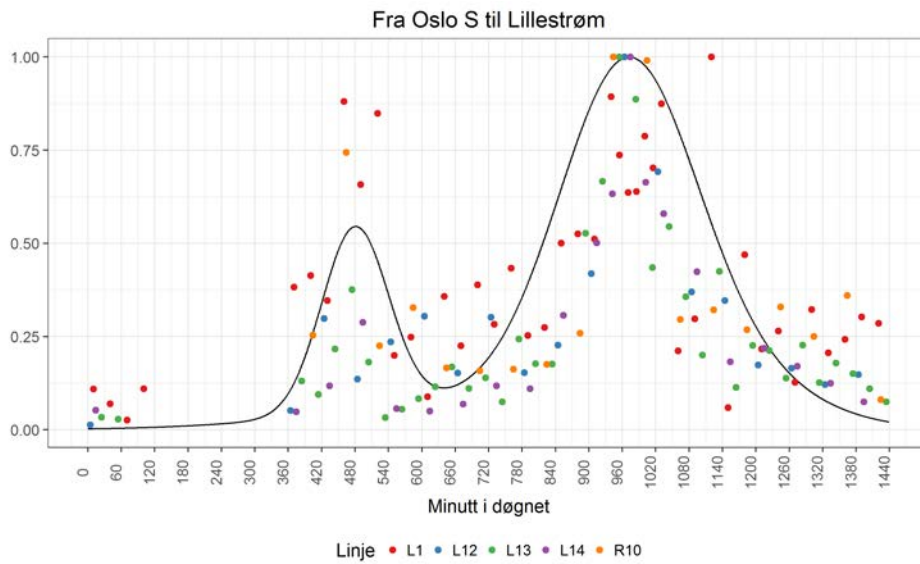
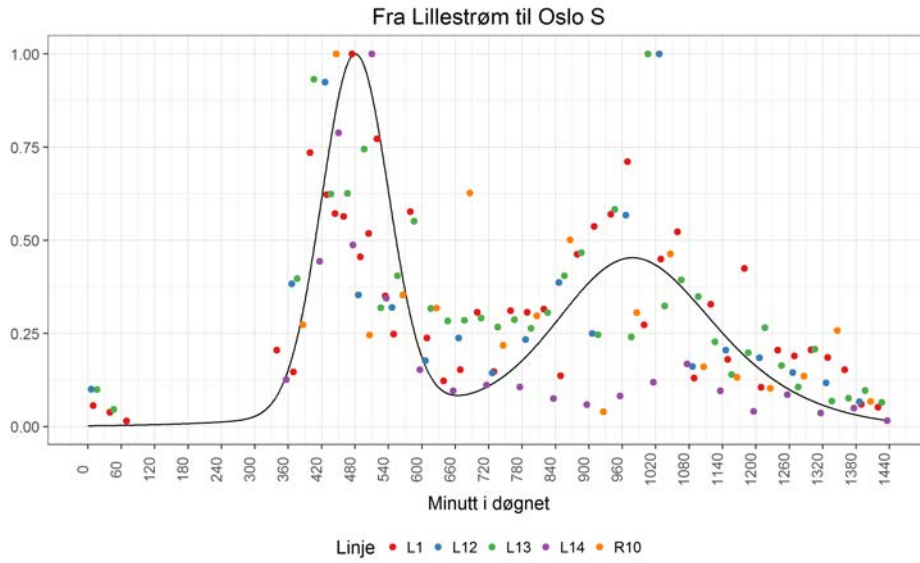


