



# Utvikling av den nasjonale persontransportmodellen i fase 5

## Del B: Estimering av modeller

Tom N Hamre  
Jens Rekdal  
Odd I Larsen

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:** Utvikling av den nasjonale persontransportmodellen i fase 5. Del B: Estimering av modeller

**Forfatter(e):** Tom N Hamre; Jens Rekdal; Odd I Larsen

TØI rapport 606/2002  
Oslo, 2002-11  
96 sider  
ISBN 82-480-0297-7  
ISSN 0802-0175

**Finansieringskilde:**

Arbeidsgruppe for Transportanalyser - Nasjonal transportplan (Vegdirektoratet, Jernbaneverket, Luftfartsverket og Kystverket)

**Prosjekt:** 2706 Estimering av langdistansemodeller i NTM5

**Prosjektleder:** Ingunn Stangeby

**Kvalitetsansvarlig:** Odd I Larsen

**Emneord:**

Logitmodeller; estimering; reisemiddel; reisemål; reisefrekvens; reiseetterspørsel; reiseformål

**Sammendrag:**

NTM5 er en videreutvikling av modellene for lange reiser fra NTM4. Modeller for valg av reisemiddel, reisemål og reisefrekvens er reformulert og estimert på RVU 97/98. Denne rapporten beskriver modellstrukturen og segmenteringsprinsippene som systemet NTM5 baserer seg på, og dokumenterer også de ulike estimeringsresultatene.

---

**Title:** Development of the Norwegian national transport model - Version 5. Estimation of models

**Author(s):** Tom N Hamre; Jens Rekdal; Odd I Larsen

TØI report 606/2002  
Oslo: 2002-11  
96 pages  
82-480-0297-7  
ISSN 0802-0175

**Financed by:**

Working Group for Transport Analysis, Norwegian National Transport Plan

**Project:** 2706 Estimation of models for longdistance trips in NTM 5

**Project manager:** Ingunn Stangeby

**Quality manager:** Odd I Larsen

**Key words:**

Logit models; esimation; travel mode; destination; travel frequency; travel demand modelling; travel purpose

**Summary:**

NTM5 represents a further development of the models for long distance trips in the previous model version (NTM4). The mode choice, destination and trip frequency models are now reformulated, and estimated on the national travel survey of 1997/1998. This report describes the structure of the model system and the principles of aggregation, and also documents the estimation results for each of the sub-models of NTM5.

**Language of report:** Norwegian

---

Rapporten kan bestilles fra:  
Transportøkonomisk institutt, Biblioteket  
Gaustadalleen 21, 0349 Oslo  
Telefon 22 57 38 00 - [www.toi.no](http://www.toi.no)

---

The report can be ordered from:  
Institute of Transport Economics, The library  
Gaustadalleen 21, NO 0349 Oslo, Norway  
Telephone +47 22 57 38 00 - [www.toi.no](http://www.toi.no)

---

# Forord

På oppdrag fra Arbeidsgruppe for Transportanalyser – Nasjonal Transportplan 2006 - 2015 (tverretattlig gruppe sammensatt av representanter fra Vegdirektoratet, Jernbaneverket, Luftfartsverket og Kystverket) har TØI videreutviklet Den nasjonale persontransportmodellen. Dette har i hovedsak dreiet seg om etablering av nye modeller for lange reiser og har resultert i det vi betegner som NTM 5. Arbeidet har vært inndelt i tre delprosjekter; A: Tilrettelegging av data (reisevaneundersøkelse, nettverksdata og sonedata), B: Estimering av modeller, C: Implementering, kalibrering og verifisering.

Denne rapporten dokumenterer valgene rundt modellstruktur, anvendelsen av de ulike datakildene og arbeidet med spesifisering og estimering av modeller i delprosjekt B. Teksten bygger på ulike arbeidsdokumenter fra prosjektgjennomføringen.

RAND Europe, ved Andrew Daly, Charlene Rohr og James Fox, har fungert som ekstern konsulent på estimeringsopplegg og spesielle data- og estimeringstekniske problemstillinger.

Arbeidet er gjennomført av forskerne Jens Rekdal (prosjektleder), Odd I Larsen og Tom N. Hamre. Professor Odd I. Larsen, Høgskolen i Molde/Møreforskning har hatt ansvaret for kvalitetssikringen. Sekretær Kari Tangen har stått for den endelige tekstbehandlingen.

Oslo, november 2002  
Transportøkonomisk institutt

*Knut Østmoe*  
instituttssjef

*Ingunn Stangeby*  
avdelingsleder



# Innhold

## Sammendrag

<b>1 Bakgrunn og innledning .....</b>	<b>1</b>
<b>2 Overordnet modellstruktur.....</b>	<b>2</b>
2.1 Soner.....	3
2.2 Transportmåter.....	4
2.3 Reiseformål .....	4
<b>3 Data.....</b>	<b>6</b>
3.1 Reisevaneundersøkelsen 1997/1998 (RVU-97/98) .....	6
3.1.1 Generelt.....	6
3.1.2 Manglende opplysninger om inntekt .....	9
3.1.3 Data for reiser kortere enn 100 kilometer .....	14
3.2 Nettverksdata .....	16
3.2.1 Nye nasjonale nettverk.....	16
3.2.2 Generering av nettverksdata.....	17
3.2.3 Reisekostnader for kollektivtransport.....	19
3.3 Sonedata .....	19
<b>4 Modeller for biltilgang og førerkort .....</b>	<b>21</b>
4.1 Segmentering i persontransportmodellen .....	21
4.2 Hvorfor segmentere?.....	21
4.3 Segmentering relatert til bilhold og førerkort.....	22
4.4 Befolkning i nye soner .....	25
4.5 Generelle prinsipper for segmentering.....	26
4.5.1 Generelt.....	26
4.5.2 Segmentering med kontinuerlige variable.....	27
4.6 Segmentering med henblikk på førerkortinnehav og biltilgang.....	29
4.6.1 Beregning av antall i hvert segment.....	29
4.6.2 Personer i husholdninger med én person over 18 (20) år .....	30
4.6.3 Personer i husholdninger med to personer over 18 (20) år .....	34
4.6.4 Personer i husholdninger med 3+ personer over 18 (20) år.....	37
4.7 Modellene .....	40
4.7.1 Husholdninger med én voksen person .....	40
4.7.2 Husholdninger med to voksne personer eller 3+ voksne personer .....	40
4.7.3 Kalibrering/justering av modellene .....	41
4.8 Beregning av antall husholdninger i de tre kategorier .....	43
4.9 Referanser.....	47
<b>Vedlegg til kapittel 4.....</b>	<b>48</b>
<b>5 Kohortmodeller for førerkortinnehav .....</b>	<b>53</b>
5.1 Tidsrekkemodell for utvikling i førerkortandeler .....	53
5.2 Nye tidsrekkemodeller.....	53
5.2.1 Datatilgang tidsrekker .....	53
5.2.2 Anskaffelse av førerkort for de yngste aldersgruppene .....	54
5.2.3 Utvikling i førerkortinnehav for de resterende aldersgrupper .....	57

5.2.4 Prognoser for førerkortandeler.....	58
<b>6 Estimering av modeller for valg av transportmåte og destinasjon .....</b>	<b>65</b>
6.1 Data .....	65
6.1.1 Behandling av reisetidskomponenter for kollektivtransport .....	65
6.1.2 Reisekostnader .....	66
6.2 Modeller for valg av reisemiddel og destinasjon .....	67
6.2.1 Modellformulering.....	68
6.2.2 Transportkvalitetsvariable .....	69
6.2.3 Attraksjonsvariable.....	69
6.2.4 Estimeringsresultater .....	69
6.2.5 De korteste lange reisene: WESML.....	72
6.2.6 Estimering med NTPL-soner.....	72
<b>7 Modeller for valg av reisefrekvens .....</b>	<b>73</b>
7.1 Alternative modellformuleringer .....	73
7.2 Valg av reisefrekvens: binomiske logitmodeller.....	74
7.2.1 Modeller for arbeids/tjenestereiser.....	76
7.2.2 Modeller for besøksreiser .....	77
7.2.3 Modeller for fritidsreiser .....	77
7.2.4 Modeller for andre private reiser .....	77
7.3 Skalering og segmenttabeller .....	78
<b>Vedlegg til kapittel 7 .....</b>	<b>81</b>
Valg av reisefrekvens: poissonmodeller .....	81
Fritidsreiser.....	81
Tjeneste- og arbeidsreiser .....	85
Besøksreiser .....	90
Andre private reiser.....	92
Sammenlikning av alternative sannsynlighetsmodeller .....	94

# Sammendrag:Utvikling av den nasjonale persontransportmodellen i fase 5

## Del B: Estimering av modeller

Utviklingen av Den nasjonale persontransportmodellen i fase 5 er inndelt i følgende tre utviklingsfaser:

1. Bearbeiding og tilrettelegging av data for lange reiser fra RVU-97/98, sonedata og nettverksdata
2. Estimering av modeller (for lange reiser, førerkortinnhav og bilhold) på nye data (RVU-97/87, sonedata, nettverksdata)
3. Implementering av nyestimerte modeller (og gamle kortdistansemodeller) i ny kildekode med forbedret brukergrensesnitt

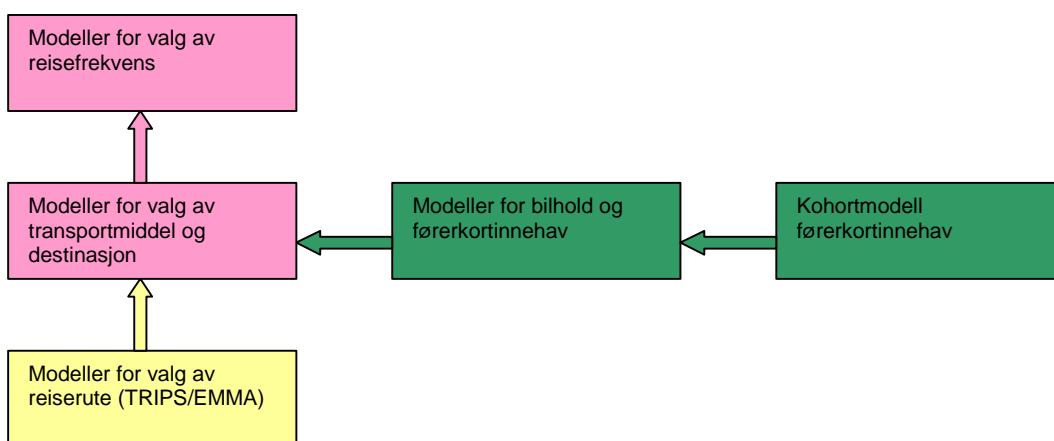
I dette dokumentet dokumenteres del B av arbeidet samt de deler av del A som foreløpig ikke er dokumentert. I dette arbeidet er det estimert modeller for valg av transportmiddel og destinasjon (multinomisk logit og strukturert multinomisk logit), for valg av reisefrekvens (binomisk vektet logit), og for førerkortinnhav og bilhold (multinomisk logit). I tillegg er det dokumentert en prosedyre for beregninger av fremtidig førerkortinnhav (kohortmodell). Den

generelle strukturen i det system av modeller som skal settes sammen i Del C av prosjektet er vist i Figur S.1.

I prinsippet viser figuren en tradisjonell 4-trinnsmodell, men med simultan estimering av modeller for valg av transportmiddel og destinasjon, og supplert med modeller for førerkortinnhav og bilhold.

Reisetider og reisekostnader beregnes i nettverksmodellene. Disse inngår sammen med prognoser for bilhold og førerkortinnhav i modellene for valg av destinasjon og transportmiddel. Bilholdsprognoser og fremskrivninger av førerkortinnhav er bare nødvendig dersom det skal gjøres langsiktige prognoser.

*Kohortmodellene* sørger for at man i prognosene får tatt hensyn til eksempelvis at dagens 40-50-åringer om 20 år vil ha vesentlig høyere førerkortinnhav enn dagens pensjonister. Kohortmodellene gir førerkortandeler fordelt på kjønn og alder for en rekke fremtidsår (2005, 2010, 2015, 2020, osv). Disse dataene benyttes til kalibreringen av førerkort/bilholdsmodellene.



Figur.S.1: Generell struktur modellsystem for lange reiser

*Fører kort/bilholdsmodellene* er tverrsnittsmønstre som i utgangspunktet gjenspeiler det førerkortinnehaver man hadde i 1997/98 når RVU ble gjennomført. Når modellene skal benyttes på fremtidsår vil ikke disse modellene ivareta kohorteffektene i førerkortinnehaver. Dataene fra kohortmodellene benyttes derfor til å kalibrere førerkortinnehaver for et gitt antall fremtidsår, slik at kohorteffektene ivaretas.

Fører kort/bilholdsmodellene ivaretar endringer i inntekt og befolknings sammensetning og predikerer biltilgjengeligheten i fremtidsårene. Dette er en variabel som er svært viktig for transportmiddelvalget i modellene for valg av transportmiddel og destinasjon.

Andre viktige variable i *modellene for valg av transportmiddel og destinasjon* er reisetider og reisekostnader med de ulike transportmidlene (fra nettverksmodellene), og diverse attraksjonsstørrelser (befolkning, arbeidsplasser, med mer) i de ulike tilgjengelige sonene.

I *modellene for valg av reisefrekvens* inngår den såkalte logsummen fra modellene for valg av transportmiddel og destinasjon (logsummen er et uttrykk for tilgjengelighet til destinasjoner med de ulike transportmåter, og attraktiviteten ved ulike destinasjoner). Andre viktige variable er alder, kjønn, inntekt, familietype, og andre kjennetegn.

I tidligere versjoner av den nasjonale transportmodellen er kommuner (435 stk) benyttet som geografisk enhet. I NTM5 er soneinndelingen forfinet og modellen omfatter nå 1428 soner (NTPL-soner). Denne soneinndelingen er basert på SSBs tettstedsdefinisjon.

Modellene for lange reiser omfatter følgende fem transportmåter:

1. Bil (passasjer og fører)
2. Buss (kun rutegående busser)
3. Tog
4. Fly (kun rutefly)
5. Båt (kun rutegående båter)

Denne inndelingen dekker 97 % av de lange reisene som er gjennomført i RVU-97/98. Det er egne modeller for følgende 4 reiseformål:

- *Arbeids-/tjenestereiser.* Omfatter reiser til/fra arbeid betalt av arbeidsgiver, og alle tjenestereiser
- *Besøksreiser.* Privat besøk hos venner og familie, etc.
- *Fritidsreiser.* Reiser til fornøyer/underholdning, organiserte fritidsaktiviteter, ferie og andre fritidsreiser

- *Andre private reiser.* Alle andre reiseformål, inkl reiser til/fra arbeid betalt av trafikanten selv

Til estimeringen av de nye langdistansmodellene trengs tre typer data som fortrinnsvis må representere samme tidsperiode:

- Reisevanedata
- Data for transporttilbud
- Data for soneinnhold

*Reisevanedata* fremskaffes gjennom intervjuundersøkelser hvor et representativt utvalg av befolkningen stilles en rekke spørsmål om sine reiseaktiviteter og en del bakgrunnsinformasjon om seg selv og sin husholdning. Modellestimeringen er basert på RVU97/98 som er bearbeidet og tilrettelagt spesielt for dette prosjektet. RVU97/98 omfatter data rapportert av 8800 respondenter som til sammen har reist vel 4800 rundturer som kunne benyttes til estimering av modeller for valg av transportmiddel og destinasjon. Datamaterialet i RVU-97/98 har imidlertid noen svakheter som har krevd spesiell oppmerksomhet i dette prosjektet:

- Lav svarprosent på spørsmål om inntekt
- Skjevt frafall
- Stedfesting på kommunenivå

Når det gjelder inntekt har vi sett oss nødt til å estimere egne modeller som beregner inntekt for personer hvor denne opplysningen mangler. Gjennom et samarbeid med kompetansepersonell ved RAND Europe har vi så langt som mulig forsøkt å ta hensyn til at frafallet har vært skjevt og at vi i stedet for kommuner som geografisk enhet ønsket å benytte NTPL-soneinndelingen.

*Data for transporttilbudet* fremskaffes gjennom såkalte nettverksmodeller som har en geografisk dimensjon representert ved en geografisk soneinndeling og nettverk for alle relevante transportmåter som knytter disse sonene sammen. Det nasjonale nettverket er foreløpig laget til programpakken EMMA og består av et vegnett generert med GISNETT og kollektivruter med tilhørende infrastruktur generert med KOLLNETT (Hamre 2001). I nettverksmodellene kan man gjøre beregninger for å ta frem reisetider og reisekostnader mellom alle geografiske områder definert gjennom soneinndelingen, med alle transportmåter som er tilgjengelig.

*Data for soneinnhold* fungerer som skapende og attraherende elementer når det gjelder reiseaktiviteter. Det er befolkningsstørrelsene i de ulike geografiske



områdene som skaper trafikken. Befolkningen deles blant annet inn etter alder, kjønn, og yrkesaktivitet. Arbeidsplasser fordelt etter næring, hotellsenger, hytter og fritidshus, med mer, fungerer som *proxy-variable* på elementer som trekker til seg turer. I tidligere versjoner av NTM har vi bare hatt tilgang på sonedata på kommunenivå. I foreliggende arbeide er det lagt stor vekt på å skaffe denne type data til veie på grunnkrets-nivå. Dataene er hovedsakelig innhentet fra Statistisk Sentralbyrå.

I avsnittene under kommenteres hver enkelt delmodell representert med en egen boks i Figur S.1.

Prinsipper for segmentering i persontransportmodellen er beskrevet i kapittel 4. Metoden som benyttes i NTM5 innebærer å anvende modeller for *bilhold og førerkortinnehav*, for å utføre en videre inndeling (segmentering) av de ulike kjønn- og alderssegmentene i hver sone – etter ulike grader av biltilgjengelighet. Dette er egne logitmodeller (en for hver av tre husholdskategorier), som beregner sannsynlighetene for at et segment av befolkningen i en gitt sone skal falle i hver av 5 ulike kategorier med hensyn til tilgang til bil. Tilgang til bil defineres gjennom førerkortinnehav (for personen selv eller medlemmer av husholdet) og antall biler i husholdet. For eksempel vil en person uten førerkort ha større “motstand” mot å velge bil i modellene for reisemiddelvalg enn en person med førerkort. Gitt at man har førerkort så vil graden av “konkurranse” om bilen(e) i et hushold være avgjørende for reisemiddelvalget (det vil si forholdet mellom antall førerkort og antall biler i et hushold).

Kategoriseringen etter fem ulike grader av biltilgjengelighet bygges inn i modellene for valg av reisemiddel ved at det estimeres dummyvariable for disse i nyttefunksjonene for bil (avsnitt 6.2).

Segmenteringsmodellene for bilhold og førerkortinnehav er imidlertid estimert på et tverrsnittsmateriale. RVU-97/98 gir bare en oversikt over hvordan situasjonen var når det gjelder befolkningens egne transportressurser, var i den tidsperioden undersøkelser ble gjennomført. I førerkortinnehavet er det en betydelig kohorteffekt. Denne effekten kan eksemplifiseres ved å se på førerkortinnehavet blant dagens pensjonister og sammenlikne dette med førerkortinnehavet blant de som var pensjonister for eksempel i 1980. Førerkortinnehavet blant menn i aldersgruppen 70-74 år i 1980 var ca 45 %. I den samme aldersgruppen har i dag ca 85 % av mennene førerkort.

For å ta hensyn til denne effekten er det laget en prosedyre som fremskriver førerkortandelene i tid.

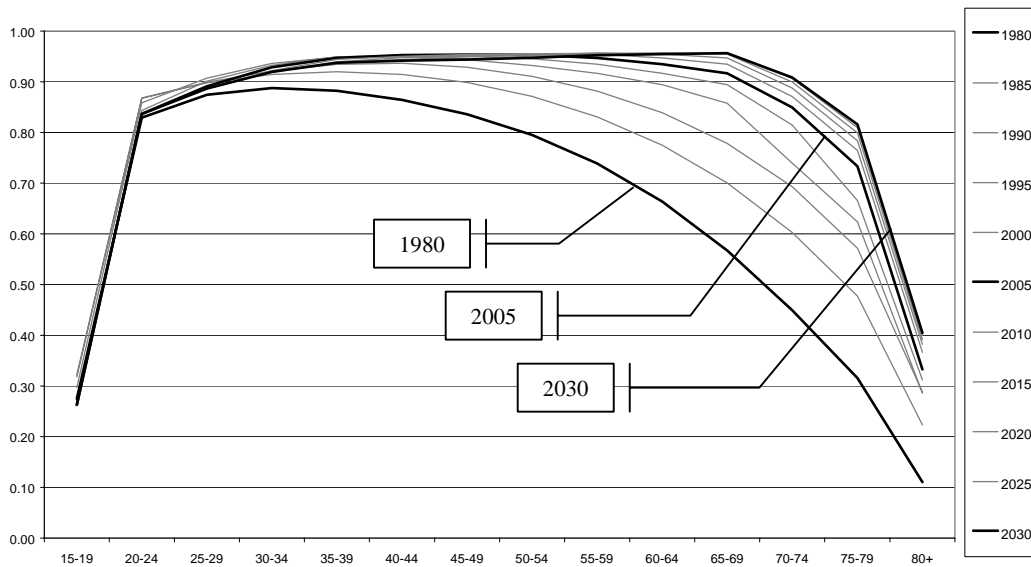
Prosedyren, som vi har kalt *kohortmodeller for førerkortinnehav*, er beskrevet i kapittel 5. Ved hjelp av denne prosedyren har vi beregnet fremtidige førerkortandeler i fem års intervaller frem til år 2030. Disse andelene skal i implementeringen av modellene for bilhold og førerkortinnehav benyttes som kalibreringsgrunnlag for å tilpasse førerkortinnehavet til tidseffekten. Tanken er at mens kohortmodellene tar seg av den rene tidseffekten, vil segmenteringsmodellene ta seg av de effekter som bl.a. skyldes endret befolkningsstruktur og den generelle økonomiske utviklingen.

Figur S.2 og Figur S.3 viser observerte (frem til år 2000) og fremskrevne (fra 2005 til 2030) førerkortandeler etter aldersgruppe for hhv menn og kvinner. Vi ser at mens det skjer relativt lite blant menn mellom 2005 og 2030, er effekten betydelig blant kvinner. I følge beregningene vil kvinners førerkortinnehav i år 2030 være svært likt mennenes. Det er grunn til å peke på at dette skyldes en ren tidseffekt som ikke kan påvirkes av samferdselspolitiske virkemidler. I perioden vil det altså skje store endringer i befolkningens egne transportressurser, og spesielt blant eldre kvinner. Dette vil naturligvis påvirke fremtidens transportomfang i relativt betydelig grad, og dette vil fanges opp i prognosene med NTM5.

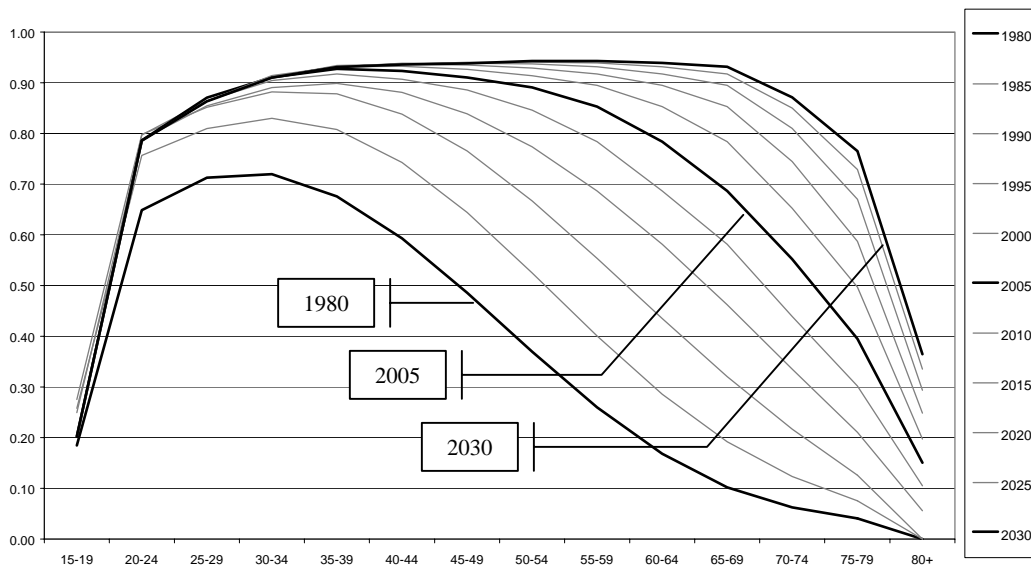
*Modellene for valg av transportmåte (reisemiddel) og destinasjon (reisemål)* er estimert på basis av de reisetidskomponentene og kostnadene som beregnes i nettverksmodellen EMMA, samt ulike sonedata (befolkning, arbeidsplasser med mer). Estimeringsarbeidet, slik det er dokumentert i kapittel 6, har bestått i å prøve ulike formuleringer av variable, og å finne de formuleringer som viser seg å passe best med data-materialet.

Frekvensen for kollektivtilbudet er nå formulert som kvadratroten av antall avganger pr døgn. Frekvensen beregnes fra total ventetid, noe som betyr at vi for relasjoner med bytter underveis må se på dette som et generalisert mål på frekvensen. En (ønsket) effekt av å bruke kvadratroten av frekvens er dessuten at effekten av en frekvensøkning blir relativt liten når frekvensen er høy fra før, og relativt høy med lave frekvenser i utgangspunktet.

Reisetid (minutter ombord i transportmiddelet) inngår uten transformasjoner i alle modeller, mens reisekostnadene (billettpriser, bom-/fergekostnader og bensin) log-transformeres.



Figur S.2: Prognoser for førerkortinnnehav blant menn



Figur S.3: Prognoser for førerkortinnnehav blant kvinner

For destinasjonsvalget benyttes ulike sonedata som "size"-variable i logitmodellene (attraksjonsvariable). Dette er spesielt befolkning, arbeidsplasser innenfor ulike næringer, hotellsenger og fritidshus – men varierer mellom de ulike delmodellene (reishensiktene). Det er estimert modeller med ulike size-variable, for å finne de som passer best med datamaterialet.

Modellene for valg av reisefrekvens er estimert med alder, kjønn, inntekt, årstid, familietype, geografiske variable og tilgjengelighet som variable. Estimeringsarbeidet og de endelige modeller er

beskrevet i avsnitt 7 og i vedlegg til kapittel 7.

Modellene er estimert som binomiske logitmodeller som er vektet etter hvor mange reiser respondentene i RVU-97/98 har gjennomført i løpet av siste måned.

I utgangspunktet er modellene estimert for å gi en best mulig reproduksjon av de data de er estimert på. Dette innebærer å benytte variable som vi i implementeringen av modellene ikke har tilgang til. For å ta hensyn til dette er modellene re-estimert før implementering. I denne estimeringen er det konstruert såkalte segmentverdier ut fra RVU-97/98 for de variable som

ikke finnes til anvendelse av modellene. Disse segmentverdiene, som egentlig er aggregerte størrelser for variablene fordelt på 7 aldersgrupper og kjønn, er benyttet i stedet for de opprinnelige variable i re-estimeringen. Koeffisientestimaterne fra den opprinnelige estimeringen er benyttet som fastlåste verdier i re-estimeringen, slik at denne egentlig bare gir nye konstantledd og skalafaktorer som tar hensyn til den aggregeringen som er gjort ved bruk av segmentverdier i stedet for de opprinnelige variablene.

Hovedpoenget med prosedyren er at vi får estimert egne konstantledd for hvert befolkningssegment i stedet for et samlet konstantledd for hele modellen. Vår oppfatning er at både verdiene på de nye konstantledd og skalafaktorer er slik at dette ikke nødvendigvis er særlig dårligere enn de opprinnelige modellene, og svært mye bedre enn å estimere modellene uten variable vi ikke har tilgang til i implementeringen.



# 1 Bakgrunn og innledning

Utviklingen av Den nasjonale persontransportmodellen i fase 5 er inndelt i følgende tre utviklingsfaser:

1. Bearbeiding og tilrettelegging av data.
2. Estimering av modeller (for lange reiser, førerkortinnehav og biltilgang) på nye data (RVU-97/87, sonedata, nettverksdata).
3. Implementering av nyestimerte modeller (og gamle kortdistansmodeller) i ny kildekode med forbedret brukergrensesnitt

I dette dokumentet dokumenteres del B av arbeidet, samt de deler av del A som foreløpig ikke er dokumentert. Modellene som er estimert er ikke de endelige modeller. Erfaringsvis vil man, i løpet av implementeringen av et nytt modellsystem, oppdage at en del forhold kunne vært gjort annerledes. Ikke minst gjelder dette segmenteringen i modellsystemet. I estimeringsprosessen er man mest opptatt av å lage gode modeller som gjenskerer og på andre måter passer best til de

data man estimerer på . Ofte vil man imidlertid spare en del beregningstid ved å gjennomføre en noenlunde ensartet segmentering mellom modellene. Dette arbeidet gjenstår i del C.

I dokumentet beskrives først den modellstruktur vi legger opp til i den nye langdistansedelen av den nasjonale transportmodellen. Deretter oppsummeres det arbeidet som er gjort med å tilrettelegge data til estimeringen. I kapittel 4 dokumenteres tverrsnittmodellene for biltilgang og førerkortinnehav, og behandlingen av kohorteffektene i førerkortinnehavet behandles i kapittel 5. I kapittel 6 og 7 beskrives foreløpige modeller for hhv transportmiddelvalg/destinasjonsvalg og reisefrekvens. En del relevante vedlegg er til sist tatt med i kapittel 8.

Arbeidet er hovedsakelig gjennomført av Tom N. Hamre (TØI), Odd I. Larsen (TØI og Høgskolen i Molde/Møreforskning), Jens Rekdal (Møreforskning, tidligere TØI), Berit Grue (TØI), Ingvil Gjelsvik (TØI) og Frode Voldmo (TØI).

## 2 Overordnet modellstruktur

I dette arbeidet er det estimert modeller for valg av transportmiddel og destinasjon (multinomisk logit og strukturert multinomisk logit), for valg av reise-frekvens (binomisk vektet logit, poisson), og for førerkortinnhav og bilhold (multinomisk logit). I tillegg er det dokumentert en prosedyre for beregninger av fremtidig førerkortinnhav (kohortmodell). Den generelle strukturen i det system av modeller som skal settes sammen i Del C av prosjektet er vist i Figur 2.1.

I prinsippet viser figuren en 4-trinnsmodell, supplert med modeller for førerkortinnhav og biltilgang. Reisetider og reisekostnader beregnes i nettverksmodellene. Disse inngår sammen med prognoser for biltilgang og førerkortinnhav i modellene for valg av destinasjon og transportmiddel.

Biltilgangsprognoser og fremskrivninger av førerkortinnhav er bare nødvendig dersom det skal gjøres langsiktige prognoser.

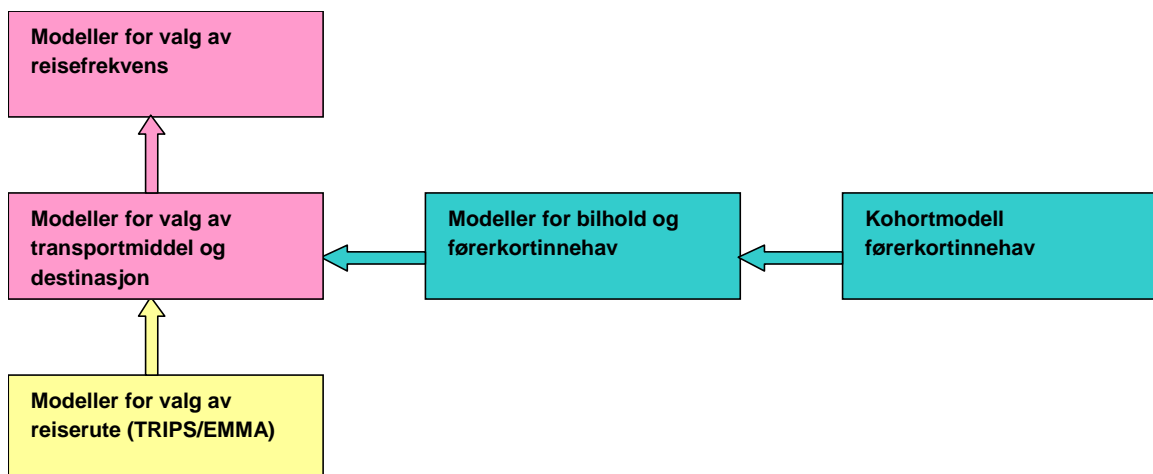
Kohortmodellene sørger for at man i prognosene får tatt hensyn til eksempelvis at dagens 40-50-åringer om 20 år vil ha vesentlig høyere førerkortinnhav enn dagens pensjonister. Kohortmodellene gir førerkortandeler fordelt på kjønn og alder for en rekke fremtidsår (2005, 2010, 2015, 2020, osv.). Disse dataene benyttes til kalibreringen av førerkort/biltilgangsmodellene.

Førerkort/bilholdsmodellene er tverrsnittsmønstre som i utgangspunktet gjenspeiler det førerkortinnhav man hadde i 1997/98 når RVU ble gjennomført. Når modellene skal benyttes på fremtidsår vil ikke disse modellene ivareta kohorteffektene i førerkortinnhavet. Dataene fra kohortmodellene benyttes derfor til å kalibrere førerkortinnhavet for et gitt antall fremtidsår, slik at kohorteffektene ivaretas.

Førerkort/biltilgangsmodellene ivaretar endringer i inntekt og befolknings sammensetning og predikerer biltilgjengeligheten i fremtidsårene. Dette er en variabel som er svært viktig for transportmiddelvalget i modellene for valg av transportmiddel og destinasjon.

Andre viktige variable i modellene for transportmiddelvalg og destinasjon er reisetider og reisekostnader med de ulike transportmidlene (fra nettverksmodellene), og diverse attraksjonsstørrelser (befolkning, arbeidsplasser, m.m.) i de ulike tilgjengelige sonene.

I modellene for valg av reise-frekvens inngår den såkalte logsummen fra modellene for valg av transportmiddel og destinasjon (logsummen er et uttrykk for tilgjengelighet til destinasjoner med de ulike transportmåter, og attraktiviteten ved ulike destinasjoner). Andre viktige variabler er alder, kjønn, inntekt, familietype og andre kjennetegn.



TØI rapport 606/2002

Figur 2.1: Generell struktur modellsystem for lange reiser

## 2.1 Soner

Dagens transportmodell (NTM 4c) benytter kommuner som geografisk enhet (435 stk). Et av hovedmålene med videreutviklingen har vært å innføre en finere soneinndeling enn kommuner. Det er derfor etablert en soneinndeling hvor landet er delt inn i 1428 soner som er ment brukt i den nye nasjonale transportmodellen for lange reiser inntil videre<sup>1</sup>.

I denne soneinndelingen er grunnkretsene brukt som byggesteiner. Fordi en rekke sentrale data oppgis på kommunenivå, er det fortsatt et viktig kriterium at sonene ikke er kommuneoverskridende.

Som et utgangspunkt for avgrensningen av NTPL-sonene, ble grunnkretsene tilordnet det tettsted som ligger nærmest, regnet i kjøretid med bil. På denne måten fikk vi et basismateriale av soner gruppert omkring hvert tettsted.

Med de ulike geografiske, befolknings- og transportmessige strukturer som finnes i Norge, samt hensynet til modelloppbygging og datatilgang, måtte det videre gjøres betydelige justeringer:

Gruppering etter nærmeste tettsted er fraveket i følgende tilfeller :

- Når nærmeste tettsted ikke ligger i kommunen. Kretsen er da tilordnet nærmeste tettsted innen kommunen, eller inngår i en egen sone uten tettsted, dersom flere av nabokretsene har liknende egenskaper.
- Når vegforhold og plassering av befolknings-tyngdepunkt medfører at en krets ikke grenser til resten av sonen. Kretsen innlemmes i en tilgrensende sone, eller mellomliggende kretser tas med i sonen, avhengig av avstander.
- Ved stor geografisk avstand eller når barrierer som f.eks. fjord, elv eller fergestrekning gjør at reisetid til nærmeste tettsted blir uforholdsmessig stor. Grupper av kretser kan i slike tilfeller danne soner uten tettsted.
- Ved mindre justeringer av sonegrenser hvor enkeltkretser med omtrent lik avstand til flere tettsteder er flyttet.
- I tett befolkede områder hvor grunnkretsene rundt byen/tettstedet utgjør et stort område både geografisk og i folketall, er det dannet flere soner av kretsene som tilhører samme tettsted.
- I noen spredtbygde strøk, er kretser som ligger langt unna nærmeste tettsted samlet i soner uten tettsted.

<sup>1</sup> Et av problemene med dette er at man i RVU bare har kjennskap til destinasjonene på kommunenivå.

- I områder med mange små tettsteder eller korte avstander mellom tettstedene, kan nærliggende kretser tilknyttet ulike tettsteder være slått sammen til soner som inneholder mer enn ett tettsted.

Etter disse justeringene har vi et sonesystem bestående av:

- 825 soner dannet rundt spesifisert tettsted. Disse sonene har fått centroide-koordinater tilsvarende næringstyngdepunktet i tettstedet.
- 526 soner uten eget spesifisert tettstedspunkt. Centroider er lagt til befolkningstyngdepunktet i sonens mest befolkede grunnkrets.
- 77 soner som inneholder to eller tre tettsteder. Disse sonene har centroiden i det mest befolkede av tettstedene.

Den første versjonen av NTPL-sonene er basert på en database bygget opp omkring GAB-data og befolkningstall for 1998-grunnkretsene.

Selv når data konsentreres innen et så lite tidsspenn som ett år, oppstår enkelte uoverensstemmelser i grunnkretsangivelser mellom de ulike datakilder. Problemene kan dukke opp der grunnkretser er sammenslått eller splittet i løpet av perioden, og flere versjoner opptrer i datakildene.

I disse tilfellene ble både gammel og ny grunnkretsinnndeling tilordnet NTPL-sonesystemet, for å sikre best mulig koblingsmuligheter mot ulike datasett uavhengig av nyanser i grunnkretsinnndelingen.

Inndelingen av grunnkretsene i Norge er i kontinuerlig endring og det brukes ikke entydige kriterier ved nummerering av nye kretser.

Ved innføring av grunnkrets som basis for soneavgrensningen, står vi derfor overfor noen utfordringer med hensyn til oppdatering og overensstemmelse mellom ulike datakilder.

Til senere versjoner av NTM, vil derfor behovet for rutiner rundt oppdatering av soneinndelingen og sikring av konsistens mellom inputdata og soneinndeling være noe større enn ved bruk av den kommunebaserte soneinndelingen.

Tabell 2.1: Antall kommuner, grunnkretser og NTPL-soner per fylke

Fylke	Beholdning i fylkene			Gjennomsnittstall		
	Antall kommuner	Antall gr.kretser	Antall NTPL-soner	Gr.kretser per sone	Gr.kretser per komm.	NTPL-soner per komm.
1 Østfold	18	689	40	17.2	38.3	2.2
2 Akershus	22	1 366	78	17.5	62.1	3.5
3 Oslo	1	552	26	21.2	552.0	26.0
4 Hedmark	22	799	50	16.0	36.3	2.3
5 Oppland	26	773	64	12.1	29.7	2.5
6 Buskerud	21	685	69	9.9	32.6	3.3
7 Vestfold	15	580	63	9.2	38.7	4.2
8 Telemark	18	595	49	12.1	33.1	2.7
9 Aust-Agder	15	366	38	9.6	24.4	2.5
10 Vest-Agder	15	446	52	8.6	29.7	3.5
11 Rogaland	26	940	110	8.5	36.2	4.2
12 Hordaland	34	1 106	137	8.1	32.5	4.0
14 Sogn og Fjordane	26	488	73	6.7	18.8	2.8
15 Møre og Romsdal	38	764	151	5.1	20.1	4.0
16 Sør-Trøndelag	25	829	104	8.0	33.2	4.2
17 Nord-Trøndelag	24	482	69	7.0	20.1	2.9
18 Nordland	45	1 065	138	7.7	23.7	3.1
19 Troms	25	585	78	7.5	23.4	3.1
20 Finnmark	19	310	39	7.9	16.3	2.1
Sum	435	13 420	1 428	9.4	30.9	3.3

TØI rapport 606/2002

## 2.2 Transportmåter

I RVU er det en inndeling på 18 svaralternativer på spørsmålet om hvilket hovedtransportmiddel som ble brukt på reisen. Disse 18 alternativene fremgår av figur 3.3. I modelleringen inngår fem alternative transportmåter:

- Bil (passasjer og fører)
- Buss (kun rutegående busser)
- Tog
- Fly (kun rutefly)
- Båt (kun rutegående båter)

Denne inndelingen dekker 97 % av de reisene som er gjennomført i RVU-97/98. På et tidspunkt var det vurdert å ta inn charterbuss som et eget transportmiddel i en modell for fritidsreiser, men dette ble droppet av ressurshensyn. De resterende transportmidlene er i følge RVU97/98 marginale når det gjelder lange reiser.

## 2.3 Reiseformål

I RVU er det videre en inndeling på 20 svaralternativer på spørsmål om hva som var hovedformålet med reisen. Disse fremgår av figur 3.4. Modellene omfatter alle disse reiseformålene, men aggregert til 4 formål med egne modeller. Disse er:

- **Arbeids-/tjenestereiser.** Omfatter reiser til/fra arbeid betalt av arbeidsgiver, og alle tjenestereiser (4 kategorier, jf. figur 3-4).
- **Besøksreiser.** Privat besøk hos venner og familie, etc.
- **Fritidsreiser.** Reiser til fornøyelser/underholdning, organiserte fritidsaktiviteter, ferie og andre fritidsreiser.
- **Andre private reiser.** Alle andre reiseformål (jf. figur 3-4), inkl. reiser til/fra arbeid betalt av trafikanten selv.

Denne aggregeringen er valgt dels av pragmatiske grunner. På et tidspunkt ble det vurdert å estimere egne modeller for arbeidsreiser. Fordi datamaterialet i RVU



97/98 her er meget tynt, bare vel 200 observasjoner, måtte man i tilfellet ha supplert med data fra andre kilder (data fra den Norske tidsverdiundersøkelsen), noe som av ulike årsaker heller ikke var direkte enkelt. Dessuten ser det i RVU-en ut til å ha hersket usikkerhet både blant intervjuere og respondenter når det gjelder forståelsen av begrepet arbeidsreise og tjenestereise. Dette har medført at tjenestereiser har blitt klassifisert som arbeidsreiser og omvendt, i et omfang som det i etterkant er vanskelig å tallfeste. Det

ble derfor besluttet at de reiser som er rapportert som arbeidsreiser og er betalt av arbeidsgiver, skulle inngå i modellene for tjenestereiser, mens de arbeidsreiser som er betalt av respondenten ble tatt inn blant andre private reiser. Disse reisene er av relativt marginalt omfang.

Den inndelingen som nå er benyttet gir et omtrent like stort omfang av observasjoner i hvert reiseformål, om lag 1000 reiser.

## 3 Data

Til estimeringen av nye langdistansemodeller, og for så vidt også for estimering av alle transportmodeller av logit-typen, trengs tre typer data som fortrinnsvis må representere samme tidsperiode:

- Reisevanedata
- Data for transporttilbud
- Data for soneinnhold

**Reisevanedata** fremskaffes gjennom intervjuundersøkelser hvor et representativt utvalg av befolkningen stilles en rekke spørsmål om sine reiseaktiviteter og en del bakgrunnsinformasjon om seg selv og sin husholdning.

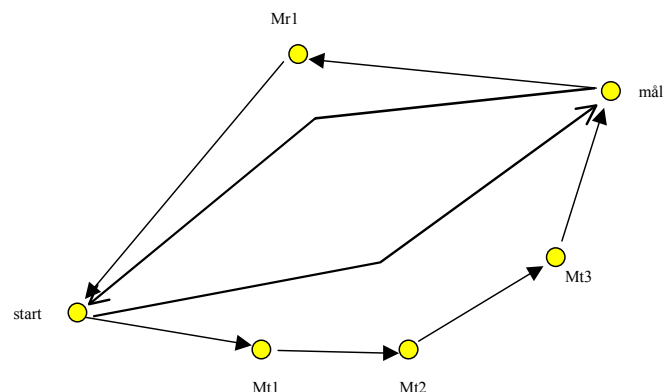
**Data for transporttilbudet** fremskaffes gjennom såkalte nettverksmodeller som har en geografisk dimensjon representert ved en geografisk soneinndeling og nettverk for alle relevante transportmåter som knytter disse sonene sammen. Det nasjonale nettverket er foreløpig laget til programpakken EMMA og består av et vegnett generert med GISNETT og kollektivruter med tilhørende infrastruktur generert med KOLLNETT (Hamre 2001). I nettverksmodellene kan man gjøre beregninger for å ta frem reisetider og reisekostnader mellom alle geografiske områder definert gjennom soneinndelingen, med alle transportmåter som er tilgjengelig.

**Data for soneinnhold** fungerer som skapende og attraherende elementer når det gjelder reiseaktiviteter. Det er befolkningsstørrelsene i de ulike geografiske områdene som skaper trafikken. Befolkningen deles blant annet inn etter alder, kjønn, og yrkesaktivitet. Arbeidsplasser fordelt etter næring, hotellsenger, hytter og fritidshus, m.m., fungerer som *proxy*-variable på elementer som trekker til seg turer. I tidligere versjoner av NTM har vi bare hatt tilgang på sonedata på kommunenivå. I foreliggende arbeide er det lagt stor vekt på å skaffe denne type data til veie på grunnkrets nivå. Dataene er hovedsakelig innhentet fra Statistisk Sentralbyrå.

### 3.1 Reisevaneundersøkelsen 1997/1998 (RVU-97/98)

#### 3.1.1 Generelt

Dette arbeidet er basert på reisevanedata fra den landsomfattende undersøkelsen gjennomført i perioden fra høsten 1997 til sommeren 1998 (RVU-97/98). Her ble om lag 8800 respondenter intervjuet blant annet om sine lange reiser lengre enn 10 mil én veg gjennomført i løpet av den siste månedperioden. Respondentene skulle rapportere alle delreiser innenfor hver turkjede de hadde gjennomført. Imidlertid manglet en del reiser i en rekke av de rapporterte turkjedene. Dette medførte at en stor andel av intervjuene måtte gås gjennom manuelt for å finne ut hvordan turkjedene egentlig er satt sammen. Figur 3.1 gir et eksempel på en relativt omfattende turkjede.



TØ1 rapport 606/2002

Figur 3.1: Eksempel på turkjede

I estimeringen må turkjeder av denne type gjøres om til rundturer med én hoveddestinasjon. Hoveddestinasjonen er definert som den destinasjon som ligger lengst unna startpunktet for reisen. Turkjeden i figur 3-1 har, som vi ser, 5 destinasjoner, tre før og én etter den destinasjonen som ligger lengst unna startpunktet. Til estimeringsformål er turkjeden kodet om til rundtur som illustrert i figuren. I aggregeringen fra delreiser til rundturer er det en rekke kompliserende forhold knyttet til at trafikantene benytter forskjellige transportmåter på de ulike delturene, at de har forskjellige formål på de ulike destinasjonene, m.m..

Her har vi benyttet ulike regler for å gjennomføre aggregeringen på en konsistent måte i datamaterialet.

Aggregeringen fra delreiser til rundturer er forsøkt gjort tilnærmedesvis slik man nå forsøker å gjøre det i RVU2001. Dette innebærer at vi så langt som mulig velger den destinasjonen som ligger lengst unna startstedet som hoveddestinasjon. Det er registrert opp til 6 delturer i en rundtur. Valg av den deltur som skal representere rundturen er forsøkt gjort ved hjelp av faste beslutningsregler.

Ved rene tur/retur-reiser velges fortrinnsvis utreisen. Hvis respondenten ikke har oppgitt utreise, velges imidlertid returen.

Ved tre delreiser velges den destinasjon som er lengst unna. Hvis de to destinasjonene ligger tilnærmedesvis like langt unna, velges den hvor respondenten oppgir å ha overnattet lengst. Samme regler er benyttet ved fire, fem og seks delturer. I tvilstilfeller er også det benyttede hovedtransportmiddel til destinasjonene brukt i en "manuell" vurdering og til valget av én "hoveddestinasjon".

Heldigvis forekommer slike omfattende turkjeder som er vist i figuren, svært sjelden i datamaterialet. Dette fremgår av tabell 3.1, som viser antallet delreiser og antallet rundturer disse er aggregert opp til etter antall mellomliggende destinasjoner.