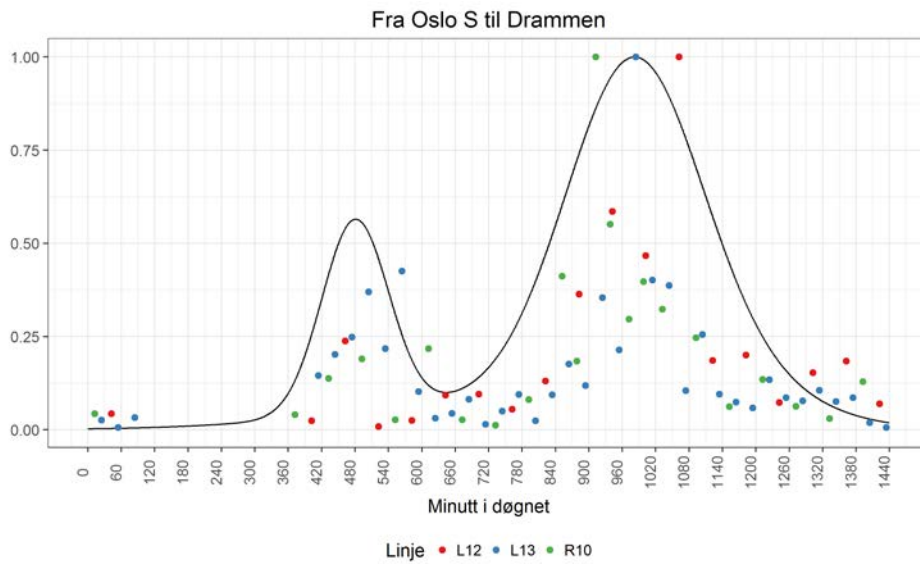
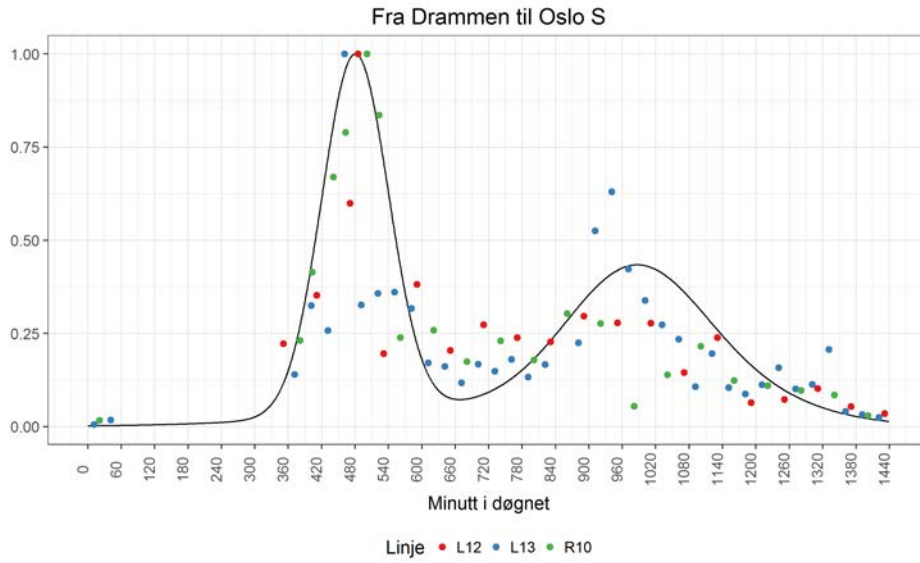
Hensikt ● Arbeid ● Tjeneste ● Andre

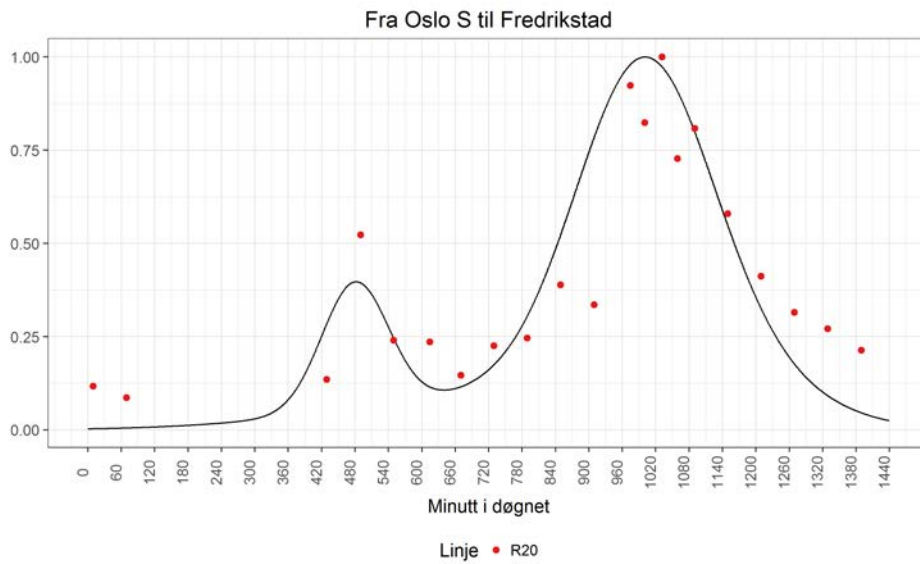
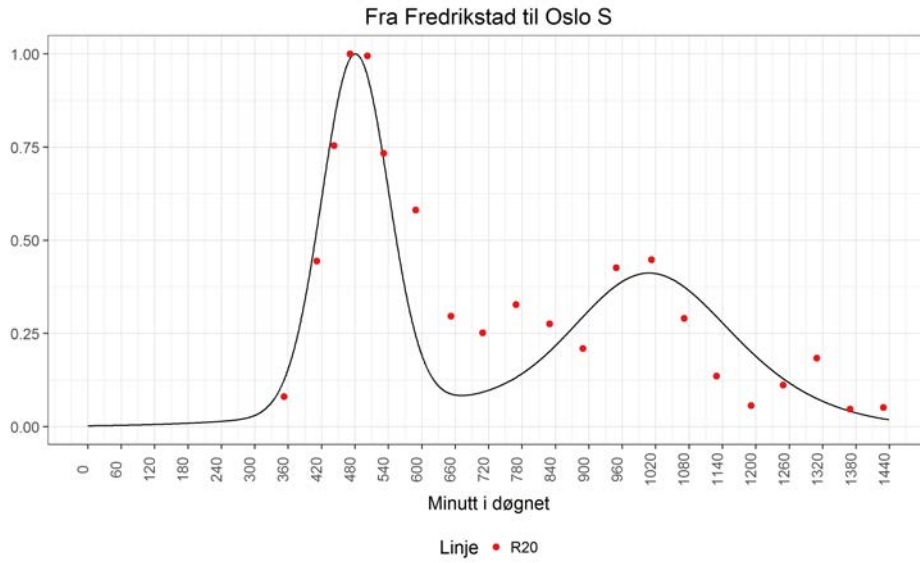
Vedlegg 4: Validering av døgnfordelinger for utvalgte stasjonsrelasjoner