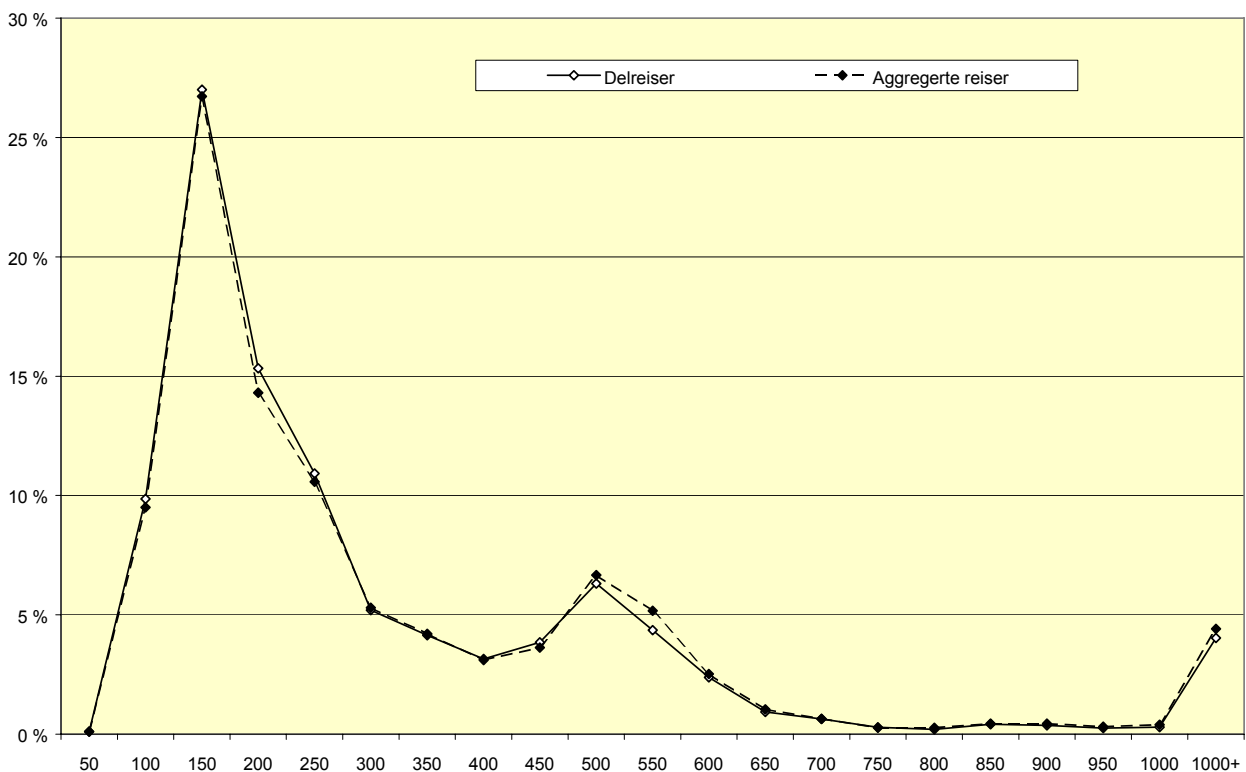
Tabell 3.1: Delreiser og rundturer etter antall mellomliggende destinasjoner i RVU 97/98

Antall mellomliggende destinasjoner	Rundturer	Delreiser
Tur/retur, 2 delturer	4 647	9 294
1 mellomliggende destinasjon *, 3 delturer	135	405
2 mellomliggende destinasjoner *, 4 delturer	39	156
3 mellomliggende destinasjoner *, 5 delturer	10	50
4 mellomliggende destinasjoner *, 6 delturer	2	12
B Sum "brukbare" rundturer/delreiser	4 833	9 917

\*Ekskl. hoveddestinasjon

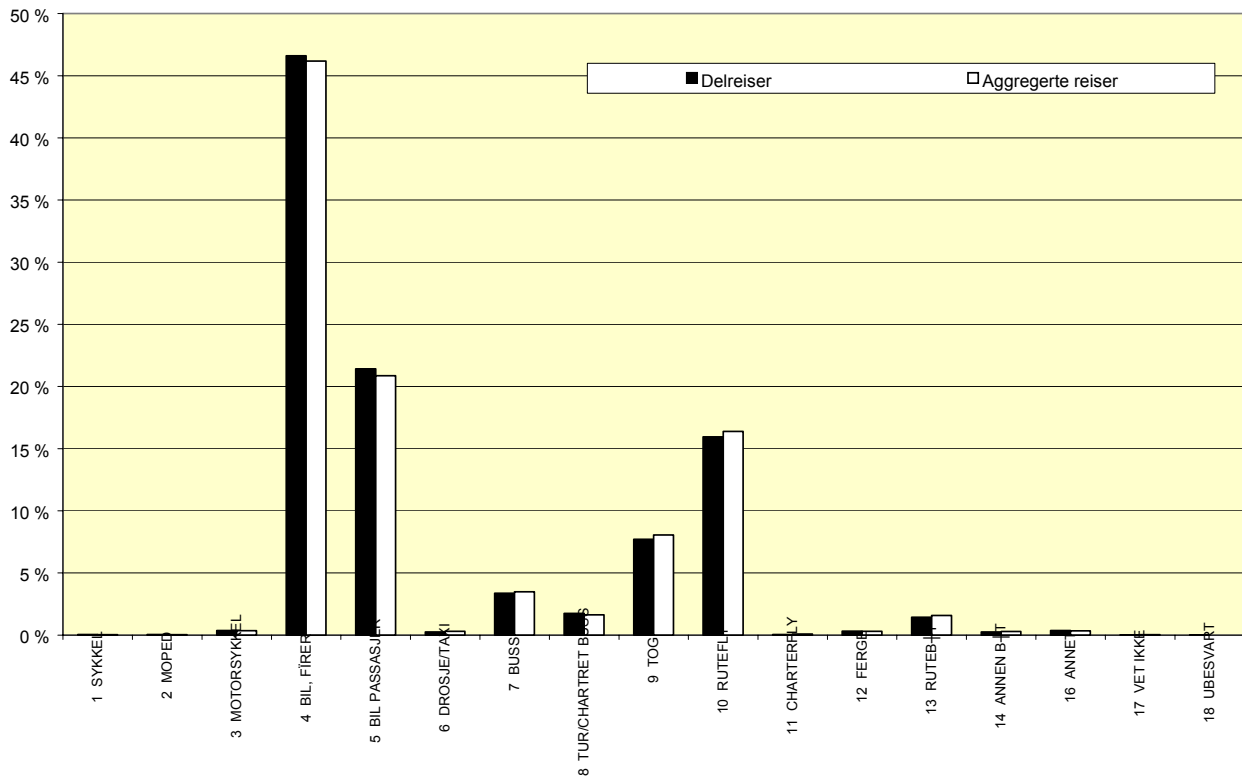
TØI rapport 606/2002

Total reiselengde for alle delturer i datafilen er 3167 mill. km. Etter aggregeringen er reiselengden 3135 mill. km. Dette betyr at aggregeringen fra delturer til tur/retur-reiser har medført at vi har "mistet" ca 1 % av total reist lengde. Figur 3.2 viser avstandsfordelingen for delreisene og for aggregerte rundturer (én veg). Vi ser at vi naturlig nok får noe færre korte reiser og en del flere lange reiser, spesielt i intervallet fra 450 til 600 km. Vi ser imidlertid at aggregeringen ikke endrer avstandsfordelingen dramatisk.



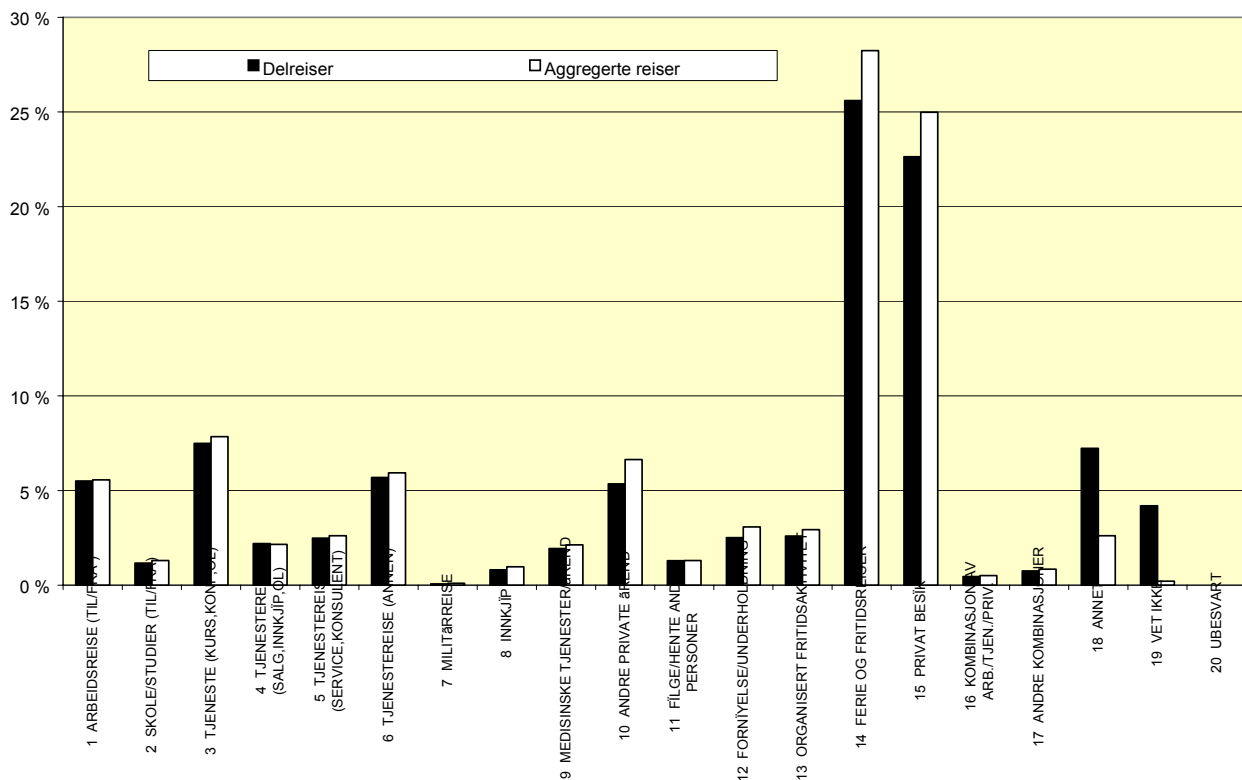
TØI rapport 606/2002

Figur 3.2: Avstandsfordeling for delreiser og for aggregerte reiser



TØI rapport 606/2002

Figur 3.3: Transportmiddelfordeling for delreiser og for aggregerte reiser



TØI rapport 606/2002

Figur 3.4: Formålsfordeling for delreiser og for aggregerte reiser

Figur 3.3 og 3.4 viser at det samme gjelder for transportmiddelfordeling og formålsfordeling. Når det gjelder transportmiddelfordelingen, blir det som vi ser forholdsmessig noe færre reiser som bilfører og passasjer, mens rutefly og tog øker noe. Formålskategoriene "annet" og "vet ikke" reduseres også ganske kraftig til fordel for de fleste spesifikke reisemål, spesielt ferie og fritidsreiser samt besøksreiser. Dette, som er vist i den siste figuren, kan tyde på at vi har truffet brukbart i arbeidet med å slå sammen delreiser til rundturer med én hoveddestinasjon.

Gjennom bearbeidelsen av datafilene fra RVU 97/98 er det laget en datafil som for de respondenter som ikke har gjennomført noen lange reiser inneholder én linje med bare person og husholdningsopplysninger, og for de som har reist, inneholder én linje for hver rundtur. For de som har reist er all informasjon om reisen (inkl. utvalgt hoveddestinasjon og mellomliggende destinasjoner) lagret. Denne filen er lagret i SPSS og fungerer som utgangspunkt for å kjøre ut reisevanedata til estimeringen.

### 3.1.2 Manglende opplysninger om inntekt

Blant de RVU-data som inngår i estimeringen, er opplysninger om personlig inntekt og husstandens samlede inntekt sentrale. Ved modellberegninger av langsiktige prognoser, vil inntekt være viktigste eksogene variabel.

Dessverre viser inntektsvariable seg ofte å være blant de bakgrunnsopplysninger som har mest mangelfull registrering i spørreundersøkelser. I RVU 97/98 mangler 29 % av respondentene opplysning om personinntekt og/eller husstandsinnntekt. For de to inntektsvariablene isolert sett, mangler 19 % av personinntektene og 27 % av husstandsinnntektene.

Ukorrigert ville dette medføre at alle opplysninger fra disse respondentene blir utelatt fra modellestimeringen til NTM5. Spesielt i langdistansemodellen, hvor datagrunnlaget i utgangspunktet er lite, gir dette et uakseptabelt tap av verdifulle reiseregistreringer.

En mulighet til å forbedre datagrunnlaget, er å utnytte den delen av RVU-materialet som *har* inntektsopplysninger, og undersøke i hvilken grad inntekten har sammenheng med andre bakgrunnsopplysninger i RVU-en.

Forklaringskraften i disse sammenhengene vil vise om manglende verdier for inntekt lar seg estimere med akseptabel nøyaktighet med utgangspunkt i andre opplysninger i RVU-en.

Vi vet at personers inntekt varierer med kjønn, alder og yrkesdeltagelse. Disse opplysningene er det

spurt etter i RVU-en. I tillegg vet vi at det er geografiske forskjeller i inntektsnivå, både mellom regioner/landsdeler og innen regioner/landsdeler. Særlig når det gjelder variasjoner innen regioner må vi regne med at bosted i noen grad er endogen i forhold til inntekt.

Andre variable som til dels kan være endogene i forhold til inntekt er førerkortinnehav og bilhold. På den annen side, hvis vi bare er interessert i en relasjon som skal benyttes til å estimere inntekt behøver ikke legge så mye vekt på forholdet eksogene/endogene variable. Dette er først og fremst relevant hvis vi skal si noe om kausale sammenhenger. I tillegg vil variable for høyeste fullførte utdanning, næringstilknytning og stilling kunne bidra til å estimere inntekt.

#### 3.1.2.1 Testkjøringer med lineær regresjonsanalyse for estimering av inntekt

Innledningsvis ble det gjort noen testkjøringer av regresjonsanalyser med de antatt viktigste forklaringsvariable nevnt ovenfor, og personinntekt i ulike former som avhengig variabel. Forklaringsvariable som ble testet var intervjuår, kjønn, førerkort- og bilinnehav, yrkesaktivitet på heltid/deltid og alder.

Forklaringsgraden ( $R^2$ ) viste seg ikke påfallende høy (maks 0,45), noe som tyder på at flere bakgrunnsopplysninger bør tas med i de endelige analyser. For de enkelte variablene viste testen gode resultater med koeffisientverdier gjennomgående signifikant forskjellig fra 0.

Når det gjelder den avhengige variabel (inntekt), er en lineære spesifisering av denne ikke nødvendigvis det beste. Av testkjøringene ga regresjonsanalysen direkte på inntektsvariabelen (IPINNT) også den laveste  $R^2$  (0,34). At IPINNT alltid vil være større eller lik 0, ville også gi et klart brudd på de vanlige forutsetninger for minste kvadraters metode.

Vi har derfor valgt å bruke logaritmisk spesifisering av inntektvariablene. Ved regresjon med  $\ln$  inntekt som avhengig variabel, antar vi at ulike variable bidrar med prosentvis endring i inntekten.

En sammenlikning av regresjonsanalyser på hhv  $\ln(\text{IPINNT}+1)$  og  $\ln(\text{IPINNT}+10)$  viser jevnt over mer presist bestemte variable og bedre forklaringskraft for sistnevnte versjon ( $R^2 = 0,45$  mot  $R^2 = 0,39$ ). Når denne versjonen ser noe bedre ut, er det trolig på grunn av at helt lave inntekter, spesielt inntekt lik null, skaper problemer med logaritmisk spesifisering.

På bakgrunn av disse testresultatene, er de endelige analysene kjørt med  $\ln(\text{IPINNT}+10)$  som avhengig variabel for personinntekt. Tilsvarende er

$\ln(\text{HUSINNT}+50)$  brukt ved estimering av husstandsinntekter.

Når det gjelder estimering av husstandsinnpekt, må vi regne med å oppnå en noe svakere modell enn for personinntekt. I tillegg til at vi har litt færre registrerte inntektsopplysninger å basere estimeringen på, inneholder RVU-en også mindre data om øvrige husstandsmedlemmer enn om intervjupersonen selv. Vi må også gå ut fra at det er noe større usikkerhet i intervjupersonenes anslag på husstandsinnpekt enn på egen inntekt.

### 3.1.2.2 Beskrivelse av forklaringsvariable

Forklaringsvariablene brukt i regresjonsanalysene for inntekter, er i hovedsak dummyvariable (verdi 0 eller 1). En del av disse var registrert som dummyvariable i det originale datamaterialet, for de øvrige er det konstruert nye variable for analysene. Den eneste forklaringsvariabel som ikke er representert som dummy, er alder, som vi har med i to varianter: ALD og ALD2 (ALD<sup>2</sup>), beregnet med utgangspunkt i den originale variabel ALDER. Respondenter i RVU 97/98 er 13 år eller eldre. For å oppnå jevne verdiskalaer fra 0 og oppover, har vi satt ALD lik ALDER-13.

Tabell 3.2: Variable brukt i regresjonsanalyse for personlig inntekt. Avhengig variabel:  $\ln(\text{IPINNT}+10)$

Variabel	Forklaring	Verdier
Opplysninger om IP:		
KJONN_D	Kjønn	Dummy, 1=kvinne
ALD	Alder-13	
ALD2	ALD <sup>2</sup>	
FKORT_D	Fører kort innehav	Dummy 1/0
Dummyer for bilbestanden i husholdet:		
BIL_D	Bil i hushold	Dummy 1/0
BREGIP_D	Bil(er) registrert på IP	Dummy 1/0
BREGA_D	Bil(er) registrert på arbeidsgiver	Dummy 1/0
Dummyer for IP's yrkesaktivitet:		
YAH_D	Yrkesaktiv, heltid (30 t eller mer pr uke)	Dummy 1/0
YAD_D	Yrkesaktiv, deltid (<30 t pr uke)	Dummy 1/0
Dummyer for IP's bosted:		
OSLO	Byregion Oslo	Dummy 1/0
BERGEN	Byregion Bergen	Dummy 1/0
NJÆREN	Byregion Nord-Jæren	Dummy 1/0
KRSAND	Byregion Kristiansand	Dummy 1/0
TØNSBERG	Byregion Tønsberg	Dummy 1/0
MINBY	Mindre byområder	Dummy 1/0
Dummyer for IP's utdanningsbakgrunn:		
VID_YRK	Videregående skole, yrkesfaglig/handel	Dummy 1/0
VID_ALM	Videregående skole, almennfaglig	Dummy 1/0
UH_LAV	Høyskole/universitet, lavere grad (inntil 15 år)	Dummy 1/0
UH_HØY	Høyskole/universitet, høyere grad (16 år eller mer)	Dummy 1/0
Dummyer for IP's stilling:		
FAGLÆRT	Faglært arbeider, formann	Dummy 1/0
FAG_FUNK	Fagfunksjonær	Dummy 1/0
OVERORDN	Overordnet stilling i privat eller off. virksomhet	Dummy 1/0
GÅRDFISK	Gårdbruker, fisker	Dummy 1/0
EIERBEDR	Eier av virksomhet/bedrift	Dummy 1/0
SJØMANN	Sjømann/skipper	Dummy 1/0
Dummyer for IP's hovedbeskjeftigelse:		
HUSARB	Husarbeid i hjemmet	Dummy 1/0
SKOLSTUD	Går på skole/studerer	Dummy 1/0
PENSJON	Alderspensionist	Dummy 1/0
UFØREP	Uføretrygdet og andre pensjonister	Dummy 1/0
ANNET	Annet	Dummy 1/0

TØI rapport 606/2002

Tabell 3.3: Variable brukt i regresjonsanalyse for husstandsinnpekt. Avhengig variabel:  $\ln(\text{HUSINNT}+50)$ 

Variabel	Forklaring	Verdier
Opplysninger om IP:		
KJONN_D	Kjønn	Dummy, 1=kvinne
ALD	Alder-13	
ALD2	ALD2	
FKORT_D	Fører kort innehav	Dummy 1/0
Dummyer for bilbestanden i husholdet:		
BIL_D	Bil i hushold	Dummy 1/0
BREGIP_D	Bil(er) registrert på IP	Dummy 1/0
BREGA_D	Bil(er) registrert på arbeidsgiver	Dummy 1/0
Dummyer for IP's yrkesaktivitet:		
YAH_D	Yrkesaktiv, heltid (30 t eller mer pr uke)	Dummy 1/0
YAD_D	Yrkesaktiv, deltid (<30 t pr uke)	Dummy 1/0
Dummyer for IP's bosted:		
OSLO	Byregion Oslo	Dummy 1/0
BERGEN	Byregion Bergen	Dummy 1/0
NJÆREN	Byregion Nord-Jæren	Dummy 1/0
KRSAND	Byregion Kristiansand	Dummy 1/0
TØNSBERG	Byregion Tønsberg	Dummy 1/0
MINBY	Mindre byområder	Dummy 1/0
Dummyer for IP's utdanningsbakgrunn:		
VID_YRK	Videregående skole, yrkesfaglig/handel	Dummy 1/0
VID_ALM	Videregående skole, almennfaglig	Dummy 1/0
UH_LAV	Høyskole/universitet, lavere grad (inntil 15 år)	Dummy 1/0
UH_HØY	Høyskole/universitet, høyere grad (16 år eller mer)	Dummy 1/0
Dummyer for IP's stilling:		
FAGLÆRT	Faglært arbeider, formann	Dummy 1/0
FAG_FUNK	Fagfunksjonær	Dummy 1/0
OVERORDN	Overordnet stilling i privat eller off. virksomhet	Dummy 1/0
GÅRDFISK	Gårdbruker, fisker	Dummy 1/0
EIERBEDR	Eier av virksomhet/bedrift	Dummy 1/0
SJØMANN	Sjømann/skipper	Dummy 1/0
Dummyer for IP's hovedbeskjeftigelse:		
HUSARB	Husarbeid i hjemmet	Dummy 1/0
SKOLSTUD	Går på skole/studerer	Dummy 1/0
PENSJON	Alderspensionist	Dummy 1/0
UFØREP	Uføretrygdet og andre pensjonister	Dummy 1/0
ANNET	Annet	Dummy 1/0

TØI rapport 606/2002

Innen variabelkategoriene for bostedsregioner, stilling og hovedbeskjeftigelse, finnes flere alternativer i RVU 97/98 enn de som er brukt i analysene. Dette skyldes at variable som ikke viser signifikante effekter er silt vekk, og inngår dermed i referansematerialet i analysen. I analysen av husstandsinnpekt finner vi variabelpar som er innen samme kategori og med liknende signifikans. Dette gjelder for eksempel byregionene Trondheim og Nedre Glomma, og disse er derfor slått sammen til en felles parameter.

En del forklaringsvariable vil naturlig nok gå igjen både i analysene av personenes og husstandenes innpekt. Man kan forvente at flere av de mest signifikante variablene for personlig innpekt viser liknende utslag også for husstandsinnpekt. Andre variable vil være representert i analysen av personinnpekt men ikke av husstandsinnpekt, og omvendt.

Slike variable finner vi for eksempel blant byregionene, hvor enkelte kan ligge generelt høyt i personinntekter, men hvor kanskje sammensetningen

av husstandene i regionen fører til at området ikke får like stort utslag for samlet husholdsinntekt. For andre regioner kan det motsatte være tilfellet.

Variablene brukt til estimering av hhv personinntekt og husstandsinnkomst, er vist i tabell 3.2 og 3.3.

Variable brukt til estimering av husstandsinnkomst, er vist i tabell 3.2.

### 3.1.2.3 Eliminering av ekstreme inntektsverdier fra datamaterialet før estimering

RVU 97/98 er basert på telefonintervjuer, og opplysningene om inntekt registreres i 1000 kr. Det er ikke utenkelig at disse dataene er mer utsatt for betydelige punchefeil enn variable med klare kategorier eller enklere verdiskalaer.

Kontroll av enkelte ekstreme verdier, tyder på at det i noen tilfeller kan ha oppstått misforståelser eller feilslag i antall nuller (for eksempel skoleelev med personlig inntekt på 3 mill. kr). Ekstreme verdier kan selvfølgelig være reelle, også om de virker usannsynlige.

På grunn av at ekstremverdier enkeltvis påvirker resultatet mer enn andre registreringer og svekker forklaringsklaringskraften, kombinert med større sannsynlighet for at de er feilregistrert, har vi valgt å bruke en statistisk metode for avgrensning av datamaterialet.

Verdiene for inntekter registrert i RVU 97/98 ligger i området 0- 7 000 000 kr. Vi har valgt å gjøre avgrensninger rettet mot høye verdier.

I forbindelse med regresjonsanalyse kan det beregnes standard residualverdier (ZRE) som angir hvordan hver enkelt record i datamaterialet er tilpasset regresjonsresultatet. De records som treffer best vil få ZRE rundt 0, hvor vi også finner hovedtyngden av observasjonene.

Vi valgte å avgrense datamaterialet, ved å ikke bruke observasjoner med standard residualverdi større eller lik 1,5 i de videre regresjonsanalysene. Metoden eliminerer ikke nødvendigvis de ekstremt høye inntektsverdiene, men de verdier som avviker mest oppad, ut fra innholdet i forklaringsvariablene i regresjonsmodellen. Likevel fører avgrensningen som ventet til størst uttynning blant de høyeste inntektene i RVU-materialet.

Med betingelsen  $ZRE < 1,5$  utelates 4-4,5 % av observasjonene før de endelige regresjonsanalysene, og forklaringsgraden økes betydelig. Ved elimineringen av et såpass lite antall ekstremverdier, økte  $R^2$  fra 0,68 til 0,77 for personinntekt og fra 0,60 til 0,68 for husstandsinnkomst.

### 3.1.2.4 Personlig inntekt - Regresjonsanalysen på $\ln(ipinnt+10)$

I tabell 3.4 vises regresjonsanalysen med  $\ln(ipinnt+10)$ , hvor  $ipinnt$  er målt i 1000 kroner.

Tabell 3.4: Regresjon med  $\ln(ipinnt + 10)$  som avhengig variabel

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,878	,771	,770	,45952

ANOVA<sup>b</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4861,478	30	162,049	767,413	,000
	Residual	1444,989	6843	,211		
	Total	6306,467	6873			

b. Dependent Variable: LN\_PINNT  $\ln(ipinnt+10)$

Coefficients<sup>a</sup>

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	3,529	,042		83,627	,000
KJONN_D	-,190	,013	-,099	-15,008	,000
ALD	4,293E-02	,002	,738	22,293	,000
ALD2	-5,412E-04	,000	-,639	-17,999	,000
FKORT_D	,315	,021	,120	15,063	,000
BIL_D	-,160	,020	-,056	-7,968	,000
BREGIP_D	,167	,015	,087	11,239	,000
BREGA_D	,171	,028	,037	6,118	,000
YAH_D	,717	,028	,359	25,196	,000
YAD_D	,384	,026	,127	14,609	,000
OSLO	,116	,014	,058	8,443	,000
BERGEN	8,440E-02	,027	,019	3,110	,002
NJÆREN	,130	,031	,025	4,193	,000
KRSAND	,198	,048	,024	4,134	,000
TØNSBERG	9,758E-02	,049	,012	1,971	,049
MINBY	3,677E-02	,016	,015	2,290	,022
VID_YRK	,131	,017	,061	7,600	,000
VID_ALM	,192	,020	,071	9,619	,000
UH_LAV	,242	,020	,094	11,931	,000
UH_HØY	,348	,021	,140	16,777	,000
FAGLÆRT	9,888E-02	,020	,034	4,942	,000
FAG_FUNK	8,065E-02	,018	,034	4,471	,000
OVERORDN	,230	,022	,077	10,665	,000
GÅRDFISK	-,166	,047	-,021	-3,552	,000
EIERBEDR	,138	,034	,025	4,078	,000
SJØMANN	,330	,104	,019	3,176	,001
HUSARB	-1,097	,040	-,191	-27,528	,000
SKOLSTUD	-,890	,033	-,274	-27,018	,000
PENSJON	,370	,039	,124	9,444	,000
UFØREP	,192	,040	,037	4,824	,000
ANNET	,348	,059	,037	5,887	,000

a. Dependent Variable: LN\_PINNT

TØ1 rapport 606/2002

Av de totalt 8838 records i materialet, har vi altså 6873 med opplysninger om de variable som er benyttet i



regresjonsanalysen. Forklaringsgraden ( $R^2$ ) på 0,77 må sies å være svært god, og effekten av forklaringsvariablene er gjennomgående signifikant forskjellig fra null.

Ser vi nærmere på noen av dummyvariablene, tyder variabelen KJONN\_D på at kvinner, når man korri-gerer for de øvrige variable, har ca 19 % lavere inntekt enn menn. Bil i husholdet er ikke nødvendigvis tegn på høy personlig inntekt, men de med høyere inntekt ser i større grad ut til å ha bil registrert på seg selv eller på arbeidsgiver.

De med førerkort har betydelig høyere inntekt (ca 32%) enn de uten. Dette er selvfølgelig til en viss grad korrelert med IP's alder, men alder i seg selv utgjør i tillegg en viktig forklaringsfaktor. Kombinasjonen av variablene ALD og ALD2 viser at inntekten øker med alderen fram til nærmere 60 år for deretter å avta.

Vi ser også at byområdene i og rundt Kristiansand, Stavanger (Nord-Jæren) og Oslo, ligger høyere i inntekt enn landet for øvrig.

Slik vi hadde forventet, har variablene for yrkesstatus og utdanning også stor betydning i en regresjonsanalyse på inntekt. Vi ser store signifikante koeffisienter for yrkesaktivitet på hhv heltid og deltid regnet i forhold til ikke yrkesaktive, samt at inntektsnivået stiger med økende utdanningsgrad.

### 3.1.2.5 Husstandens inntekt - Regresjonsanalysen på $\ln(\text{husinnt} + 50)$

I tabell 3.5 vises regresjonsanalysen med  $\ln(\text{husinnt}+50)$ , med husinnt målt i 1000 kroner. Som ventet gir regresjonsanalysen på husstandsinn-tekter en noe lavere forklaringsgrad enn den vi fikk for personinntekter. Data som beskriver husstand og familietype viser seg svært nyttige i analysen, selv om disse variable blir litt for generelle til å frambringe en like høy forklaringsgrad som personinntektsmodellen oppnådde. Analysen av husstandsinn-tekter har også færre observasjoner å bygge på (6162). Forklaringsgraden ( $R^2=0,68$ ) må likevel anses som fullt ut akseptabel for estimering av de manglende verdier i RVU-en.

Tabell 3.5: Regresjon med  $\ln(\text{husinnt} + 50)$  som avhengig variabel

#### Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,823	,678	,675	,28449

#### ANOVA<sup>b</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1040,594	44	23,650	292,217	,000
	Residual	495,145	6118	,081		
	Total	1535,739	6162			

b. Dependent Variable: LN\_HINNT

#### Coefficients<sup>a</sup>

	Unstandardized Coefficients		Standard-ized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	4,949	,030		165,646	,000
IAAR_D	-2,964E-02	,008	-,026	-3,522	,000
ALD	1,238E-02	,001	,387	9,664	,000
ALD2	-1,869E-04	,000	-,422	-9,412	,000
KJONN_D	-4,171E-02	,008	-,042	-5,193	,000
FKORT_D	-8,135E-02	,017	-,053	-4,822	,000
ANTFK	8,722E-02	,010	,135	8,731	,000
SQRT_BIL	,163	,012	,146	13,958	,000
BREGA_D	8,681E-02	,018	,037	4,722	,000
ANTYAKT	,162	,009	,219	17,577	,000
FAMTYP02	,321	,013	,286	25,413	,000
FAMTYP05	8,718E-02	,038	,017	2,292	,022
FAMTYP06	,313	,014	,259	21,905	,000
FAMTYP07	,319	,018	,183	18,078	,000
FAMTYP08	,315	,022	,142	14,447	,000
FAMTYP09	,221	,021	,123	10,749	,000
FAMTYP10	-,252	,051	-,037	-4,978	,000
AGOD0203	,104	,017	,049	6,259	,000
AGODUT04	-,148	,041	-1,279	-3,648	,000
AGODUT06	,144	,041	1,242	3,541	,000
YAH_D	,129	,013	,122	9,727	,000
HINNTSON	5,247E-02	,013	,036	4,114	,000
OSLO	,115	,011	,111	10,757	,000
BERGEN	7,624E-02	,018	,033	4,222	,000
NJÆREN	,167	,021	,061	7,990	,000
GRENLAND	8,344E-02	,031	,020	2,656	,008
DRAMMEN	9,032E-02	,026	,026	3,412	,001
KRSAND	9,314E-02	,032	,022	2,912	,004
TØNSBERG	,114	,033	,026	3,469	,001
TRH_NG	4,514E-02	,018	,019	2,492	,013
MINBY	6,045E-02	,011	,047	5,449	,000
VID_YRK	4,141E-02	,011	,038	3,654	,000
VID_ALM	7,571E-02	,014	,053	5,594	,000
UH_LAV	,123	,013	,095	9,305	,000
UH_HØY	,210	,014	,168	15,313	,000
FAG_FUNK	5,230E-02	,011	,043	4,769	,000
OVERORDN	,132	,013	,087	10,094	,000
GARDFISK	-,116	,030	-,030	-3,927	,000
FRITTERV	4,522E-02	,027	,013	1,670	,095
JOURNALI	,192	,077	,018	2,509	,012
HUSARB	-,109	,023	-,042	-4,713	,000
SKOLSTUD	-,154	,022	-,072	-6,911	,000
PENSJON	-,116	,025	-,075	-4,729	,000
UFØREP	-,217	,024	-,083	-9,006	,000
ARBLED	-,186	,048	-,029	-3,872	,000

a. Dependent Variable: LN\_HINNT

TØ1 rapport 606/2002

Flere av variablene med personlige bakgrunnsdata om IP har fortsatt liknende effekt som for personinntekt, men naturlig nok med noe lavere koeffisientverdi og signifikans for husstandsinn-tekten. Dette gjelder for eksempel variable for alder, kjønn, yrkesaktivitet

(heltid), utdanningsnivå, stillingskategori og hovedbeskjeftigelse.

For sistnevnte kategori ser vi også at alders- og uførepensjonister, som lå noe høyere i personinntekt sammenliknet med yrkesaktive generelt, nå slår ut med motsatt fortegn når det gjelder husstandsinnpekt. At IP har førerkort viser heller ikke like sterk sammenheng med husstandens inntekt som med personlig inntekt, alt annet likt. Tvert imot har FKORT\_D nå fått en liten, men signifikant negativ koeffisient, men husstandsinnpekten vil øke med antall førerkort i husholdningen.

Husstanden har, ikke uventet, også høyere inntekt jo flere yrkesaktive og biler den inneholder. Inntekten er også høyere i husstander hvor bilutgifter er helt eller delvis dekket av arbeidsgiver.

Variablene for familietype er svært beskrivende for husstandssammensetningen, og noen av disse er også blant de mest signifikante i denne analysen. Basisgruppen for denne variabelkategorien utgjøres av enslige og enslige med barn under 13 år. I forhold til disse gruppene, ligger par med eller uten barn 31-32 % høyere i husstandsinnpekt. Husstander med flere voksne i samme familie har ca 22 % høyere inntekt enn enslige.

For byregioner er det kommet med flere områder enn vi hadde blant forklaringsvariablene for personinntekt. Basisområdet er nå redusert til Tromsø + resten av landet, og signifikansverdiene har dermed i de fleste tilfeller økt noe. Nord-Jæren, Oslo og Tønsberg er de byområdene som viser høyest nivå på husstandsinnpektene. I tillegg til byregionene, har vi gruppert noen bydeler/kommuner i Oslo og Akershus i en egen "høyinntektssone". For husstandsinnpektene har denne en signifikant koeffisientverdi på toppen av den effekt vi allerede har for Oslo-regionen.

Dummy-variabelen for intervjuår kommer ut med en signifikant verdi som forteller at det i intervjuene gjort i 1997 ble registrert nær 3 % lavere verdier for husstandsinnpekte enn året etter.

### 3.1.2.6 Konstruksjon av nye inntektsvariable til RVU 97/98 på grunnlag av regresjonsanalysene

Ved bruk av inntektsfunksjoner hvor koeffisientverdiene fra regresjonsanalysene inngår, er estimater for hhv person- og husstandsinnpekt beregnet for hver observasjon i RVU-fila. Av disse er vi spesielt ute etter å bruke estimatene for de observasjoner som fra før av mangler registrering av person- og/eller husstandsinnpekt. De nye variablene som skal inngå i datagrunnlaget for estimering av persontransportmodellen,

vil med andre ord bestå av en kombinasjon av registrerte (oppgitte) inntektsverdier og estimater.

Konstruksjonen av nye variable for hhv. person- og husstandsinnpekt, er beskrevet nedenfor:

#### Personinntekt

- I utgangspunktet mangler RVU 97/98 opplysninger om personlig inntekt i nær 19 % av observasjonene.
- Ikke oppgitt personinntekt, enslige og enslige med barn under 13 år:  
I noen få tilfeller (6 stk) er husstandsinnpekten registrert. For disse settes personinntekt lik husstandsinnpekt.
- For observasjoner med manglende personinntekt hvor også husstandsinnpekt mangler, brukes estimert personinntekt. Etter denne operasjonen er andel "missing" nede i 2,5%.
- For øvrige observasjoner med manglende personinntekt brukes også estimert verdi, dersom denne ikke overstiger ev. registrert husstandsinnpekt.

#### Husstandsinnpekt

- Opplysning om husstandens samlede inntekt mangler i nær 27 % av observasjonene i RVU 97/98
- Ikke oppgitt husstandsinnpekt: Vi godtar ikke estimerte verdier for husstandsinnpekt hvis lavere enn personinntekt. Dette gjelder 186 observasjoner. Fordi 80 % av disse er enslige eller enslige forsørgere, gjør vi ikke store feil ved å sette husstandsinnpekt lik personinntekt for samtlige.
- Det samme gjøres generelt for alle enslige og enslige forsørgere med barn under 13 år, dersom husstandsinnpekt ikke er angitt. Etter denne operasjonen er andel "missing" 21%.
- For de ovennevnte tilfeller brukes altså personinntekt - registrert eller estimert - som rettesnor framfor estimert husstandsinnpekt. Dette gjøres fordi personinntekt både har høyest svarprosent og best regresjonsmodell.
- For øvrige observasjoner med manglende husstandsinnpekt, brukes estimert verdi for denne.

Andel "missing" i de nye variablene for personinntekt og husstandsinnpekt ender rundt 0,3 % i begge tilfeller.

### 3.1.3 Data for reiser kortere enn 100 kilometer

Vi har to kilder for å studere de mellomlange reisene; "de lengste daglige" reiser i turdagboken og "de korteste lange reiser" gjennomført siste måned i datafilen for lange reiser. Det er flere grunner til nøye å studere fordelingen av disse reisene.

- Det er regionalt og politisk viktige reiser (intercity-markedet for tog, konkurranseflater buss/tog, pendlingsomland til regionale sentra, pendling vs telependling, m.m.).
- Lange arbeidsreiser er subsidierte reiser (skattefradrag, firmabilordninger, rushproblemer rundt store byer, dyr infrastruktur både på jernbane og veg, m.m.).
- Det er reiser som vi egentlig ikke samler inn gode data for (tynt representert i turdagboken og sannsynligvis beheftet med frafall og skjevheter i seksjonen for lange reiser).

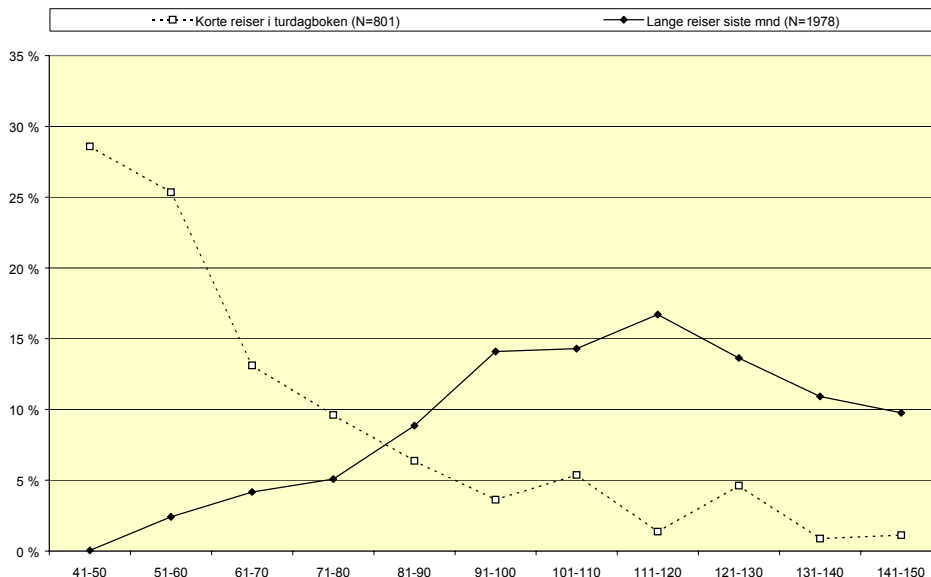
Det er derfor viktig og prioritert innenfor videreutviklingen av langdistansemodellene at vi får et bedre grep om disse reisene enn i dagens NTM. Spørsmålet blir hvordan dette skal gjøres. I figur 3.5 sammenliknes avstandsfordelingen i turdagboken med avstandsfordelingen i datafilen for lange reiser. Fra begge kilder har vi bare tatt med reiser i intervallet 40 til 150 km. I turdagboken ser vi at fordelingen synker frem til ca 100 km hvor den blir litt ujevn, som til dels kan skyldes litt få observasjoner innenfor 10 km intervaller. Fordelingen i datafilen for lange reiser øker relativt jevnt frem til 100 km hvor den også flater ut, men får en topp ved om lag 120 km for så å synke jevnt frem mot 150 km.

For å få et litt bedre sammenlikningsgrunnlag, kan vi aggregere avstandsintervallene og blåse tallene fra

turdagboken opp til månedstall (det vil si grovt sett å multiplisere med 30). Dette er gjort i tabell 3.6. Selv om det på grunn av få observasjoner er beheftet betydelig statistisk usikkerhet ved tallene fra turdagboken, viser tabellen ganske tydelig at de korteste lange reisene rapportert fra siste måned er klart underrapportert. Antallet fra 50 til 100 km kan vi ikke legge særlig vekt på, fordi disse reisene egentlig ikke skulle rapporteres i denne delen av undersøkelsen.

Skal vi tro tallene i tabellen, er imidlertid antallet reiser i intervallet 100 - 150 km underrapportert med ca 20 % i reisevaneundersøkelsens del for lange reiser siste måned. I intervallet 150 - 200 km er underrapporteringen for reiser siste måned lavere, men likevel 7 %. Først fra 200 km og oppover er det turdagboken som har et lavere antall turer enn delen for lange reiser siste måned. Dette skyldes sannsynligvis at mange av de som reiste langt dagen før intervjuet ikke var kommet hjem enda når de ble ringt opp av intervjumannskapet.

Tallene i tabellen viser det som må karakteriseres som en alvorlig feilkilde i reisevaneundersøkelsen. Dette kommer noe overraskende på oss. For å korrigere for dette i estimeringen av modellene, er det forsøkt å vekte observasjonene i datafilen for lange reiser slik at avstandsfordelingen blir mer korrekt. Dette kommer vi tilbake til i kapittel 6.



TØI rapport 606/2002

Figur 3.5: Sammenlikning av avstandsfordeling for reiser fra 40 til 150 km fra turdagboken og for lange reiser

Tabell 3.6: Sammenlikning av antall lange reiser i turdagboken og i datafilen for lange reiser (siste måned)

	Antall lange reiser i turdagboken	Lange reiser i turdagboken oppblåst til månedstall	Lange reiser siste måned (t/r)	Forhold mellom datakilder: siste måned/turdagbok	estimert frafall: (+=fracfall siste mnd, -=fracfall turdagbok)
50-100	465	13 950	1 376	0.1	12 574
100-150	107	3 210	2 596	0.8	614
150-200	49	1 470	1 366	0.9	104
200-250	33	990	1 006	1.0	-16
250-300	14	420	488	1.2	-68
300+	54	1 620	2 820	1.7	-1 200
Sum	722	21 660	9 652	0.4	12 008

TØI rapport 606/2002

Tabell 3.7: Definisjoner i de nasjonale transportnettene

Nettverk	Nodenummer	Lenketyper	Modes	Funksjoner
Vegnett	5000-939999 990000-999999	1=Europaveg 2=Riksveg 3=Fylkesveg 999=Sonetilknytninger	a= bil b= buss p= gang (5/km/t) g= tilbringer tog (10 km/t) l= tilbringer fly (40 km/t)	vd11: sonetilknytninger vd30, vd40, vd50, vd60, vd70, vd80, vd90: funksjoner som gir kjøretider tilsvarende hastighetsgrensen
Ferger	5000-939999 990000-999999	1=Europaveg 2=Riksveg 3=Fylkesveg	a= bil b= buss y= ferge som tilbringer (20 km/t)	vd1-vd10: funksjoner som gir overfartstid og ventetid avhengig av frekvens.
Bussnett	5000-939999 990000-999999	1=Europaveg 2=Riksveg 3=Fylkesveg	b= buss p= tilbringer (5 km/t)	ft99: avhengig av skiltet hastighet på veg
Tognett	952121-955219	950=toglenker 952=tilknytningslenker	t= vanlig tog e= ekspresstog g= tilbringer (10 km/t)	ft1-ft97: kjøretid i minutter
Flynett	958001-958055 959001-959055	956=flylenker 957=lenker mellom ankomst og avgang 958=tilknytningslenker fra veg til avgang 959=tilknytningslenker fra ankomst til veg	f= fly l= tilbringer (40 km/t) p= transfer (5 km/t)	ft1-ft97: Kjøretid i minutter
Båtnett	970000-970561	997=båtlenker 970= tilknytningslenker	s= rutebåt h= hurtigbåt p= tilbringer (5 km/t)	ft98: OBS: "length" = tid på disse lenkene (s), ft1-ft97: Kjøretid i minutter (h)

TØI rapport 606/2002

## 3.2 Nettverksdata

### 3.2.1 Nye nasjonale nettverk

Nye nasjonale nettverk er laget hovedsakelig med utgangspunkt i de TØI-utviklede applikasjonene GISNETT og KOLLNETT. Disse to applikasjonene benytter data fra hhv ELVEG (levert av Transport Telematikk AS) og elektroniske data fra Rutebok for Norge (levert av Norsk Reiseinformasjon AS).

Det er etablert nettverk som skal tilsvare infrastruktur i 1997/98 for vegtrafikk, fly, båt og tog. Nettverkene trafikkeres av biler, bussruter, båtruter, flyruter og togruter. Kollektivtilbudet skal representere

en gjennomsnittlig dag ved årsskiftet 1997/98. Tabell 3.8 gir en oversikt over elementer og definisjoner i de ulike nasjonale nettverkene.

Ved hjelp av de etablerte nettverksmodellene beregnes reisetider (ombordtid, frekvenser, tilbringer-tid, m.m.) og reisekostnader (kjørekostnader, billettpriser fergekostnader, m.m.) som skal benyttes til estimeringen av nye langdistans modeller.

Det er viktig å understreke at nettverksmodeller av denne type aldri vil være perfekte. For det første er beregningsmetodene for trafikkfordeling basert på forutsetninger som sjelden vil være oppfylt. For det andre vil såpass store mengder av data sannsynligvis alltid inneholde mer eller mindre alvorlige feil som

kan være ganske ressurskrevende å eliminere. Arbeidet med drifting og utvikling av slike databaser vil derfor være en kontinuerlig prosess, hvorav det nå er laget en første versjon på nasjonalt nivå.

### 3.2.2 Generering av nettverksdata

God kvalitet også på fordelingen av trafikk i de nye nettverkene er av sentral betydning for å få korrekte parametere for de ulike tids- og kostnadsvariable når transportmodellene skal estimeres. I dette avsnittet skal vi bare foreslå noen preliminære forutsetninger i rutevalgsmoellene. Disse vil benyttes i estimeringsarbeidet, men man bør endre disse hvis det senere oppdages problemer eller vi på andre måter finner ut at det ikke passer å benytte disse.

Det er laget makroer som beregner de data vi trenger fra nettverkene. Det er ett sett som beregner kommune – kommune matriser og ett sett for beregninger på NTPL-sone nivå. Makroene heter:

For kommune – kommune matriser:	For NTPL – NTPL matriser:
trd5_bil.mac	NTPL5_bil.mac
trd5_buss.mac	NTPL5_buss.mac
trd5_baat.mac	NTPL5_baat.mac
trd5_fly.mac	NTPL5_fly.mac
trd5_tog.mac	NTPL5_tog.mac

Disse makroene beregner følgende data og skriver dem ut til filer:

#### For kollektivtransport:

- Kjøretid
- Tilbringetid
- Total ventetid
- Første ventetid
- Antall påstigninger
- Antall ferger på tilbringerstrekning
- Billettpris (full pris)
- Billettpris (månedskort tog)
- Eventuelle fergekostnader

#### For bil (to sett avhengig av om reisen betales av andre eller av bilisten selv):

- kjøretid
- kjørekostnad (km avhengig)
- fergekostnad og bomkostnad fører
- fergekostnad og bomkostnad passasjer
- antall ferger

For bilreiser vil vi foreslå å benytte to forskjellige fordelingsprinsipper. Hovedgrepet er å skille mellom bilister som selv betaler for reisen og bilister som får dekket alle kostnadene (tjenestereiser, noen typer arbeidsreiser). Disse to grupper vil ha forskjellig vegvalg i vegnettet. For den første gruppen benyttes et generalisert kostnadsbegrep hvor både tidsbruk og monetære utlegg inngår, mens den siste gruppen bare opplever reisetid som reisemotstand i vegvalget. I det første tilfellet må reisetid og -kostnader vektet sammen ved hjelp av en tidsverdi som vi i lys av Killi (1999) setter til kr 40,- per time. I tillegg må vi ha en kilometerkostnad for bilkjøring som vi på vegvalgsnivå kan sette til 1 kr/km. En såpass lav kilometerkostnad kan tolkes som en *adferdsrelevant* kilometerkostnad når bilister skal velge reiserute. I tillegg kommer eventuelle bompenger og fergebilletter. Dette innebærer vi definisjonsmessig får følgende sammenheng for bilister som selv betaler reisen:

$$40 \text{ kr} = 40 \text{ km} = 60 \text{ min}$$

eller

$$1 \text{ kr} = 1 \text{ km} = 1,5 \text{ min}$$

eller

$$0,67 \text{ kr} = 0,67 \text{ km} = 1 \text{ min}$$

Når det gjelder kollektivreiser er veiing av reisetidskomponenter avgjørende for rutevalget. Det gjelder her å benytte et sett med vektforaktorer og assignmentsparametre for øvrig som gir realistiske rutevalg i de kollektive nettverkene. Det er viktig at de rutevalgene som beregnes for de ulike transportmåtene stemmer bra overens med de valg informantene i reisevaneundersøkelsen har gjort og rapportert. Hvis man ikke klarer dette vil man kunne få ”ukorrekte” verdier på de parametere som senere skal estimeres.

Tabell 3.8 viser de verdier som benyttes for lange reiser i SAMPERS-modellene. Dette er verdier som man i Sverige har benyttet lenge. TØI har gjennomført analyser for fly og for tog med sikte på å fastsette tilsvarende verdier som passer til norske forhold og nettverk<sup>2</sup>. Analysene er gjennomført ved å legge matriser ut på nettverkene med ulike forutsetninger, for så å sammenlikne med statistikk for kommet og reist på flyplassene og passasjertellinger på tog. Disse verdiene er vist i Tabell 3.9. Som vi ser, er disse verdiene en del forskjellig fra de svenske.