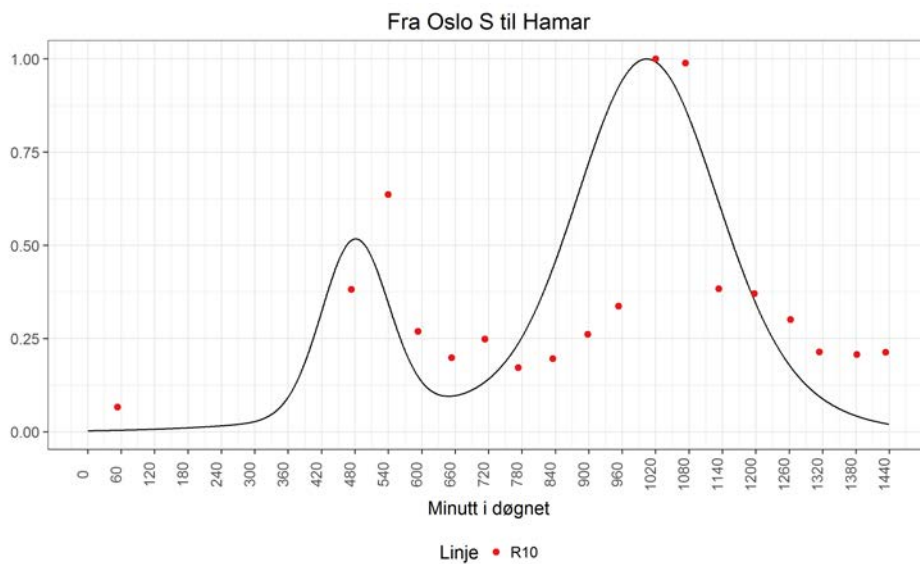
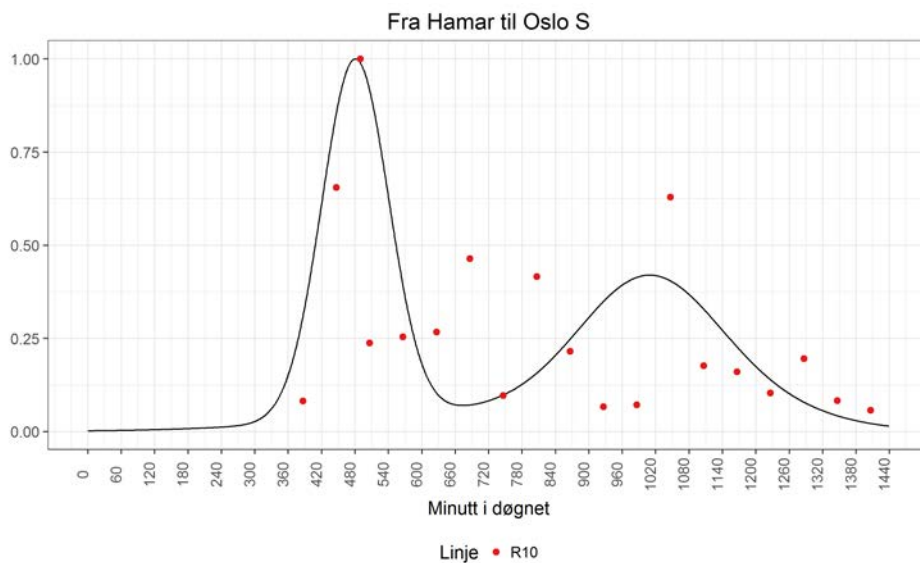
I figurene under sammenlignes døgnfordelingen for alle reisehensikter med transformerte (fra snitt- til OD-format) passasjertellinger for utvalgte stasjonsrelasjoner. Den heltrukne linjen er den estimerte døgnfordelingen og punktene er passasjertellinger fra aktuelle toglinjer. Fordelingen og passasjertellingene er normert ved å dele på respektive maksimumsverdier for visualiseringens skyld.