<sup>2</sup> Test av ulike forutsetninger for rutevalgsmoeller for fly (PT/1422/00), Testing av rutevalgsmoeller for tog på nasjonalt nivå (PT/1432/00)

Tabell 3.8: Assignmentsparametre for lange reiser i SAMPERS etter transportmiddel<sup>3</sup>

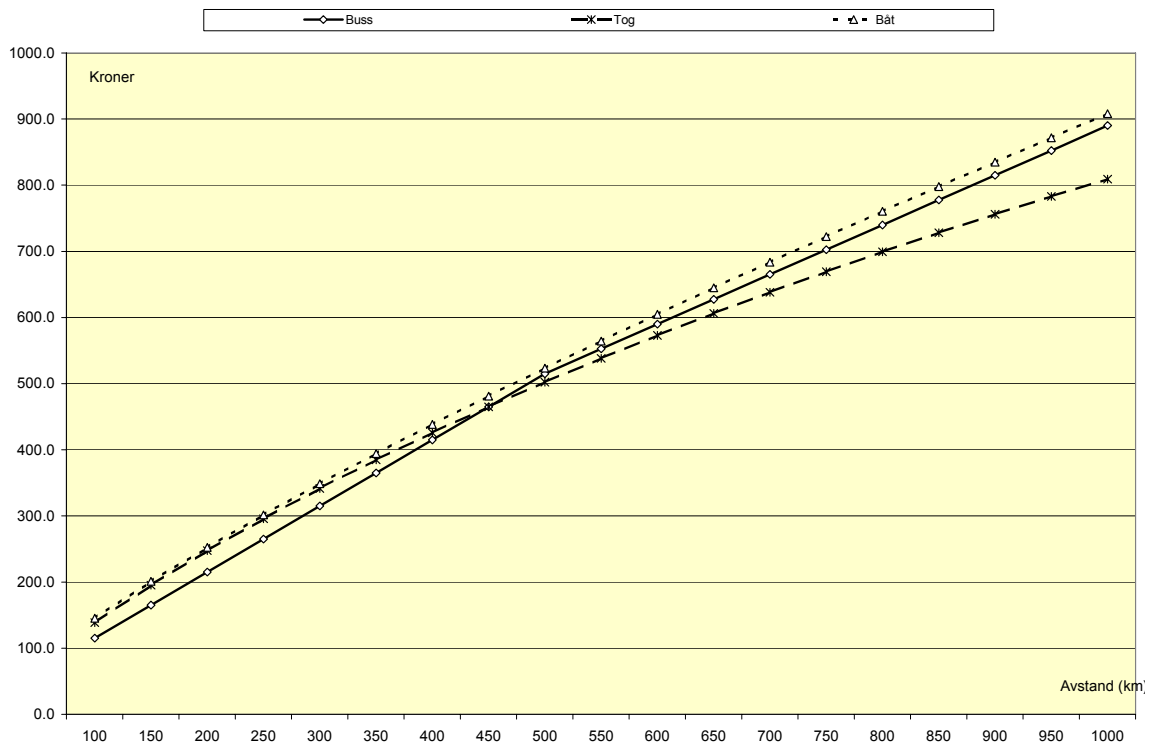
	Fly	Tog (tjenestereiser)	Tog (private reiser)	Buss
Tilbringerhastighet	30 km/t	6 km/t	6 km/t	5 km/t
Påstigningsulempe	15 min	10 min	20 min	60 min
Ventetidsfaktor	0.5	0.5	0.5	0.5
Vektfaktor for tilbringertid	5.0	5.0	5.0	3.0
Vektfaktor for påstigningsulempe	2.0	5.0	5.0	3.0
Vektfaktor for ventetid	2.0	2.0	2.0	2.0

TØI rapport 606/2002

Tabell 3.9: Preliminært anbefalte assignmentsparametre for lange reiser i NTM5

	Fly	Tog	Buss	Båt
Tilbringerhastighet	40 km/t	10 km/t	5 km/t	5 km/t
Påstigningsulempe	30 min	40 min	40 min	40 min
Ventetidsfaktor	0.2	0.5	0.5	0.5
Vektfaktor for tilbringertid	1.0	2.0	2.0	2.0
Vektfaktor for påstigningsulempe	1.0	1.0	1.0	1.0
Vektfaktor for ventetid	1.0	1.0	1.0	1.0

TØI rapport 606/2002



TØI rapport 606/2002

Figur 3.6: Estimerte sammenhenger mellom billettpriser og reiseavstand for buss-, båt-, og togtransport

<sup>3</sup> Kilde: Peter Roming (SIKA)

For fly er den største forskjellen knyttet til tilbringertid. I Sverige benyttes 30 km/t med vektfaktor på 5. Dette innebærer en tidsbruk på 10 minutter på 1 km. De norske forutsetningene med hastighet 40 km/t og vektfaktor 1 innebærer en tidsbruk på 1,5 minutt på samme strekning. Påstigningsulempe og vektfaktor utgjør til sammen identiske forutsetninger, mens ventetid blir vektlagt vesentlig lavere i Norge. En relasjon med 4 avganger per døgn vil i det svenske nettverket få en veiet ventetid på 180 minutter, mens tilsvarende i det norske vil være 36 minutter. Vi har forsøkt med de svenske forutsetningene også i det norske nettverket, men med svært dårlig resultat. Det kan være mange årsaker til dette, for eksempel forskjeller i nettverkskoding, problemer med turmatrisen, eller rett og slett at norske flytrafikanter har andre preferanser enn de svenske.

For tog er forutsetningene mer like. I Sverige benyttes forutsetninger som gir en noe lavere veiet tilbringertid og en noe høyere veiet ventetid. For buss og båt har vi foreløpig valgt å benytte samme forutsetninger som for tog, med unntak av halvert tilbringerhastighet. Som nevnt vil forutsetningene angitt i tabellen kunne endres hvis det oppstår problemer med nettverksdataene i estimeringen.

### 3.2.3 Reisekostnader for kollektivtransport

For de kollektive transportmåter er Takstregulativene i Rutebok for Norge 1998, benyttet til å estimere sammenhenger mellom avstand (tid for hurtigbåt) og billettpris (uten moderasjoner). De estimerte modellene gir tilnærmet perfekt føyning til datasettene som er benyttet. De estimerte sammenhengene vises under.

$$K_{\text{buss}} = 15 + km_{\text{buss}} - 0.25 (km_{\text{buss}} > 500) ( km_{\text{buss}} - 500)$$

$$K_{\text{tog}} = -3.78 + 3.79 km_{\text{tog}} - 1.492 (km_{\text{tog}})^{1.1}$$

$$K_{\text{båt}} = 3.52 (T_{\text{båt}})^{0.795}$$

Billettprisene beregnes i makroene i EMMA på grunnlag av matriser for reiseavstand (tid for båt) ombord transportmiddelet. Billettprisene vil dermed avhenge av kvaliteten på rutevalgsberegningene i EMMA. Figur 3-6 viser sammenhengene grafisk (ved en antatt seilingshastighet på 30 knop med båt). Som vi ser er det svært små forskjeller i fullpris mellom de tre transportmåter. Vi ser også at tog er den billigste transportmåten på de noe lengre relasjonene.

For fly har vi lenge hatt en matrise med priser for 1997/98. Matrisen er etablert med bakgrunn i informasjon fra konkurransetilsynet (Voldmo 1998).

Denne er kvalitetssikret og ligger til grunn for billettprisanslagene i dette arbeidet.

## 3.3 Sonedata

Det er samlet inn en rekke data, fortrinnsvis på grunnkrets nivå, som beskriver innholdet i hver enkelt av landets om lag 13000 grunnkretser. Disse data er aggregert til NTPL-soner direkte. En del av variablene finnes bare på kommunenivå eller med postnummer. Her er det benyttet data fra en del andre registre og datakilder til å fordele dataene på NTPL-soner. Dataene er mottatt med svært ulikt format og det er derfor laget applikasjoner som bearbeider formateringen til en ensartet inndeling. Følgende data er innhentet og bearbeidet (G = grunnkrets, K = kommune, P = postnummer):

- (G) Grunnkretsinnndeling med endringer
- (G) Areal på grunnkretser (30.12.00)
- (G) Folkemengde etter kjønn og alder (1.1.98)
- (K) Befolkningsframskrivninger kjønn og alder
- (K) Sentralitetsindikatorer
- (G) Ant. bedrifter og ansatte fordelt på næringshovedområde
- (G) Sysselsatte etter bosted
- (K) Sysselsatte etter bosted, etter 10 hovednæringer
- (G) Gjennomsnittlig bruttoinntekt pr pers 17 år og over inntektsåret 97.
- (G) Familiestatistikk (1.1.98)
- (P) Antall hoteller/hotellsenger (20 senger eller mer, 97)
- (K) Hytter og fritidshus (98)
- (K) Antall elever og studenter (97/98)

Sonedata for NTPL-soner er etablert ved å summere grunnkretsdata for alle grunnkretser som inngår i hver av NTPL-sonene. Dessuten er informasjon fra GAB benyttet for å koble hotelldataene (5.1, angitt på postnummer) til grunnkrets. GAB er videre benyttet til å ta ut antall hytter og fritidshus på grunnkrets nivå. Aggregering av disse tallene viser avvik fra SSB-statistikken i 5.2, dette fordi vi definerer et videre begrep av bygninger som kan attrahere fritidsreiser (dessuten tilsvarer tallene omtrent dagens situasjon).

Bedriftsstatistikken (antall bedrifter) er detaljert i to dimensjoner – både på næring og størrelseskategorier. Antall ansatte totalt pr næring er også angitt. Se nedenfor.

Næringer:	Størrelseskategorier:
I alt	Bedrifter i alt
Uoppgitt	Ansatte i alt
A Jord- og skogbruk	Uten ansatte
B Fiske	1 – 4 ansatte
C Oljeutv., bergv.	5 – 9 ansatte
D Industri	10 – 19 ansatte
E Kraft- og vannf.	20 – 49 ansatte
F Bygge- og anl.v.	50 – 99 ansatte
G Varehandel mv.	100+ ansatte
H Hotell, restaur.	
I Samferdsel	
J Finanstjenester	
K Forr.tjen, eiend.	
L Off.adm, forsvar	
M Undervisning	
N Helse-, sosialtj.	
O A sos. og pers. tj.	
P Lønnet husarbeid	
Q Internasj. org.	

For data om inntekt (4.1) er det beregnet et gjennomsnitt over alle grunnkretser som inngår i hver NTPL-sone.

Familiestatistikken (4.2) er inndelt i følgende kategorier:

- 1 Ektepar uten barn
- 2 Ektepar med barn uansett alder
- 3 Ektepar med barn under 18 år
- 4 Ektepar med bare barn 18 år og over
- 5 Ektepar med 1 barn under 18 år
- 6 Ektepar med 2 barn under 18 år
- 7 Ektepar med 3+ barn under 18 år
- 8 Samboere med felles barn uansett alder
- 9 Samboere med felles barn under 18 år
- 10 Samboere med bare felles barn 18 år og over
- 11 Samboere med 1 felles barn under 18 år
- 12 Samboere med 2 felles barn under 18 år
- 13 Samboere med 3+ felles barn under 18 år

Næringskategorier og/eller størrelseskategorier aggregeres opp etter hva som er hensiktsmessig til modellbruk.



## 4 Modeller for biltilgang og førerkort

### 4.1 Segmentering i persontransportmodellen

Persontransportmodellen har i dag en segmentering og rutiner for beregning av segmenter som på grunn av vanskelig tilgjengelig programkode er nesten umulig å gjennomskue. Ikke minst gjelder dette segmentering i forhold til førerkortinnehav og bilhold. Det er også et spørsmål om den segmentering som benyttes er unødig komplisert. I forbindelse med estimering og implementering av nye modeller er det derfor grunn til å se på segmenteringsproblemet fra grunnen av.

### 4.2 Hvorfor segmentere?

I prinsippet kunne vi tenke oss en modell praktisk talt uten segmentering. Det vil si at vi hadde én modell f.eks. for alle personer over 15 år. Med dette utgangspunkt kunne vi ha modeller for ulike reisemål. En slik modell ville imidlertid ikke kunne inneholde variable som beskriver individuelle kjennetegn, men den ville blitt meget enkel å estimere og implementere. Man kunne multiplisere beregnede sannsynligheter for å foreta reiser og valg av destinasjon og reisemåte med antall bosatte i sonen over 15 år. I den utstrekning man skulle ta hensyn til variable som inntekt eller biltilgjengelighet måtte slike variable inngå i modellen som gjennomsnitt for bosatte i sonen. Man ville altså kunne ha en form for geografisk segmentering. Skulle slike variable inngå måtte man også ha statistikk/prognoser for disse variable og man ville ikke komme utenom en eller annen form for modeller for førerkortinnehav og bilhold.

Den enkleste form for segmentering ville være bare å ta hensyn til kjønn og alder. Dette er variable som er lette tilgjengelige. Vi kunne for eksempel ha 3 - 5 aldersgrupper. Sammen med kjønn ville dette gi 6 til 10 segmenter. Beregnede sannsynligheter innenfor disse segmenter kunne så multipliseres med antall personer innenfor de ulike segmenter. Som ovenfor kunne dette suppleres med gjennomsnittsverdier for aktuelle variable for bosatte i sonen i de ulike segmenter.

Sett at vi hadde 8 segmenter definert over kjønn og aldersgrupper. Hvordan skulle man da i modellen skille mellom disse segmenter? Dette kunne gjøres på

ulike måter som beregningsmessig og estimerings-teknisk ville ha litt ulike implikasjoner:

- Man kunne ha 7 dummies i frekvensmodellen slik at antall genererte reiser pr person varierte mellom segmentene.
- Man kunne ha tilsvarende dummies i reise-middelvalgmodellene for en eller flere reisemåter. Dette ville bety at en person i et segment ville ha en større sannsynlighet for å velge en bestemt reisemåte enn en person i et annet segment.
- Man kunne ha interaksjonsvariable som konstrueres ved at transportstandardvariable multipliseres med segment dummies. Dette vil innebære at effekten av pris og/eller tid (eventuelt frekvens) vurderes ulikt av personer i ulike segmenter.
- Man kan ha ulike kombinasjoner av a-c.
- Man kan estimere helt separate modeller for hvert segment. Dette er ikke noe problem hvis man har tilstrekkelig mange observasjoner.

Ved å ha ”segmenteringsdummies” i reise-middelvalgmodellen vil disse også påvirke destinasjonsvalget, men man kan i prinsippet også tenke seg segmentering trukket direkte inn i destinasjonsvalget. Her er imidlertid mulighetene begrenset og det er vanskeligere å tenke seg en fornuftig begrunnelse for dette.

Beregningsteknisk (i form av antall regneoperasjoner) er det tyngre å ha med interaksjonsvariable enn rene dummies. Med rene dummies beregnes nyttefunksjonen for ett segment og for de øvrige segmenter beregnes nyttefunksjonen ved å legge til eller trekke fra en konstant til den først beregnede.

Med interaksjonsvariable er det i realiteten parameterverdier for transportstandardvariable som varierer mellom segmenter.

[  $U_s = a + bX + a_s$  eller  $U_s = a + (b+b_s)X$  eller  $U_s = a + bXb_s$  (ulik verdi på X f.eks. pga av rabattprosent som varierer mellom segmenter, dette kan også modelleres med den mellomste varianten) ]

Det er altså mulig å lage relativt enkle modeller med segmentering på kjønn og alder og samtidig ta hensyn til geografiske forskjeller i form av gjennomsnittsverdier for bosatte i ulike soner og/eller eventuelt

sonespesifikke variable utformet som dummies. Det ville være store praktiske fordelene ved en slik modellstruktur.

#### 4.2.1.1 Hvorfor trenger man eventuelt mer og mer komplisert segmentering?

I bunn og grunn henger behovet for ytterligere og mer komplisert segmentering sammen med tre forhold:

1. Et kjennetegn som f.eks. førerkortinnehav påvirker hvilke alternativer som er tilgjengelig.
2. Aggregeringsproblemet som innebærer at hvis man har undergrupper som ut fra visse kjennetegn har forskjellig adferd eller alternativmengde, så vil bruk av gjennomsnittsverdier for disse kjennetegn – eller at kjennetegnene ikke spesifiseres, men inngår i konstantledd – innebærer at det oppstår feil ved modellen anvendelsen som har karakter av aggregeringsfeil.<sup>4</sup>
3. Ved estimering av en modell bør vi i størst mulig grad inkludere variable med ”god” forklaringskraft når det gjelder observerte valg fordi dette også kan gi bedre estimater på parametere for andre variable.

Hadde det ikke vært for 2 kunne vi i prinsippet tenkt oss et opplegg hvor vi benyttet mange ”segmenteringsvariable” ved estimering. Deretter kunne vi beholde de estimerte parametere for transportstandardvariable og andre variable som ikke hadde noe med segmentering å gjøre, sløyfe de segmenteringsvariable vi ikke ville benytte ved implementering og re-estimere konstantleddet. I tillegg til det generelle aggregeringsproblemet er det to andre problemer med en slik fremgangsmåte.

- a. Det er ikke sikkert at datamaterialet er representativt for totalpopulasjonen når det gjelder fordelingen på segmenter. Da blir den estimerte konstant feil i forhold til anvendelse på totalpopulasjonen.

<sup>4</sup> Sett at vi har to like store grupper med sannsynligheter for å velge et bestemt alternativ på hhv 0, 1 og 0,6. Hvis vi ikke skiller mellom disse to grupper ved estimering og implementering av en modell så vil vi få tilpasset en konstant som gjør at modellen reproducerer den aggregerte markedsandel (P) som er 0,35. Hvis det i modellen inngår en variabel X med parameter a og vi endrer denne variabel så vil vi for en modell hvor gruppene er aggregert få:  $[dP/dX=P(1-P)a]$

$dP/dX=0,35*0,65*a=0,2275*a$

Med en modell hvor det er segmentert på de to grupper vil vi få:

$dP/dX = 0,5*0,1*0,9*a + 0,5*0,4*0,6*a=0,165*a$

Begge varianter vil altså reproducere den aggregerte markedsandel i utgangspunktet, men vi vil få relativt ulike resultater når vi ser på effekten av endringer i variable i ettertid, og modellen uten segmentering gir vesentlig større følsomhet for endring i X.

- b. Over tid kan fordelingen mellom segmenter endres og da blir også det estimerte konstantledd feil og vi kan få en nivåfeil i forbindelse med prognoser.

Punkt a. kan vi i noen grad komme rundt i forbindelse med etterkalibrering av modellen og en slik etterkalibrering vil ofte måtte gjøres uansett selv om vi ikke alltid har gode data å kalibrere mot.

Konklusjonen på denne lille drøfting er at hvis vi finner ”segmenteringsvariable” med god forklaringskraft i forbindelse med estimering av modeller så bør vi i størst mulig grad også forsøke å ha med disse variable i forbindelse med implementeringen selv om mye og komplisert segmentering er noe ”herk” fra et praktisk synspunkt.

I praksis kan vi nok regne med at aggregeringsproblemet som oftest er av mindre betydning enn det vi kan få inntrykk av fra eksemplet i fotnoten. Dessuten vil man kunne ha unøyaktigheter i segmentering både i forhold til ulike geografiske områder og over tid. Altså: Det er i prinsippet riktig å segmentere, men vi har unøyaktighet når det gjelder fordelingen av personer på ulike segmenter. Den økte nøyaktighet vi oppnår ved utstrakt segmentering kan derfor bli fiktiv dersom vi driver dette for langt.

## 4.3 Segmentering relatert til bilhold og førerkort

Segmentering etter kjønn og alder er normalt uproblematisk. Segmentering blir imidlertid mer problematisk for en del andre variable som f.eks. kan være relatert til husholdningskarakteristika eller andre forhold hvor vi bare har opplysninger fra RVU, men ikke statistikk eller prognoser på detaljert nivå. Da må vi ha en eller flere ”modeller” for å splitte befolkningen i ulike segmenter eller eventuelt gjøre eksplisitte forutsetninger om hvordan befolkningen er fordelt på de ulike segmenter.

Vi må imidlertid huske på at siden det vi primært skal modellere dreier seg om reiser, så stammer behovet for segmentering fra de modeller vi har estimert for reisemiddelvalg, destinasjonsvalg og turfrekvenser. Bruker vi her få personlige kjennetegn utover kjønn og alder som forklaringsvariable blir også behovet for segmentering lite.

Når det gjelder forhold knyttet til førerkortinnehav og bilhold ser det ut til at det er to aspekter som er viktig ved segmentering.

1. *Hvorvidt en person har førerkort eller ikke.* Uten førerkort er ikke ”bilfører” tilgjengelig som reisemåte.

Det er imidlertid mulig at antall biler og eventuelt antall førerkort i husholdningen har betydning for sannsynligheten for at vedkommende skal være "bilpassasjer". Behovet for å ha en segmenteringsvariabel som skiller personer *uten førerkort* i to grupper er derfor et empirisk spørsmål. Selv om vi ikke skiller mellom bilfører og bilpassasjer vil det allikevel også være slik at personer med førerkort og "god" tilgang til bil vil ha stor sannsynlighet for å reise med bil i forhold til en person uten førerkort og tilgang til bil i husholdningen.

2. Hvis vedkommende har førerkort, kan vi ha følgende alternativer:

- Husholdningen disponerer ikke bil.

(I prinsippet kan vedkommende låne eller leie bil. Det vil si bilfører er ikke logisk utelukket som reisemåte, og det er mulig at vi kan for praktiske formål kanskje se bort fra bilturer som bilfører foretatt av slike personer. Dette er også et empirisk spørsmål.)

- Husholdningen har 1 fk og 1 bil, dvs. biltilgjengelighet 1/1
- Husholdningen har 2 fk og 1 bil, dvs. biltilgj  $\frac{1}{2}$
- Husholdningen har 3+fk og 1 bil, dvs. biltilgj  $\leq 1/3$
- Husholdningen har 2 fk og 2+ biler, dvs. biltilgj =1
- Husholdningen har 3+fk og 2 biler, dvs. biltilgj  $\leq 2/3$
- Husholdningen har 3+fk og 3+ biler, dvs. biltilgj  $\approx 1$

Ut fra dette skulle det man tro at når det gjelder segmentering relatert til førerkort og bilhold, så vil man klare seg med 3-7 segmenter. Minimum på 3 fremkommer hvis vi grupperer sammen de som ikke har førerkort med dem som har førerkort, men ikke bil i husholdningen og sier at disse ikke kan reise som bilfører. For de øvrige kunne vi skille mellom:

- hfk=biler
- hfk**≤**biler

Hvis man differensierer i siste gruppe og i tillegg for fk=0 skiller mellom bil=0 og bil**≥**1 får man 7 grupper.

I den nye langdistansemodellen skiller vi ikke mellom bilfører og bilpassasjer som reisemåte. Vi har imidlertid en dummy for førerkortinnehav som gir større sannsynlighet for å reise med bil. I tillegg har vi med en variabel for biltilgjengelighet.

Problemet er så hvordan vi skal fordele antall personer i en befolkningsgruppe på disse kategorier, eventuelt i kombinasjon med andre segmenteringsvariable som kan være mer eller mindre korrelert med

segmenteringen når det gjelder førerkort og bilhold. Det er i denne forbindelse vi trenger modeller for førerkortinnehav og bilhold og en metode for å fordele personer i ulike befolkningsgrupper på husholdningstyper.

Det siste gjøres i dag (NTM 4c) med en prosedyre som er programmert i GAUSS. Her beregnes antall husholdninger av ulike typer med bibetingelser som gjør at denne fordeling er konsistent med befolkningsprognosen og samtidig mest mulig lik en "prior"-fordeling basert på FoB 1990.

Antall personer i befolkningsgruppe "i" som tilhører husholdningstype "j" ( $N_{ij}$ ) kan så beregnes som  $N_{ij} = a_{ij}H_j$ . Koeffisientene  $a_{ij}$  inngår i en matrise "A" som også stammer fra FoB 1990 og benyttes ved beregning av antall husholdninger av ulike typer. Dette er nærmere beskrevet i arbeidsdokumentet TRU/0437/94. Beregnet antall husholdninger tilfredsstiller ligningsystemet  $AH=B$  hvor B er en kolonnevektor ( $n \times 1$ ) med antall personer i ulike aldersgrupper fordelt på kjønn og H er en kolonnevektor ( $h \times 1$ ) med antall husholdninger av ulik type og A er ( $n \times h$ ).

Sannsynligheten for at en vilkårlig person i gruppe "i" tilhører husholdningstype "j" kan da beregnes som  $P(i,j) = N_{ij}/N_i$ .

Siden vi ikke har klart å trenge gjennom den eksisterende kode når det gjelder segmentering er det følgende basert på hva jeg vil betrakte som en naturlig fremgangsmåte:

La fk være en dummy for førerkortinnehav for en person og hfk en variabel som angir antall førerkort i en husholdning. La videre b stå for antall biler i husholdningen og s for segment mht biltilgang. La oss videre anta at s består av følgende segmenter:

- s=1 : Personer uten førerkort og b=0
- s=2 : Personer uten førerkort og b  $\geq 1$
- s=3 : Personer med førerkort og b=0
- s=4 : Personer med førerkort og b  $\geq$ hfk
- s=5 : Personer med førerkort og b<hfk

Det siste segment kan også deles i undersegmenter avhengig av forholdstallet b/hfk, men dette er trolig unødvendig.

Det vi da trenger er et opplegg som gir:

$P(s=s^*|ij)$  ( $ij$  =person i befolkningsgruppe "i" som tilhører husholdningstype "j")

Det enkleste hadde her egentlig vært å ha en egen modell som direkte fordelte personer på disse

segmenter. Da kunne man unngå å koble denne fordelingen direkte mot separate modeller for førerkort og bilhold og en komplisert konstruksjon av segmenter.

En modell som foretar en slik fordeling mellom segmenter må ha en "nyttefunksjon" for hvert segment. For segment  $S=1$  kan denne være null. For de øvrige nyttefunksjoner ville vi da få en kombinasjon av alternativspesifikke konstanter og parametre for inntekt og variable som beskriver bostedssone, individuelle kjennetegn og husholdningskjennetegn utover inntekt.

Det man mister med dette opplegg i forhold til det eksisterende er det forhold at sannsynligheten for å tilhøre en husholdning med bil i utgangspunktet er uavhengig av faste og variable kostnader ved å ha bil. Dette ivaretas i dag med bilholdsmodellen og bør fortrinnsvis også ivaretas i en ny modell.

I dag foretas også en fremskrivning av førerkort-andeler basert på kohorteffekter og dette bør man også søke å ivareta i forbindelse med en ny modell. Sett at vi har en modell som for hvert segment ( $ij$ ) beregner sannsynligheter:

$P(s=1), \dots, P(s=S)$  slik at sannsynlighetene summerer seg til 1. Kohorteffekten når det gjelder førerkort innebærer at vi for enkelte segmenter ( $i,j$ ) bør justere fordelingen slik at  $P(s=1)$  og  $P(s=2)$  reduseres noe på bekostning av en motsvarende økning av de øvrige sannsynligheter. Denne justering må være slik at sannsynlighetene fremdeles summerer seg til 1. Justeringen kan gjøres direkte på fordelingen eller på parametere i nyttefunksjonene.

Tilsvarende har vi at endringer i faste eller variable kostnader kan påvirke bilholdet og dette må i så fall ivaretas ved en justering av  $P(s=2)$ ,  $P(s=4)$  og  $P(s=5)$  og en motsvarende justering av de øvrige sannsynligheter slik at de fremdeles summerer seg til 1. Vi bør derfor opprettholde en bilholdsmodell som kan benyttes i forbindelse med en slik justering. Her er det mer problematisk å gjøre justeringen direkte på nyttefunksjonene, men det går sikkert å finne et annet opplegg.

I NTM 4c opererer vi med 31 husholdningstyper og 10 grupper for kombinasjon av kjønn og alder. Antall segmenter er imidlertid mye lavere enn 310 fordi ikke alle befolkningsgrupper inngår i alle husholdningstyper. Antall segmenter skulle derfor kunne bli overkommelig. Inndelingen i husholdningstyper er i dag også basert på at man

Fordelingen av IO i den file som er laget for estimering av bilholdsmodellen og hvor vi har kuttet ut observasjoner hvor ingen har førerkort, inntekts-

opplysninger m.m. mangler, har følgende fordeling på de aktuelle segmenter:

Tabell 4.1: IO i RVU 1997/98 fordelt etter førerkort og biltilgang i husholdningen<sup>1</sup> (Hushold med førerkort, silt)

Segment	Antall IO	Prosent	Variable <sup>2</sup>
1 (fk=0, b=0)	59 (17)	1,3	bilf=na
2 (fk=0, b≥1)	151 (34)	3,2	bilf=na, $\delta_1=1$ (bilpass)
3 (fk=1, b=0)	391	8,3	$\delta_2=1$ (bilf)
4 (fk=1, b≥hfk og b ≥1)	2320	49,5	$\delta_1=1$ (bilpass)
5 (fk=1, b≤hfk og b ≥1)	1767	37,7	$\delta_1=1$ (bilpass), $\delta_3=1$ (bilf)
Sum	4688 (51)	100,0	

1) Tallene i parentes gjelder personer under 18 år 2)  $\delta_1$  = dummy i nyttefunksjon for bilpassasjer  $\delta_2$  = dummy i nyttefunksjon for bilfører  $\delta_3$  = dummy i nyttefunksjon for bilfører.

TØI rapport 606/2002

Som vi ser, gir dette grunnlag for 4-dummies hvorav den ene sørger for at bilfører ikke er tilgjengelig som reisemåte, dvs. den inngår ikke i selve nyttefunksjonene.

Tabell 4.2: IO i RVU 1997/98 fordelt etter førerkort og biltilgang i husholdningen<sup>1</sup> (alle records uveiet, ved missing på antall biler er forutsatt én bil)

Segment	Prosent	Variable <sup>2</sup>
1 (fk=0, b=0)	8,45	bilf=na
2 (fk=0, b≥1)	11,50	bilf=na, $\delta_1=1$ (bilpass)
3 (fk=1, b=0)	5,95	$\delta_2=1$ (bilf)
4 (fk=1, b≥hfk og b ≥1)	41,75	$\delta_1=1$ (bilpass)
5 (fk=1, b≤hfk og b ≥1)	32,35	$\delta_1=1$ (bilpass), $\delta_3=1$ (bilf)
Sum	100,00	8838 records

TØI rapport 606/2002

Tabell 4.3: IO i RVU 1997/98 fordelt etter førerkort og biltilgang i husholdningen<sup>1</sup> (alle records veiet m/WVAR, miss på antall biler er forutsatt en bil)

Segment	Prosent	Variable <sup>2</sup>
1 (fk=0, b=0)	5,82	bilf=na
2 (fk=0, b≥1)	16,01	bilf=na, $\delta_1=1$ (bilpass)
3 (fk=1, b=0)	3,87	$\delta_2=1$ (bilf)
4 (fk=1, b≥hfk og b ≥1)	36,69	$\delta_1=1$ (bilpass)
5 (fk=1, b≤hfk og b ≥1)	37,61	$\delta_1=1$ (bilpass), $\delta_3=1$ (bilf)
Sum	100,00	8838 records

TØI rapport 606/2002

Tabell 4.3 gir trolig den riktigste fordeling på antall segmenter. Her har vi imidlertid også med personer

under 18 år. Tabell 4.4 viser fordelingen på segmenter for personer 18+ år.

Tabell 4.4: IO over 18 år i RVU 1997/98 fordelt etter førerkort og biltilgang i husholdningen<sup>1</sup> (alle records veiet m/WVAR, miss på antall biler er forutsatt en bil)

Segment	Prosent, uveiet	Prosent, veiet (WVAR)	Variable <sup>2</sup>
1 (fk=0, b=0)	8,16	5,24	bilf=na
2 (fk=0, b≥1)	5,58	6,76	bilf=na, δ <sub>1</sub> =1 (bilpass)
3 (fk=1, b=0)	6,41	4,35	δ <sub>2</sub> =1 (bilf)
4 (fk=1, b≥hfk og b≥1)	44,99	41,30	δ <sub>1</sub> =1 (bilpass)
5 (fk=1, b≤hfk og b≥1)	34,86	42,34	δ <sub>1</sub> =1 (bilpass), δ <sub>3</sub> =1 (bilf)
Sum	100,00	99,99	8202 records

TØI rapport 606/2002

Ifølge kolonne 3, som burde gi et tilnærmet korrekt resultat for totalbefolkningen, er det altså vel 16 % av befolkningen over 18 år som ikke har førerkort og/eller ikke bil i husholdningen.

Hvis vi skal estimere modeller som fordeler befolkningen på disse segmenter bør vi trolig ha tre separate modeller:

1. For personer i husholdninger med 1 person over 18 år
2. For personer i husholdninger med 2 personer over 18 år
3. For personer i husholdninger med 3+ personer over 18 år

Modellene bør bare estimeres for personer som er 18 år og eldre. Avhengig av hvilke variable som inngår i de ulike modeller, trenger vi et opplegg som fordeler personer på av forskjellig kjønn og alder på husholdningsgruppene 1-3 med undergrupper.

Arbeidsdokument PT/1400/00 (JR) inneholder allerede modeller for førerkortinnhav estimert på individnivå i tillegg til modeller for antall førerkort i husholdninger av forskjellig type. Ved å utvide ”individmodellene” med flere undersegmenter og re-estimere kan man lett få estimert en ny modell.

## 4.4 Befolkning i nye soner

Utgangspunktet er at vi har estimert nye modeller for reismiddelvalg, destinasjonsvalg og turfrekvenser. I tillegg er det en modell som fordeler personer på 5 segmenter når det gjelder førerkortinnhav og

biltilgang. Disse modeller skal implementeres med NTPL-soner som geografisk enhet.

Et første problem her er at befolkningsprognosene bare går på kommuner. Selv om vi har statistikk på NTPL-soner så utarbeides det ikke ”offisielle” prognoser på dette nivå. Vi må derfor ha en metode som på en konsistent måte fordeler den kommunale prognose mellom NTPL soner. Dette vil bl.a. bety at når vi summerer over NTPL-soner i en kommune så skal vi treffe prognosetallene for kommunen fordelt på kjønn og aldersgrupper.

Ideelt sett skulle denne fordeling bli påvirket av kommunale utbyggingsprogrammer m.m., men å basere fordelingen på informasjon fra de enkelte kommuner vil skape mange praktiske problemer og det blir uansett meget usikkert når vi skal lage prognoser med 20-25 års tidshorisont.

Det er derfor mest aktuelt å bruke en metode som fordeler befolkningen relativt mekanisk. Det minimum av informasjon som man her bør benytte i tillegg til dagens befolkningsstruktur i de enkelt NTPL-soner er at dagens befolkning blir eldre. Siden det neppe er aktuelt å ha finere inndeling enn 5 års aldersgrupper og 5 års intervaller mellom prognoseår, så skulle dette passe bra. Formelt trenger vi da den prosedyre som gir:

$$B_{ks}^T = f(B_{ks}^{T-1}, B_k^T)$$

der

$B_{ks}^T$  = en vektor med befolkning på alder og kjønn på tidspunkt ”T” i sone ”s” i kommune ”k”.

$B_k^T$  = en vektor med samme dimensjon for kommunen (befolkningsprognosen)  
Som bibetingelse har vi:

$$\sum_s B_{ks}^T = B_k^T$$

Antall personer 0-5 år bør vi trolig fordele i henhold til antall kvinner i aldersgruppene opp til 40 år etter en eller annen formel.

Når man har befolkningen fordelt på kjønn og alder i ulike soner blir et neste skritt å segmentere befolkningen i ulike grupper mht kjønn og alder videre på husholdningstyper og andre variable som benyttes i modellene. Hvilke variable dette dreier seg om er foreløpig ikke helt klart, selv om modellestimeringen hittil har gitt visse holdepunkter.

## 4.5 Generelle prinsipper for segmentering

### 4.5.1 Generelt

Generelt bør vi ta en gjennomgang av estimerte modeller og bestemme den segmentering som skal benyttes ved modellimplementering. Poenget med segmentering er altså å unngå aggregeringsfeil både når det gjelder fordeling i en 0-situasjon og når det gjelder prediksjon av endringer. For øvrig er segmentering bare en komplikasjon som øker kompleksitet og regnetid i modeller.

Ulike segmenteringsprinsipper og aggregeringsproblemet generelt er behandlet i Ben-Akiva og Lerman (1985), kap 6. Den følgende bygger delvis på dette.

Et segmentering som har ”liten” betydning for aggregerte virkninger av å endre en ”policyvariabel” bør i utgangspunktet kunne sløyfes. Et mulig unntak er hvis det dreier seg om et markedssegment som det er knyttet stor politisk interesse til f eks av fordelingspolitiske grunner.

Normalt vil det være 2 kriterier som hver for seg eller i kombinasjon tilsier at man bør/ikke bør segmentere:

1. Segmentet eller segmentene utgjør en relativt stor andel av det totale marked. Har man ett segment som utgjør f eks 95 % av totalen og et annet som utgjør 5 %, så vil denne segmentering neppe ha særlig stor betydning for totalresultater. Total-effekten av å endre en ”policyvariabel” vil normalt domineres kraftig av effektene for det største segment.
2. Hvis segmenteringsvariabelen - som vanligvis er en dummy variabel - har ”lav” absoluttverdi på parameteren, så indikerer dette små adferdsforskjeller mellom segmenter selv om parameteren er meget signifikant i form av høy t-verdi. ”Lav” må her forstås slik at variabelen har *relativt* liten betydning for verdien på nyttefunksjonene. Hva dette innebærer f eks i form av prosentvis påslag eller fratrukk i ulike nyttefunksjoner kan man få en formening om med litt eksperimentering.

I tillegg til disse to kriterier må det også benyttes et visst pragmatisk skjønn. Segmenteringsvariable som det er vanskelig eller umulig å lage særlig gode prognoser for bør kunne sløyfes ved implementering. Segmentering vil da bare gi en fiktiv økning i presisjon. *Dette betyr at når alle modeller er estimert så bør man ta en vurdering av mulige forenklinger og re-estimere modellene samtidig som vi låser parametrene på de variable vi vil beholde i modellen og estimerer nye konstantledd og en skaleringsfaktor.*

En viktig konklusjon i Ben-Akiva og Lerman (1985), kap 6 er at man bør ivareta segmentering som har betydning for settet av valgbare alternativer. Det vil si at bl.a. segmentering etter førerkort- og biltilgang er viktig for å unngå aggregeringsfeil.

Når vi segmenterer etter kategoriske variable bør vi kunne forhåndsberegne faste tillegg (positive eller negativ) til konstantleddet i ulike nyttefunksjoner, dvs. at vi ved implementering har en tabell av nedenstående type:

Tabell 4.5: Tillegg/fradrag i nyttefunksjon

Segment	Andel av befolkningen	Nyttefunksjon				
		A	B	..	..	M
1	$a_1$					
2	$a_2$					
..					$\Delta_{sm}$	
..						
S	$a_s$					

TØI rapport 606/2002

Dette vil redusere antall regneoperasjoner og redusere kjøretiden på modellen.

Når det gjelder segmentering er det generelt viktig hvor segmenteringen skjer. Vi kan her skille mellom det vi litt løst kan betegne som ”logsum-segmentering” og annen segmentering. En segmentering som influerer på log-summene må trekkes med gjennom nesten alle beregninger i den forstand at modellen må kjøres en full ”loop” over hvert segment. Dette innebærer at eksekveringstiden vil være tilnærmet proporsjonal med antall ”logsum-segmenter”.

Dette kan illustreres ved tabell 4.6:

Tabell 4.6: Kryssgruppering av "demografiske segmenter" (D) og "logsumsegmenter" (L)

"Demografiske segmenter"	"Logsum-segmenter"				Antall personer
	1	2	k	L	
1	..	..	$a_{1l}f_{1k}$	..	$N_1$
2			..	..	$N_2$
.			..	..	.
d	$a_{d1}f_{d1}$	..	$a_{dl}f_{dk}$	..	$N_d$
.	..		..	..	.
.	..		..	..	.
.	..		..	..	.
.	..		..	..	.
D	..		$a_{Dl}f_{Dk}$	..	$N_D$
Antall reiser:	$R_1 = \sum_d a_{d1} f_{d1} N_d$	$R_2 = \sum_d a_{d2} f_{d2} N_d$	$R_k = \sum_d a_{dk} f_{dk} N_d$	$R_L = \sum_d a_{dL} f_{dL} N_d$	

$a_{dl}$  = andel av personene i demografisk segment "d" som skal tilordnes logsum "l".  $f_{dl}$  = reisefrekvens for personer i segment "d" med logsum "l". "Demografisk segment" behøver ikke nødvendigvis bare dreie seg om kjønn og alder. Det kan også dreie seg om husholdningstype eller andre forhold. Holder vi oss bare til demografi er det imidlertid lett å få nøyaktige anslag på antall personer i hvert segment.

TØI rapport 606/2002

Siste linje i tabellen vil da være antall reiser som skal fordeles på destinasjon og reisemåte med logsum "k". Utregning av denne linje vil normalt gå meget raskt, det er beregning av sannsynligheter for destinasjons- og reisemiddelvalg som er krevende mht maskintid og dette må gjøres L ganger.

Størrelsen på D, dvs. antall "demografiske" segmenter, er på den annen side ikke noe kritisk element og kan i og for seg være relativt stort dersom dette bidrar til en god frekvensmodell og generelt til gode estimater på siste linje i tabellen.

Dette peker på de hovedoppgaver man står ovenfor når det gjelder segmentering.

1. Vi må bestemme oss for antall L- og D-segmenter (og selvsagt hva som karakteriserer hvert segment)
2. Vi må beregne størrelsen på  $N_d$ . Det enkleste er selvsagt om D-segmentene bare dannes ved en gruppering av befolkningen etter kjønn og alder, men i de fleste tilfeller blir det også aktuelt å kryssgruppere kjønn og alder mot andre variable og vi trenger et opplegg for å foreta en slik kryssgruppering..
3. Vi må bestemme andelene  $a_{dk} \forall d,k$ . Her kan det i større eller mindre grad bli tale om å benytte delmodeller, men også andeler direkte utregnet fra tilgjengelig statistikk kan være aktuelt.
4. Vi må bestemme  $f_{dk} \forall d,k$ . Dette er oppgaven for frekvensmodeller.

Hvis 2-4 er beheftet med relativt stor usikkerhet vil dette også gi stor usikkerhet i beregnet antall reiser for hvert L-segment på siste linje i Tabell 4.6. I så fall kan

vi regne med at vi bare oppnår en fiktiv presisjon ved å operere med mange L-segmenter. Med mange L-segmenter vil vi også få noen segmenter som utgjør en meget liten andel i form av antall reiser. Har vi f eks så mange som 10 L-segmenter vil vi helt sikkert kunne regne med at ett eller flere segmenter utgjør mindre enn 5 % av totalt antall reiser og får relativt liten betydning for aggregerte resultater.

#### 4.5.2 Segmentering med kontinuerlige variable

Inntekt er en variabel som er problematisk i forbindelse med segmentering. Inntekt varierer kontinuerlig i befolkningen både innen og mellom segmenter som dannes ved kombinasjoner av andre kjennetegn.

Vi har her flere muligheter:

- Vi kan helt unnlate å segmentere og operere med gjennomsnittlig inntekt pr person for hele populasjonen som modellen omfatter - enten dette gjelder personlig inntekt eller husholdningsinntekt.
- Vi kan segmentere etter andre kriterier og operere med gjennomsnittsinntekt pr segment, eventuelt la denne gjennomsnittsinntekt pr segment variere bare langs den geografiske dimensjon ved at den multipliseres med en sonespesifikk faktor.
- Vi kan segmentere f eks i høy, middels og lav inntekt og operere med en gjennomsnittsinntekt for de respektive kategorier. Dvs. at hvis vi ut fra andre kriterier har "S" segmenter så vil dette øke til 3S segmenter. Innen hvert av de S segmenter må vi da anta at fordelingen mellom de 3 inntekts-



segmenter holder seg konstant og at vi bare øker gjennomsnittinntekten i hvert segment. Ellers må vi også ha prognoser for fordelingen mellom inntektssegmenter for hvert segment som vi ellers har. I nåværende modell segmenteres etter inntektskategorier, men det gjøres ikke noe for å endre fordelingen mellom kategorier.

- Vi kan operere med en gjennomsnittsinntekt pr segment og korrigere for inntektsvariasjoner.

Implikasjonene av pkt 2 og 4 kan illustreres på følgende måte:

Anta at sannsynligheten for at person nr "i" i et gitt segment har en sannsynlighet for å velge et bestemt alternativ og at denne sannsynlighet kan skrives:

$$P_i = P(z, r_i) \quad (1)$$

$z$  = verdien på andre segmenteringsvariable som definerer segmentet og  $r_i$  = personens inntekt.

Antall personer i segmentet som velger dette alternativ vil "korrekt" beregnet da være:

$$N = \sum_i P_i \quad (2)$$

Summeringen går her over antall personer i segmentet. Sett at vi foretar en rekkeutvikling med 3 ledd av  $P_i$  rundt gjennomsnittsinntekten for segmentet,  $r^*$ . Da får vi:

$$P_i = P(z, r^*) + P'_r(z, r^*) \cdot (r_i - r^*) + \frac{1}{2} P''_r(z, r^*) \cdot (r_i - r^*)^2 \quad (3)$$

Hvis vi nå summerer sannsynlighetene over personer i segmentet får vi:

$$N^* = S \cdot P(z, r^*) + \frac{S}{2} P''_r(z, r^*) \cdot \sigma_r^2 \quad (4)$$

Der  $S$  = antall personer i segmentet. Hvis  $P''_r$  og/eller  $\sigma_r^2$  er "små" i tallverdi vil det andre leddet kunne sløyfes og vi får en løsning som antydte i pkt 2. Hvis begge ledd er relativt høye i tallverdi bør man strengt tatt ta med 2-gradsleddet som en korreksjonsfaktor. Dette vil tilsvare løsningen i pkt 4.

Det analytiske uttrykk for  $P_r$  blir rimelig enkelt hvis inntekten bare inngår i en av nyttefunksjonene, men blir mer komplisert dersom inntekten inngår i alle

nyttefunksjonene. I sistnevnte tilfellet må minst to nyttefunksjoner ha forskjellig verdi på parameteren for inntekt.

Gitt at sannsynligheten for å velge et bestemt alternativ vil ha en S-form som funksjon av  $r$  må man regne med at tallverdien på den 2. deriverte er relativt lav over det meste av variasjonsområdet for  $r$ , og at tallverdien når et maksimum for to verdier av  $r$ .

Det kan være av interesse å finne et uttrykk for den maksimale feil som kan oppstå dersom man sløyfer 2. gradsleddet. Dersom inntekt inngår i flere nyttefunksjoner og parameteren har samme fortegn vil som regel effekten bli mindre – men ikke nødvendigvis - enn om inntekt inngår i bare én nyttefunksjon. For tilfellet hvor en variabel ( $r$ ) inngår bare i én nyttefunksjon og der lineært med parameter  $\beta$  vil vi ha:

$$\frac{\partial^2 P}{\partial r^2} = \beta^2 P(1 - P)(1 - 2P) \quad (5)$$

Her har den 2. deriverte et maksimum og et minimum for hhv  $P = 0,5 \pm 0,289$ . Av dette følger også maksimalfeilen:

$$\pm \frac{1}{2} \beta^2 0.211 \cdot 0.789 \cdot 0.578 \cdot \sigma_r^2 = \pm 0.048 \cdot \beta^2 \cdot \sigma_r^2 \quad (6)$$

Det kritiske er altså produktet av parameteren kvadrert og variansen innefor det segment vi ser på. Dette produktet kan vi i praksis regne ut og således få et uttrykk for den maksimale prosentvise feil vi kan få ved utregning av aggregert markedsandel når 2.-gradsleddet sløyfes.

Hvis  $r$  i stedet for å være en kontinuerlig variabel er en binær variabel vil  $\sigma^2$  nødvendigvis være mindre enn 1 og kan maksimalt være 0,5. Selv om  $\beta^2$  er relativt stor i tallverdi skal det da relativt mye til før man får mer enn  $\pm 5-10\%$  feil i beregning av aggregert markedsandel for et segment. Dette er i og for seg en relativt stor feil, men i praksis kan det selvsagt også være slik at når vi summerer over segmenter så kan feilen gå i ulike retninger for ulike segmenter slik at totalresultatet allikevel blir forholdsvis nøyaktig.

Kombinasjonen av parameterverdi over 1 i tallverdi og stor varians for den variable vi betrakter kan imidlertid lett gi store aggregeringsfeil.

En kontinuerlig variabel kan også inngå i 2 eller flere nyttefunksjoner med forskjellig parameterverdi. Vi kan definere gjennomsnittlig parameterverdi som:



$$\bar{\beta} = \sum_{m=1}^M P_m \beta_m \quad (7)$$

og variansen på parameterverdiene som:

$$\sigma_{\beta}^2 = \sum_{m=1}^M P_m (\beta_m - \bar{\beta})^2 \quad (8)$$

Her refererer  $P_i$  seg til de sannsynligheter vi får når vi benytter gjennomsnittsverdier for den kontinuerlige variable. Da kan vi utlede følgende uttrykk for den 2. deriverte (forutsatt at inntekten inngår lineært i nyttefunksjonene):

$$\frac{\partial^2 P_m}{\partial r^2} = P_m \{(\beta_m - \bar{\beta})^2 - \sigma_{\beta}^2\} \quad (9)$$

Det ikke lett å få en klar konklusjon om den maksimale feil fra dette uttrykket. Isolert sett vil tallverdien på uttrykket øke med  $P_m$ , men hvis  $P_m$  går mot 1 så vil også uttrykket i parentesene gå mot null. Hvis  $P_m$  går mot null vil også aggregeringsfeilen gå mot null. Det kan derfor hende at vi her har et entydig maksimum eller minimum avhengig av verdien på uttrykket i klammeparentesen.

På den annen side: Det er lett å regne ut verdien på uttrykke for enhver verdi på  $r$  inkludert gjennomsnittsverdien i samplet og på den måte få en ide om størrelsen på aggregeringsfeilen som da vil være tilnærmet gitt ved:

$$\frac{1}{2} \frac{\partial^2 P_m}{\partial r^2} \sigma_r^2 = \frac{1}{2} P_m \{(\beta_m - \bar{\beta})^2 - \sigma_{\beta}^2\} \sigma_r^2. \quad (10)$$

NB! vi har her to varianser. Denne ene definert ved variansen på en kontinuerlig variabel for personer i et markedssegment og den andre er definert for parameterverdier i nyttefunksjoner for ulike alternativer.

Hvis vi har 2 eller flere kontinuerlige variable og opererer med gjennomsnittsverdier for disse innenfor hvert segment som dannes av kategoriske variable blir uttrykkene for "feilen" vi gjør mer komplisert fordi vi får et uttrykk som inneholder flere varianser pluss kovarianser. Metoden i pkt 4 ovenfor blir da svært tungvint å benytte.

Inntekt i seg selv vil normalt ikke være en "policy-variabel" i transportmodeller og man kan normalt kalibrere modeller slik at man i en 0-situasjon treffer

markedsandelen for ulike alternativer relativt brukbart hvis man har et godt datagrunnlag å kalibrere mot. Virkningen av inntektsendringer vil normalt bare ha betydning i forbindelse med prognoser.