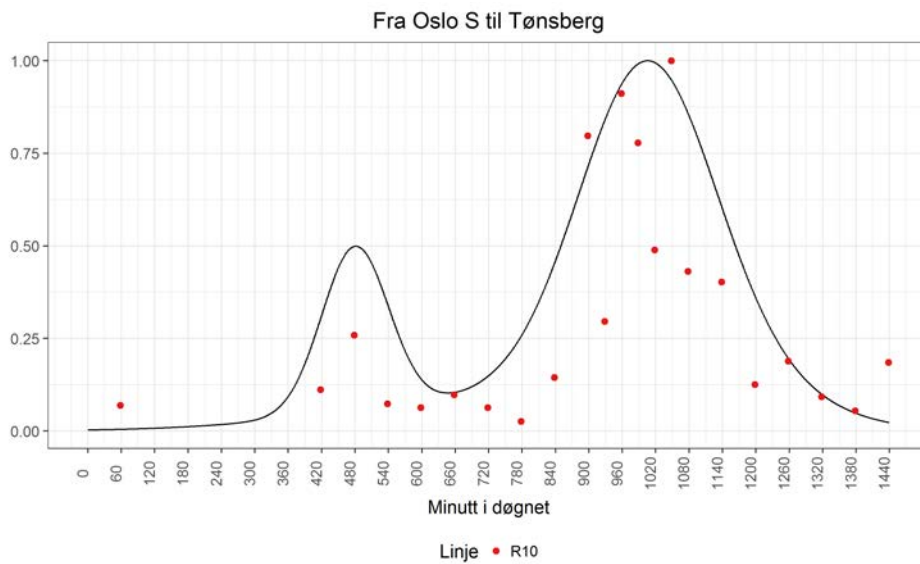
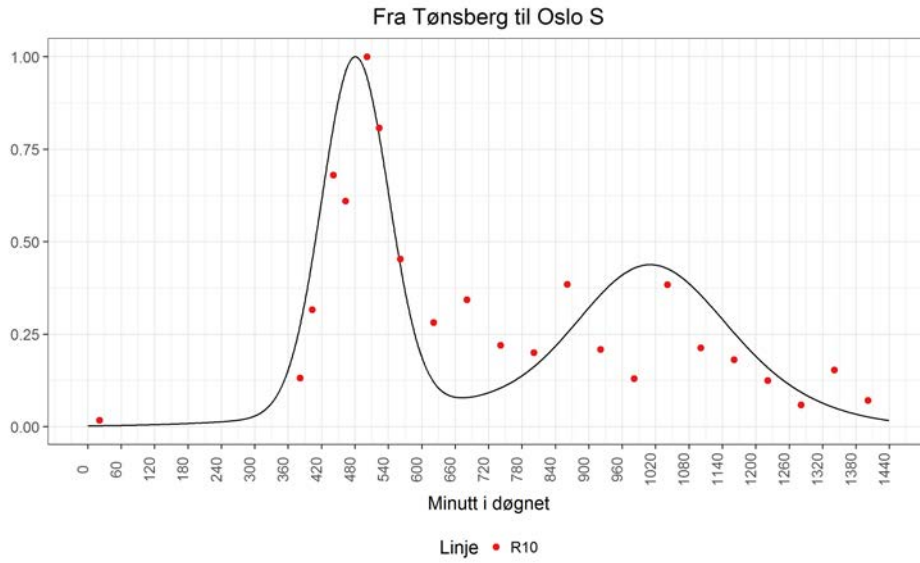












Transportøkonomisk institutt (TØI) Stiftelsen Norsk senter for samferdselsforskning

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 tverrfaglig miljø med rundt 70 høyt spesialiserte forskere.

Instituttet utgir tidsskriftet Samferdsel med 10 nummer i året 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
Gaustadalléen 21
NO-0349 Oslo

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