I Ben-Akiva og Lerman (1985), kapittel 6 fremheves "sample enumeration" som den "beste" metode. Dette innebærer at man anvender en estimert modell på et representativt utvalg av befolkningen og deretter "blåser" opp resultatene til totalpopulasjonen. Samtidig påpekes det at denne metode kan være problematisk i forbindelse med sonebaserte modeller på grunn av beregningsomfanget. Dette var også konklusjonen etter noen vurderinger og beregninger som ble gjort i forbindelse med implementeringen av NTM-4.

Den metode som alternativt anbefales er at man estimerer en modell. Deretter danner man et antall segmenter basert på kombinasjoner av diskrete variable. For hvert segment beregnes gjennomsnittsverdien på aktuelle kontinuerlige variable. Nyttfunksjonene beregnes for hvert individ i samplet med disse gjennomsnittsverdier i stedet for observerte verdier. De beregnede verdier på nyttefunksjonene benyttes så til å estimere en skalaparameter som nyttefunksjonene skal multipliseres med. Dette skal ta hensyn til at de "målefeil" i de kontinuerlige variable som introduseres ved å benytte gjennomsnittsverdier i stedet for de verdier som er observert for hvert individ, vil øke variansen på restleddet i nyttefunksjonene. Skalaparameteren skal ligge mellom 0 og 1. Dette er et opplegg som burde passe godt i langdistansemodellen.

Dette innebærer at vi benytter metoden i pkt 2, men med den ekstra vri at vi estimerer en skalaparameter for nyttefunksjonen. Spørsmålet er om man også bør estimere et nytt konstantledd, dvs. at vi benytter kalkulerte nyttefunksjoner som uavhengige variable og estimerer parametre  $A$  og  $B_i$  slik at  $U^*_i = AU_i + B_i$ . Man skulle nemlig tro at de "nye" feilledd i nyttefunksjonene også kunne få endret forventningsverdi.

## 4.6 Segmentering med henblikk på førerkortinnehav og biltilgang

### 4.6.1 Beregning av antall i hvert segment

La  $N_d$  stå for antall personer i demografisk segment nr "d" i en sone. La  $b_{dh}$  være andelen av disse som tilhører husholdningskategori "h". La  $P_{dhs}$  være andelen av befolkningssegment "dh" som mht førerkortinnehav og bilhold faller i segment "s".

Avhengig av hvordan vi ellers segmenterer i modellen vil vi være interessert i:

Totalt antall personer i sonen som tilhører segment "s":

$$N(s) = \sum_d N_d \sum_h b_{dh} P_{dhs}$$

Antall personer i befolkningssegment "d" som tilhører segment "s":

$$N(s | d) = N_d \sum_h b_{dh} P_{dhs}$$

Antall personer i befolkningssegment "d" og husholdingstype "h" som tilhører segment "s":

$$N(s | d, h) = N_d b_{dh} P_{dhs}$$

Her skal vi også ha:

$$\sum_h b_{dh} = 1 \quad \text{og} \quad \sum_s P_{dhs} = 1$$

De tre modeller for kombinasjonen av førerkortinnhav og biltilgang gir altså  $P_{dhs}$ , mens vi mp benytte en annen metode for å anslå "vektene"  $b_{dh}$ . Dette kommer jeg tilbake til. (Her vil  $P_{dhs}$  kunne motsvare et element  $a_{dk}$  i Tabell 2).

#### 4.6.2 Personer i husholdninger med én person over 18 (20) år

Modeller for kombinasjoner av førerkortinnhav og bilhold ble estimert separat for hushold med hhv 1, 2 og 3+ personer over 18 år. Disse modeller inneholder en blanding av personspekifikke variable (kjønn, alder, personlig inntekt), variable som avhenger av bosted og variable som karakteriserer husholdningen (antall barn, husholdningsinntekt, par m/barn osv). Skal modellene benyttes til å segmentere befolkningen etter førerkortinnhav og bilhold kreves det at man først segmenterer etter de variable som benyttes i modellen. Av de modeller som ble estimert for personer i hushold med én person 18+ år hadde modellen som er gjengitt i Tabell 4.7 den beste likelihoodverdi. Vi ser at alle parametere har t-verdi større enn 2 og alle er forholdsvis høye i tallverdi.

Tabell 4.7: Modell for personer over 18 (20) år i husholdninger med én person over 18 (20) år. Beregnet inntekt for "missing". Min. inntekt = 80000<sup>1)</sup>

```

=====
MAXLIK: Version 3.1.3 10/08/01 12:31 pm
=====
se_i1pka.pgm 18+ 20+ 3)
mean log-likelihood -0.703579 -0.701781
Number of cases 2114 2104
-----
Parameters Estimates Est./s.e. Estimates Est./s.e.
-----
U-fkort
P01 konst_fk 3.0603 5.789 3.0054 5.607
P02 mk18&19 -2.2381 -3.218
P03 k65+ -0.5232 -2.606 -0.5513 -2.726
P04 sqrt(alder) -0.4206 -6.506 -0.4173 -6.367
P05 kvinne -0.6920 -4.097 -0.6377 -3.757
P06 OBTS -0.7013 -5.090 -0.7165 -5.184
P07 pinnt 100' 1.3128 12.501 1.3266 12.579
U-bil
P08 and.tett -2.4639 -5.892 -2.3002 -4.590
P09 ant.barn0-12 0.6080 3.525 0.6217 3.572
P10 OBTS -0.8737 -5.477 -0.8824 -5.388
P11 pinnt 100' 0.4997 5.739 0.5131 5.847
P12 sqrt(alder) 0.4888 7.728 0.5852 5.968
P13 (alder-35)^22) -0.5153 -2.869 -0.6505 -3.051
P14 kvinne -0.6003 -4.164 -0.5805 -3.999
P15 konstant -0.7888 -1.171
=====

```

- 1) P01-P07 er parametere i nyttefunksjon for førerkort
- 2) Den variable er dividert med 1000
- 3) Estimert med alder = oppgitt alder + 0,5

TØI rapport 606/2002

I modellen er det 4 dummy-variable. Én av disse (OBTS) gjelder imidlertid for "storby" og vil slå inn for alle segmenter i soner som ligger i disse fire byer. Hvis vi segmenterer etter de øvrige dummy-variable får vi følgende segmenter:

1. Menn 18 eller 19
2. Kvinner 18 eller 19
3. Menn 20+ år
4. Kvinner 20-64 år
5. Kvinner 65+ år

Vi har i tillegg følgende variable som ikke er binære:

6. sqrt(alder)
7. personlig inntekt
8. (alder-35)<sup>2</sup>/1000
9. antall barn 0-12

Variabel nr 6 og 8 kan benyttes til ytterligere demografisk segmentering. Dette er særlig aktuelt innenfor aldersintervall over 20 år for menn og kvinner. Vi kan derfor tenke oss følgende segmenter:

- a1: Menn 18 eller 19
- a2: Kvinner 18 eller 19
- a3: Menn 20-24 år
- a4: Kvinner 20-24 år
- a5: Menn 25-34 år
- a6: Kvinner 25-34 år
- a7: Menn 35-64 år
- a8: Kvinner 35-64 år
- a9: Menn 65+ år
- a10: Kvinner 65+ år

Vi kunne også tenke oss ytterligere segmentering over 65 år, men dette betyr neppe så mye siden reise-frekvens (og antall personer) ikke er så stort. For hver av disse segmenter kan vi beregne gjennomsnittsverdier for de variable som varierer kontinuerlig. I de data som er benyttet for estimering av modellen har vi gjennomsnittverdier som vist i Tabell 4.8.

Tabell 4.8: Gjennomsnittsverdier for segmenter i samplet

Segment	Antall	Gj.sn alder	(ald-35) <sup>^</sup> 2/1000	Sqrt (alder)	p- inntekt	ant. barn<12
a1: m18-19	6	18.833	0.262	4.340	0.895	0.167
a2: k18-19	4	18.500	0.273	4.301	1.250	0.000
a3: m20-24	41	22.610	0.155	4.753	1.510	0.024
a4: k20-24	50	22.560	0.156	4.748	1.222	0.160
a5: m25-34	240	29.646	0.037	5.438	2.512	0.058
a6: k25-34	250	29.844	0.035	5.457	1.913	0.556
a7: m35-64	449	47.517	0.227	6.866	2.725	0.111
a8: k35-64	499	50.405	0.313	7.072	2.055	0.194
a9: m65+	168	74.262	1.584	8.609	1.590	0.000
a10: k65+	407	74.803	1.624	8.641	1.166	0.000
Alle	2114	50.249	0.577	6.963	1.989	0.147

TØI rapport 606/2002

Av Tabell 4.8 fremgår det at i husholdninger med én person over 18 år så er det bare 0,5 % hvor personen er 18 eller 19 år. Disse to segmenter er altså helt ubetydelige og bør for denne type husholdninger kunne sløyfes ved implementering. Det vil si at vi later som om alle personer over 18 år i denne type husholdninger er 20 år eller eldre. Modellen i Tabell 4.\*7 ble derfor re-estimert på personer over 18 år (kolonne 4 og 5). Det gir en også en marginal forbedring i likelihood-verdien pr observasjon. Tabell 4.9 viser hvordan den re-

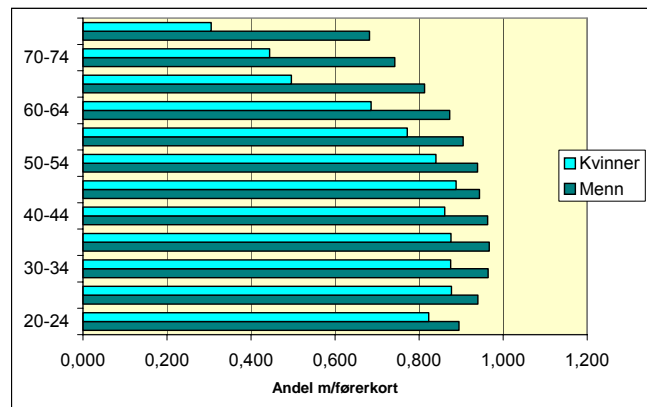
estimerte modell reproducerer fordelingen og de implisitte inntektselastisiteter.

Tabell 4.9: Effekten av inntektsøkning på fordelingen mellom segmenter for modellen i Tabell4.7. (20+)

Segment	Observert	Modell	%-avvik	Innt. elast.
s=1	467	467.0	0.00	-1.000
s=3	343	342.5	-0.15	-0.380
s=4	1294	1294.5	0.04	0.434
Antall	2104	2104.0		

TØI rapport 606/2002

Figur 4.1 viser førerkortandeler som modellen produserer i ulike aldersgrupper. De modellberegnete sannsynligheter er her veiet med WVAR. For alder isolert gir modellen her et maksimum i aldersgruppen 20-24 år og deretter synkende førerkortandel. Når dette ikke reproduseres av modellen skyldes det trolig at inntekt og andelen som bor i storbyer endres med alderen.



TØI rapport 606/2002

Figur 4.1: Modellproduserte førerkortandeler, veiet med WVAR

I den re-estimerte modell får vi 8 segmenter definert ved kjønn og alder. Spørsmålet er så om vi skal segmentere videre i den forstand at vi opererer med forskjellig gjennomsnittsinntekt i ulike segmenter avhengig av bosted, dvs. ”storby” og ”andel av befolkningen i tettbygde strøk”. Problemet er da at antall observasjoner i hvert segment begynner å bli lite når vi splitter opp ytterligere de 8 segmenter. Estimert på gjennomsnittsinntekt begynner også å bli relativt usikkert hvis vi skal basere oss på RVU-data. Et alternativ kan være at vi på grunnlag av gjennomsnittsinntekt i NTPL – soner lager en indeks som angir avviket fra landsgjennomsnittet. Når modellen anvendes på soner kan da gjennomsnittsinntekten for

segmentet multipliseres med denne indeks. Vi får da tatt hensyn til geografiske variasjoner i gjennomsnittsinntekt.

Som nevnt ovenfor vil den maksimale feil på aggregert markedsandel for den enkelte segmenter, jf. (5) og (6), avhenge av variansen på den variable og parameterverdien kvadrert. Av de aktuelle parametrene er det bare parameteren for inntekt i nyttefunksjonen for førerkort som har en tallverdi større enn 1.

Tabell 4.10 viser at det er visse systematiske forskjeller mellom storbyene og resten av landet. Ved en ytterligere oppdeling av disse to kategorier begynner gjennomsnittsverdien å bli forholdsvis usikre når vi tar hensyn til antall personer i de enkelte segmenter. Ytterligere segmentering med hensyn på disse to variable er derfor neppe tilrådelig. For resten av landet – og eventuelt for soner innenfor storbyene – bør man heller supplere med en indeks som inntekten kan multipliseres med for bedre å ivareta inntektsvariasjoner.

Tabell 4.11 gir variansen for de variable hvor vi operer med gjennomsnittsverdier i nyttefunksjonene. Bortsett fra når det gjelder inntekt er variansene relativt små. I kombinasjon med parameterverdier på under én i tallverdi skulle man derfor ikke tro at det særlig store feil ved aggregeringen bortsett fra det som kan skyldes inntekt.

Siden vi her opererer m.m. enn én kontinuerlig variabel burde vi ved vurdering av maksimalmeilen også tatt hensyn til kovariansen mellom de kontinuerlige variable innenfor hvert segment. Særlig menn i aldersgruppene fra 20 år har relativt stor varians på inntekten og dette indikerer at bruk av gjennomsnittsverdier for inntekt kan være en kilde til aggregeringsfeil spesielt for disse grupper. I første rekke vil dette gjelde for førerkortinnehav hvor parameteren for inntekt er relativt høy. På den annen side er førerkortinnehavet relativt høyt og vi vil jevnt over ligge et godt stykke fra de sannsynligheter som gir maksimal feil ved aggregering.

Tabell 4.10: Segmentert etter storbyer og resten av landet

Segment	Oslo, Bergen, Tr.heim, Stvngr			Resten av landet		
	Antall	Gj.sn.innt	Gj.sn ant barn	Antall	Gj.sn.innt	Gj.sn ant barn
a3: m20-24	24	1.302	0.000	17	1.804	0.059
a4: k20-24	20	1.159	0.200	30	1.264	0.133
a5: m25-34	106	2.505	0.000	134	2.518	0.104
a6: k25-34	114	2.044	0.325	136	1.803	0.750
a7: m35-64	122	2.948	0.057	327	2.641	0.131
a8: k35-64	186	2.242	0.156	313	1.943	0.217
a9: m65+	49	1.858	0.000	119	1.479	0.000
a10:k65+	120	1.353	0.000	287	1.088	0.000

TØI rapport 606/2002

Tabell 4.11: Varianser på variable innenfor de enkelte segmenter

Segment	Oslo, Bergen, Tr.heim, Stvngr				Resten av landet			
	Barn<12	Pinntekt	sqrt(ald)	(ald-35)^2	Barn<12	Pinntekt	sqrt(ald)	(ald-35)^2
a3	0.000	0.609	0.019	0.001	0.055	0.701	0.016	0.001
a4	0.160	0.356	0.021	0.001	0.116	0.399	0.017	0.001
a5	0.000	1.309	0.064	0.001	0.168	1.415	0.073	0.001
a6	0.377	0.729	0.075	0.001	0.835	0.570	0.062	0.001
a7	0.054	1.908	0.366	0.051	0.237	1.705	0.367	0.059
a8	0.196	0.794	0.403	0.064	0.285	0.815	0.374	0.067
a9	0.000	1.527	0.188	0.428	0.000	0.651	0.120	0.247
a10	0.000	0.691	0.119	0.251	0.000	0.300	0.133	0.276

TØI rapport 606/2002

Tabell 4.12: Tillegg til konstantleddet i nyttefunksjonene for personer i husholdninger med 1 person 20+ år

Segment	U-førerkort	U-bil
a3: m20-24	-2.005	2.711
a4: k20-24	-2.641	2.127
a5: m25-34	-2.288	3.185
a6: k25-34	-2.934	2.617
a7: m35-64	-2.880	3.892
a8: k35-64	-3.603	3.375
a9: m65+	-3.604	4.024
a10: k65+	-4.807	3.437
Tillegg OBTS	-0.717	-0.882

TØI rapport 606/2002

For hver av de 8 segmentene mhp. kjønn og alder kan vi nå regne ut et tillegg (positivt eller negativt) til nyttefunksjonen for hhv førerkort og bil og vi kan benytte gjennomsnittsinntekt og gjennomsnittlig antall barn under 12 år som vist i Tabell 4.10. Tilleggene er basert på parametere fra den re-estimering som er nevnt og er vist i kolonne 4 og 5 i Tabell 4.14. Tilleggene til konstantleddet for de ulike segmenter er vist i Tabell 4.12. Videre har vi benytte de gjennomsnittsverdier for inntekt og antall barn 0-12 som er vist i Tabell 4.10 (justert for alder=oppgitt alder + 0,5) til å beregne nyttefunksjoner for hver observasjon. Disse nyttefunksjoner er så benyttet til å estimere en skalaparameter og et konstantledd for hver av nyttefunksjonene. Konstantleddene ble imidlertid meget små i tallverdi og ikke signifikante. Tabell 4.14 viser derfor bare resultatet for estimering med skalaparametere for hver nyttefunksjon.

Som vi ser, får skalaparametrene høye t-verdier og ligger som teorien sier de skal gjøre mellom 0 og 1. Som ventet blir likelihood-verdien høyere i tallverdi (ca 10 %) når vi for alle kontinuerlige variable, bortsett fra "andel i tettbygde strøk" opererer med gjennomsnittsverdier for hvert segment. Tabell 4.14 viser fordeling på S1, S3 og S4 for de to modellene samt beregnede inntektselastisiteter.

Tabell 4.13: Estimering av skalparametere uten konstantledd

```

=====
Mean log-likelihood  -0.777336
Number of cases     2104
=====
Parameters      Estimates      Std. err.      Est./s.e.
-----
P01 skala Ufk   0.7879         0.0315         25.038
P02 skala Ubil  0.9755         0.0454         21.465
=====

```

TØI rapport 606/2002

Tabell 4.14: Prediksjon og inntektselastisiteter for opprinnelig modell og skalert modell med gjennomsnittsverdier for hvert segment

Segment	Antall obs.	Opprinnelig modell (Tab 4.7)		Skalert modell (Tab.4.13)	
		Predikert	Innt. elast.	Predikert	Innt. elast.
S1	467	467.0	-1.000	468.7	-1.156
S3	343	342.5	-0.380	344.7	-0.375
S4	1294	1294.5	0.434	1290.6	0.485
I alt	2104	2104.0		2104.	

TØI rapport 606/2002

Naturlig nok ser vi at de skalert modeller er dårligere når det gjelder reprodusere fordelingen på segmenter i samplet, men forskjellen er ikke avskrekkende. Vi ser også at skaleringen ikke er tilstrekkelig til å unngå at den skalert modell gir inntektselastisiteter som er høyere i tallverdi enn for den opprinnelige modell bortsett fra for S2. Uten skalering ville imidlertid forskjellene trolig blitt enda større.

Begge modeller er altså meget bra når det gjelder å reprodusere fordelingen på S1, S3 og S4 totalt. Tabell 4.15 viser fordelingen innenfor de enkelte demografiske segmenter. Som vi ser blir denne selvsagt dårligere enn for totalen, men bør allikevel være fullt akseptabel.

Et opplegg hvor vi varierer gjennomsnittsinntekten mellom soner både innenfor "storbyer" og i resten av landet vil sannsynligvis redusere modellens følsomhet noe slik at denne "aggregeringsfeil" i forhold til effekt av endringer reduseres. Med modellen med konstantledd gir bedre fordeling mellom segmenter, men fordi denne også har høyere verdi på skalaparametrene får vi også en tendens til at elastisitetene blir høyere i tallverdi og ligger lenger fra den opprinnelige modell. I denne modellen har vi nå altså 16 segmenter.

Tabell 4.15: Fordeling i sampel og modeller, prosent

	Pers 20+ år i hush.								
	med 1 voksen			Opprinnelig modell			Skalert modell		
	S1	S3	S4	S1	S3	S4	S1	S3	S4
a3: m20-24	7.3	46.3	46.3	10.5	40.5	49.0	12.6	40.0	47.4
a4: k20-24	14.0	42.0	44.0	18.2	42.3	39.5	21.3	41.3	37.5
a5: m25-34	5.8	20.4	73.8	4.8	22.0	73.2	5.2	21.7	73.1
a6: k25-34	11.6	30.8	57.6	12.3	29.8	57.9	14.3	29.1	56.6
a7: m35-64	8.2	9.6	82.2	6.1	9.0	85.0	5.8	8.7	85.5
a8: k35-64	19.0	15.2	65.7	18.3	16.6	65.2	18.8	16.8	64.4
a9: m64+	19.6	9.5	70.8	26.2	9.9	63.9	26.4	10.3	63.2
a10: k64+	61.2	10.3	28.5	61.2	9.2	29.6	59.0	10.6	30.4

TØI rapport 606/2002

#### 4.6.3 Personer i husholdninger med to personer over 18 (20) år

Modellen som er vist i Tabell 4.16 ser ut til å være den beste når det gjelder segmentering på førerkortinnnehav og bilhold for denne gruppe. For førerkort opererer vi her med kontinuerlig alder for IO. Dette innebærer at når det gjelder demografi (kjønn, alder) så kan vi operere med de samme segmenter som i foregående modell. Vi har også med storby (OBTS) som en dummyvariabel her.

Kombinasjonen av OBTS og demografisk segmentering gir da i utgangspunktet 20 segmenter. I tillegg

har vi med en dummy for par m/barn. Ren multiplikasjon gir da 40 segmenter. I nyttefunksjonene for bil kommer det inn ytterligere variable og man er nødt til å gjøre en vurdering av ytterligere segmentering kontra gjennomsnittsverdier. Tabell 4.18 viser at modellen reproducerer fordelingen på segmenter i samplet på en tilfredsstillende måte. De implisitte inntektselastisiteter medfører at segmentene med dårlig biltilgang (S1, S2 og S3) har negative inntektselastisiteter, mens S4 som medfører full tilgang til bil har en rimelig inntektselastisitet.

Tabell 4.16: Modell for segmentering i husholdninger med to personer over 18 år. Records med inntekt "missing" er med. Bil betinget av førerkort. Minimum inntekt = 100 000.1)

s2p_endx.pgm					
Mean log-likelihood -0.956986					
Number of cases 4817					
	Variabel	Estimat	St.avvik	"t-verdi"	Nyttefunksjon
P01	IO=k20-24	-0.4035	0.2582	-1.563	Førerkort
P02	m_ald/18	1.0050	0.2022	4.971	Førerkort
P03	k_ald/18	1.3343	0.1567	8.514	Førerkort
P04	(m_ald/18)^2	-0.1959	0.0491	-3.988	Førerkort
P05	(k_ald/18)^2	-0.3970	0.0402	-9.863	Førerkort
P06	OBTS	-0.7536	0.1286	-5.859	Førerkort
P07	pint 100'	1.0453	0.0779	13.421	Førerkort
P08	par m/barn	0.5991	0.1572	3.812	Førerkort
P09	OBTS	-0.7834	0.2553	-3.069	bil-ikke fk
P10	hhinnt 100'	0.7824	0.1209	6.474	bil-ikke fk
P11	par m/barn	0.8017	0.3314	2.419	bil-ikke fk
P12	k_ald/18	0.3492	0.0777	4.491	bil-ikke fk
P13	konstant	-2.2166	0.4212	-5.262	bil-ikke fk
P14	IO=k20-24	1.1965	0.4030	2.969	fk-ikke bil
P15	OBTS	2.4526	0.2049	11.970	fk-ikke bil
P16	ikke par	0.7991	0.2709	2.950	fk-ikke bil
P17	konstant	-2.1092	0.2843	-7.419	fk-ikke bil
P18	and.tett	-0.8238	0.1078	-7.641	biler $\geq$ førerkort
P19	IO=m35-65	0.4321	0.0772	5.598	biler $\geq$ førerkort
P20	IO=m65+	0.7806	0.1291	6.048	biler $\geq$ førerkort
P21	IO=k35-65	0.2783	0.0779	3.571	biler $\geq$ førerkort
P22	hhinnt 100'	0.5700	0.0699	8.151	biler $\geq$ førerkort
P23	IO=k20-24	0.6880	0.2476	2.779	biler<førerkort
P24	OBTS	0.3045	0.0840	3.623	biler<førerkort
P25	hhinnt 100'	0.4750	0.0691	6.871	biler<førerkort
P26	ikke par	-0.7147	0.1772	-4.033	biler<førerkort
P27	IO=m64+	0.4824	0.1995	2.418	biler<førerkort

1) Estimert med alder = oppgitt alder + 0.5

TØI rapport 606/2002

Tabell 4.17: Fordeling på segmenter og gjennomsnittsalder i hvert segment. Personer i husholdninger med 2 personer 18+ år

Segment	Antall observasjoner			Gjennomsnittsalder		
	Totalt	I "storbyer"	Resten	Totalt	I "storbyer"	Resten
b1: m18-19	24	3	21	18.92	19.17	18.88
b2: k18-19	21	5	16	19.17	19.50	19.06
b3: m20-24	84	34	50	23.19	23.41	23.04
b4: k20-24	122	35	87	22.93	23.13	22.84
b5: m25-34	564	159	405	30.57	30.25	30.70
b6: k25-34	710	179	531	30.44	30.28	30.50
b7: m35-64	1429	292	1137	47.76	47.58	47.81
b8: k35-64	1310	263	1047	47.46	47.18	47.53
b9: m64+	331	65	266	72.40	73.07	72.24
b10: k64+	267	54	213	72.24	71.46	72.44

TØI rapport 606/2002

Tabell 4.18: Fordeling på segmenter og inntektselastisiteter. Personer i husholdninger med 2 personer 18+ år

Segment	Ant. obs	Modell	%-avvik	Innt. elast.
S1	152	150.6	-0.9	-1.84
S2	288	290.4	0.8	-0.42
S3	140	137.7	-1.6	-1.35
S4	2119	2121.6	0.1	0.32
S5	2118	2116.6	-0.1	-0.07

TØI rapport 606/2002

Tabell 4.19: Fordelingen innen demografiske segmenter i samplet. Personer i husholdninger med 2 personer 18+ år. Prosent

	S1	S2	S3	S4	S5
B3	6.0	6.0	13.1	35.7	39.3
B4	8.2	10.7	11.5	19.7	50.0
B5	0.9	1.6	5.0	42.0	50.5
B6	2.3	4.5	3.5	37.6	52.1
B7	1.0	1.6	2.3	53.4	41.7
B8	2.3	8.1	1.1	45.0	43.4
B9	5.7	2.4	2.4	52.6	36.9
B10	19.9	34.5	2.2	12.7	30.7

TØI rapport 606/2002

Sammenholder vi Tabell 4.19 og 4.20 ser vi også at fordelingen mellom S1-S5 innen det enkelte demografiske segment også blir ganske bra.

Tabell 4.17 viser at også når det gjelder husholdninger med 2 personer over 18 år så utgjør husholdninger hvor en eller begge personer er under 20 år en meget liten andel. Totalt dreier det seg i utvalget om 45 personer eller 0,9 %. Det vil si at totalresultatene vil

bli fullstendig dominert av de øvrige segmenter. Modellen i Tabell 4.16 er derfor estimert på personer 20 år og over.

Tabell 4.20: Modellens fordeling innen demografiske segmenter. Personer i husholdninger med 2 personer 18+ år. Prosent

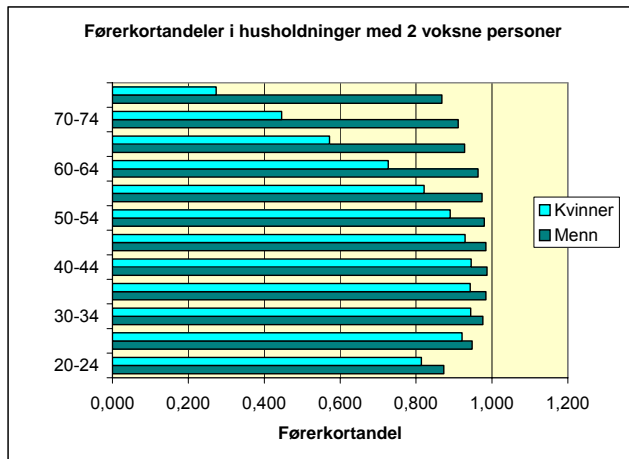
	S1	S2	S3	S4	S5
B3	8.4	4.9	10.0	35.0	41.8
B4	9.0	9.8	10.2	20.2	50.8
B5	1.6	1.8	3.8	40.0	52.7
B6	1.7	4.8	3.2	39.2	51.0
B7	0.8	1.2	1.8	53.7	42.5
B8	2.3	8.1	2.3	44.9	42.4
B9	6.1	3.2	3.2	52.3	35.3
B10	18.8	36.1	1.9	13.0	30.2

TØI rapport 606/2002

For inntekt (Tabell 4.21) og husholdningstyper (Tabell 4.22) er det noe større forskjell og ved beregning av gjennomsnittsverdier opprettholder vi her et skille.

Med 1. og 2. gradsledd i nyttefunksjonen for førerkort vil man kunne beregne alder for maksimalt førerkortinnhav korrigert for andre variable som eventuelt samvarierer med alder. For modellen i Tabell 10 er dette maksimum på 46,2 år for menn og 30,2 år for kvinner. Dette vil ikke nødvendigvis falle sammen med aldersintervallet for maksimalt førerkortinnhav, verken i observasjonsmaterialet eller i modellproduserte resultater. Grunnen er at andre variable som har større eller mindre samvariasjon med alder også spiller inn. Figur 4.2 viser imidlertid at maksimum også reproduseres relativt godt av modellen.





TØI rapport 606/2002

Figur 4.2: Modellproduserte førerkortandeler, veiet med WVAR

Modellen i Tabell 4.16 har 5 nyttefunksjoner. Når denne skal implementeres vil vi kunne forhånds-beregne konstantleddene for hvert segment og nyttefunksjon. Tabell 4.18 viser at det er relativt liten forskjell på gjennomsnittsalderen innenfor det enkelte segment i storbyene og resten av landet. Når det gjelder de variable "ald/18" og "(ald/18)<sup>2</sup>" er det derfor ikke grunn til å skille her.

Tabell 4.21: Gjennomsnittverdier innenfor segmenter. Personer i husholdninger med 2 personer 18+ år

Segment	Oslo,Bergen, Tr.heim,Stavanger		Resten	
	Pers.innt	Hush. inntekt	Pers.innt	Hush. inntekt
b3: m20-24	1.393	3.241	1.810	3.656
b4: k20-24	0.999	3.208	1.126	3.462
b5: m25-34	2.767	4.763	2.873	4.499
b6: k25-34	2.031	5.138	1.668	4.615
b7: m35-64	3.903	6.044	3.395	5.219
b8: k35-64	2.270	5.372	1.895	4.724
b9: m64+	2.281	3.412	1.957	2.973
b10: k64+	1.389	3.022	0.971	2.408

TØI rapport 606/2002

På samme måte som for husholdninger med én person over 20 år kan vi nå beregne et konstantledd for hvert demografiske segment og nyttefunksjon. Disse er vist i Tabell 14.23.

Disse konstantledd sammen med gjennomsnittsverdiene vist ovenfor og variable som varierer geografisk (OBTS og and.tett) er benyttet til å beregne nyttefunksjoner. Disse nyttefunksjoner er så benyttet ved estimering av skalaparametere og konstantledd. Det

vist seg at det bare konstantledd for "Ubil0|fk=1" ble signifikant. Resultatene fra denne estimering er vist i Tabell 4.24.

Tabell 4.22: Gjennomsnittverdier innenfor segmenter. Personer i husholdninger med 2 personer 18+ år

Segment:	Oslo,Bergen, Tr.heim,Stavanger		Resten	
	Par m/barn	Flere Voksne	Par m/barn	Flere Voksne
b3: m20-24	0.088	0.500	0.220	0.340
b4: k20-24	0.143	0.371	0.437	0.149
b5: m25-34	0.434	0.082	0.733	0.032
b6: k25-34	0.520	0.050	0.798	0.019
b7: m35-64	0.510	0.027	0.524	0.026
b8: k35-64	0.426	0.072	0.448	0.038
b9: m64+	0.015	0.015	0.011	0.034
b10: k64+	0.000	0.074	0.000	0.094

TØI rapport 606/2002

Tabell 4.23: Konstantledd i nyttefunksjonene. Personer i husholdninger med 2 personer 18+ år

Segment:	Nyttefunksjon				
	U-fk	Ubil fk=0	Ubil0 fk=1	Ubil1 fk=1	Ubil2 fk=1
b3: m20-24	0.969	-2.217	-2.109	0.000	0.000
b4: k20-24	0.650	-1.772	-0.913	0.000	0.688
b5: m25-34	1.137	-2.217	-2.109	0.000	0.000
b6: k25-34	1.111	-1.626	-2.109	0.000	0.000
b7: m35-64	1.241	-2.217	-2.109	0.432	0.000
b8: k35-64	0.672	-1.290	-2.109	0.278	0.000
b9: m64+	0.858	-2.217	-2.109	1.213	0.000
b10: k64+	-1.072	-0.815	-2.109	0.000	0.482

TØI rapport 606/2002

Tabell 4.24: Estimerer på skalaparametre og konstantledd for nyttefunksjoner. Person 20+ år i hushold med to personer over 18 år

Mean log-likelihood	-1.01333		
Number of cases	4817		
Parameters	Estimates	Std. err.	Est./s.e.
P01 skala_ufk	0.8357	0.0202	41.347
P02 skala_ubil fk=0	0.5038	0.0655	7.691
P03 skala_ubil0 fk=1	0.8533	0.0630	13.551
P04 skala_ubil1 fk=1	0.7854	0.0675	11.640
P05 skala_ubil2 fk=1	0.7925	0.0692	11.459
P06 konst_ubil0 fk=1	1.5086	0.2221	6.793

TØI rapport 606/2002



Også den skalerte modell med gjennomsnittsverdier gir en akseptabel fordeling på segmenter og avvikene mellom beregnede inntektselastisiteter i Tabell 4.18 i forhold til Tabell 8.19 er moderate.

Tabell 4.25: Fordeling på segmenter og inntektselastisiteter. Personer i husholdninger med 2 personer 18+ år.

Segment	Ant. obs	Modell	%-avvik	Innt. elast.
S1	152	151.6	-0.3	-1.97
S2	288	293.7	2.0	-0.78
S3	140	137.8	-1.6	-1.42
S4	2119	2126.3	0.3	0.30
S5	2118	2107.6	-0.5	0.01

TØI rapport 606/2002

Tabell 4.26: Modellens fordeling innen demografiske segmenter. Personer i husholdninger med 2 personer 18+ år. Prosent

	S1	S2	S3	S4	S5
B3	7.1	6.5	8.6	36.1	41.7
B4	8.9	11.5	18.5	19.9	41.2
B5	1.3	2.3	3.6	41.0	51.8
B6	2.0	5.5	2.8	39.7	49.9
B7	0.5	1.4	1.8	52.3	44.1
B8	2.3	7.5	2.1	44.6	43.4
B9	5.3	4.7	2.7	57.0	30.2
B10	21.7	32.8	2.1	14.6	28.8

TØI rapport 606/2002

Fordelingen innen de enkelte demografiske segmenter blir naturlig nok litt dårligere enn med den opprinnelige modell.

Alt i alt ser det ut som at bruk av gjennomsnittsverdier for det enkelte segment også i denne modellen fungerer tilfredsstillende. Også for denne modellen vil resultatene kunne bli bedre ved implementering dersom inntekten justeres med en indeks som er spesifikk for hver sone.

#### 4.6.4 Personer i husholdninger med 3+ personer over 18 (20) år

Resultatene for den endelige modell for personer i hushold med flere enn 3 personer over 18 år er vist i Tabell 4.27.

I denne modellen er ikke personer på 18 og 19 år en ubetydelig gruppe som i de to foregående, og vi bør derfor operere med 10 demografiske segmenter. Det er

på den annen side bare ca. 3 % av IO i denne type husholdninger som er over 65 år eller eldre.

Tabell 4.27: Modell estimert for personer over 18+ år i husholdninger med 3+ voksne personer. Records med inntekt "missing" er med. Bil betinget av førerkort. Minimum husholdningsinntekt = 100 000

Program:sp3_end.pgm					
Mean log-likelihood -0.917176					
Number of cases 1145					
Variabel	Estimat	St.avvik	"t-verdi"	Nyttefunksjon	
P01 andel i tettb.strøk	-1.2096	0.3814	-3.172	Førerkort	
P02 m_ald/18	2.6249	0.4078	6.437	Førerkort	
P03 k_ald/18	2.7890	0.4052	6.883	Førerkort	
P04 (m_ald/18)^2	-0.5572	0.1007	-5.533	Førerkort	
P05 (k_ald/18)^2	-0.7778	0.1132	-6.868	Førerkort	
P06 IO.innt i 100'	0.7171	0.1384	5.182	Førerkort	
P07 IO alder<20	-1.4580	0.2313	-6.305	Førerkort	
P08 hhinnt i 100'	0.3459	0.0706	4.897	bil-ikke fk	
P09 IO alder<20	0.9428	0.4707	2.003	bil-ikke fk	
P10 m20-35,k20-24	2.8488	0.5626	5.063	fk-ikke bil	
P11 k25-34	3.6669	0.6925	5.295	fk-ikke bil	
Oslo, Bergen,				fk-ikke bil	
P12 Tr.h, Stvng	1.4914	0.4022	3.708		
P13 konstant_Ubil10	-2.4907	0.7059	-3.529	fk-ikke bil	
P14 IO=mann	0.5273	0.1534	3.437	biler≥førerkort	
P15 andel i tettb.strøk	-0.7745	0.2962	-2.615	biler≥førerkort	
P16 hhinnt i 100'	0.7192	0.1355	5.308	biler≥førerkort	
P17 konstant_Ubil11	-1.0853	0.2707	-4.010	biler≥førerkort	
P18 hhinnt i 100'	0.6622	0.1330	4.978	biler<førerkort	

TØI rapport 606/2002

Tabell 4.28: Fordeling på segmenter og inntektselastisiteter. Personer i husholdninger med 3+ personer 18+ år

Segment	Ant. obs	Modell	%-avvik	Innt. elast.
S1	31	26.8	-13.5	-1.44
S2	139	143.2	3.0	-0.25
S3	39	41.7	6.9	-1.40
S4	242	241.0	-0.4	0.34
S5	694	692.3	-0.2	0.06

TØI rapport 606/2002

Selv om modellen har relativt store avvik når det gjelder segment S1 og S3, så er disse segmenter relativt sett så små at dette har liten betydning.

Sammenholder vi Tabell 4.29 og Tabell 4.30 ser vi at modellen fordeler brukbart innenfor de enkelte demografiske segmenter, men med noen få store avvik der hvor det er få observasjoner

For segmentene C9 og C10 er det så få observasjoner at det ikke er noen grunn til å skille mellom gjennomsnittsverdier for ”storbyene” og resten av landet. Ved beregning av gjennomsnittinntekt for hvert demografisk segment er det imidlertid benyttet en beregnet inntekt for hhv storbyer og resten for disse to segmenter. Dette for å behandle alle segmenter likt.

Tabell 4.29: Fordelingen innen demografiske segmenter i samplet. Personer i husholdninger med 3+ personer 18+ år Prosent

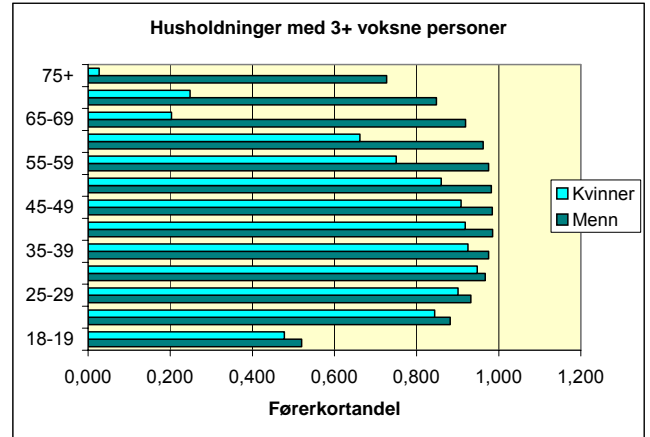
	S1	S2	S3	S4	S5	Ant. Obs
C1	4.3	42.9	0.0	14.3	38.6	70
C2	4.9	47.6	0.0	7.3	40.2	82
C3	4.6	6.5	10.2	25.0	53.7	108
C4	1.3	14.7	13.3	14.7	56.0	75
C5	0.0	5.4	8.9	35.7	50.0	56
C6	2.7	10.8	24.3	18.9	43.2	37
C7	0.9	1.3	0.9	28.3	68.6	318
C8	1.9	10.1	0.3	18.0	69.8	367
C9	12.5	0.0	0.0	20.8	66.7	24
C10	50.0	50.0	0.0	0.0	0.0	8

TØI rapport 606/2002

Tabell 4.30: Fordelingen innen demografiske segmenter i modellen. Personer i husholdninger med 3+ personer 18+ år . Prosent

	S1	S2	S3	S4	S5	Ant. Obs
C1	3.7	43.3	0.3	16.6	36.0	70
C2	4.4	47.1	0.4	9.8	38.3	82
C3	2.0	9.5	8.9	24.6	55.0	108
C4	3.2	12.7	13.0	14.9	56.2	75
C5	0.9	4.2	8.1	27.6	59.3	56
C6	1.9	6.5	24.7	14.2	52.7	37
C7	0.3	1.3	0.4	30.4	67.6	318
C8	2.1	9.6	0.4	18.5	69.4	367
C9	3.4	8.5	1.0	24.5	62.5	24
C10	26.5	59.1	0.2	2.8	11.5	8

TØI rapport 606/2002



TØI rapport 606/2002

Figur 4.3: Modellproduserte førerfordel, veiet med WVAR

Modellen gir for alder isolert sett maksimalt førerfordel ved 42,4 år for menn og 32,3 år for kvinner. Dette stemmer godt med Figur 4.3 hvor førerfordelen i ulike aldersgrupper i tillegg også påvirkes av andre variable som kan samvariere med alder.

Tabell 4.31: Gjennomsnittsinntekt innenfor de enkelte demografiske segmenter. Personer i husholdninger med 3+ personer 18+ år

Segment	Personlig inntekt i 100 000 kr			Husholdningsinntekt i 100 000 kr		
	Hele landet	Storbyer	Resten	Hele landet	Storbyer	Resten
C1:m18-19	0.584	0.250	0.634	4.777	5.422	4.682
C2:k18-19	0.279	0.278	0.280	4.120	4.265	4.096
C3:m20-24	0.961	1.038	0.937	5.058	4.615	5.198
C4:k20-24	0.530	0.431	0.576	4.256	3.353	4.681
C5:m25-34	1.917	1.524	2.074	5.406	4.658	5.705
C6:k25-34	1.447	1.316	1.527	4.642	3.822	5.140
C7:m35-64	3.111	3.633	3.004	5.166	5.927	5.010
C8:k35-64	1.727	2.334	1.629	4.838	5.851	4.675
C9:m65+	1.915	2.236*	1.185*	2.969	3.406*	2.879*
C10:k65+	0.739	0.999*	0.697*	2.246	2.716*	2.170*

\* Beregnet ved å benytte samme forholdstall mellom hele landet og storbyer-resten som for segmentene C7 og C8

TØI rapport 606/2002

Tabell 4.32: Gjennomsnittsverdier for alder og alder<sup>2</sup>i de enkelte demografiske segmenter. Personer i husholdninger med 3+ personer 18+ år

Segment	(alder+0,5)/18	((alder+0,5)/18) <sup>2</sup>
C1	1.052	1.107
C2	1.050	1.104
C3	1.234	1.529
C4	1.223	1.502
OC5	1.562	2.461
C6	1.579	2.517
C7	2.765	7.770
C8	2.685	7.305
C9	3.919	15.418
C10	4.208	17.903

TØI rapport 606/2002

Gjennomsnittsverdiene i Tabell 4.31 og 4.32 sammen med de parametre som er estimert i Tabell 21 er benyttet til å beregne konstantledd og nye verdier på nyttefunksjonene. Deretter er det estimert 3 skalaparametere og ett konstantledd som vist i Tabell 4.34.

Tabell 4.33: Konstantledd i nyttefunksjonene. Personer i husholdninger med 3+ personer 18+ år

Segment	Nyttefunksjon				
	U-fk	Ubil fk=0	Ubil0 fk=1	Ubil1 fk=1	Ubil2 fk=1
C1:m18-19	0.6858	0.9428	-2.4907	-0.5580	0.0000
C2:k18-19	0.6125	0.9428	-2.4907	-1.0853	0.0000
C3:m20-24	2.3874	0.0000	0.3581	-0.5580	0.0000
C4:k20-24	2.2439	0.0000	0.3581	-1.0853	0.0000
C5:m25-34	2.7279	0.0000	0.3581	-0.5580	0.0000
C6:k25-34	2.4454	0.0000	1.1762	-1.0853	0.0000
C7:m35-64	2.9292	0.0000	-2.4907	-0.5580	0.0000
C8:k35-64	1.8048	0.0000	-2.4907	-1.0853	0.0000
C9:m65+	1.6968	0.0000	-2.4907	-0.5580	0.0000
C10:k65+	-2.1879	0.0000	-2.4907	-1.0853	0.0000

TØI rapport 606/2002

Tabell 4.34: Estimerte skalaparametere og konstantledd. Personer i husholdninger med 3+ personer 18+ år

Program:sp3_end.pgm			
Mean log-likelihood -0.952782			
Number of cases 1145			
Parameter	Estimat	Std.avvik	"t-verdi"
P01 ufk	0.9236	0.0481	19.192
P02 bil fk=0, bil1 fk=1, bil2 fk=1	0.8326	0.0552	15.081
P03 ikke bil fk=1	0.9471	0.1166	8.124
P04 konstant-bil1 fk=1	-0.1653	0.0895	-1.846

TØI rapport 606/2002

Naturlig nok får vi også her en litt dårligere føyning målt ved likelihood-verdien, men modellen reproducerer fordelingen i samplet relativt bra (Tabell 4.35 og 4.36) og har inntektselastisiteter for de "tunge" segmenter som ligger svært nær den opprinnelige modell.

Tabell 4.35: Fordeling i samplet og modell. personer i husholdninger med 3+ personer 18+ år

Segment	Ant. obs	Modell	%-avvik	Innt. elast.
S1	31	28.9	-6.9	-1.67
S2	139	139.2	0.2	-0.32
S3	39	40.0	2.7	-1.92
S4	242	242.4	0.2	0.34
S5	694	694.5	0.1	0.11

TØI rapport 606/2002

Tabell 4.36: Modellens fordeling innenfor de enkelte demografiske segmenter. Personer i husholdninger med 3+ personer 18+ år

	S1	S2	S3	S4	S5
C1:m18-19	4.6	40.4	0.3	16.4	38.2
C2:k18-19	6.3	45.5	0.5	10.1	37.7
C3:m20-24	2.4	10.2	8.7	23.1	55.6
C4:k20-24	4.1	13.4	15.1	14.4	53.0
C5:m25-34	1.0	4.3	9.2	25.8	59.7
C6:k25-34	2.0	7.2	25.6	13.8	51.4
C7:m35-64	0.4	1.6	0.5	29.1	68.4
C8:k35-64	2.4	9.7	0.6	18.8	68.5
C9:m65+	3.8	8.7	1.3	23.8	62.4
C10:k65+	31.6	59.0	0.2	1.9	7.3

TØI rapport 606/2002

Ut fra de resultater som er vist ovenfor ser det altså ut til at et opplegg hvor vi segmenterer demografisk (kjønn og aldersgrupper) og benytter gjennomsnittsverdier innen de enkelte demografiske segment for inntekt og en del andre variable vil fungere tilfredsstillende. Ved re-estimering med gjennomsnittsverdier får vi skalaparametere som er mindre enn 1 og som med ett unntak er større enn 0,7.

De skalerte modeller reproducerer fordelingen i samplet relativt bra – men selvsagt litt dårligere enn modellene hvor vi ikke benytter gjennomsnittsverdier for hvert segment. De implisitte inntektselastisiteter – spesielt for de "tunge" segmenter – ligger meget nær de opprinnelige modeller. Dette tyder også på at bruk av gjennomsnittsverdier for hvert segment ikke introduserer aggregeringsskjevheter av betydning.

## 4.7 Modellene

Segmenteringen når det gjelder førerkort og bilhold går altså på 5 segmenter:

S1	Personen har ikke førerkort og ikke bil i husholdet
S2	Personen har ikke førerkort, men bil i husholdet
S3	Personen har førerkort, men ikke bil i husholdet
S4	Personen har førerkort og biltilgang $\geq 1$
S5	Personen har førerkort og biltilgang $< 1$ (og bil $> 0$ )

TØI rapport 606/2002

Av disse er S1 et segment hvor det er 0 sannsynlighet for å være bilfører (alternativet er ikke tilgjengelig) og hvor sannsynligheten for å være bilpassasjer er liten.

S2 har også 0 sannsynlighet for "bilfører" som alternativ for reisemåte, men vi må regne med større sannsynlighet for at en person i dette segment vil være bilpassasjer.

En person som tilhører S3 har ikke bilfører logisk utelukket som reisemåte, men vi må regne med en lav sannsynlighet. I likhet med S1 vil det også være liten sannsynlighet for å være bilpassasjer.

En person i S4 vil ha stor sannsynlighet for å ha bilfører som reisemåte og trolig relativt liten sannsynlighet for å være bilpassasjer.

En person i S5 vil normalt ha mindre sannsynlighet for å være bilfører enn en person i S4, men kanskje større sannsynlighet for å være bilpassasjer.

I langdistansemodellen skiller vi ikke mellom bilfører og bilpassasjer, men det er en dummy variabel for førerkort og en variabel for biltilgang som i praksis erstatter dummy variable for S3-S5. Modellen kan imidlertid lett re-estimeres hvor man da har en dummy-variabel for biltilgang=0 og én dummy-variabel for biltilgang  $\geq 1$ , eventuelt for  $0 < \text{biltilgang} < 1$ .

### 4.7.1 Husholdninger med én voksen person

Over har vi altså 3 modeller der skillet går på antall voksne i husholdningen. I husholdninger med én voksen person (18+ år) vil det bare være én person som har førerkort og hvis husholdningen har bil så vil denne person (IO) ha full tilgang til bil. Det vil si at det bare er segmentene S1, S3 og S4 som er aktuelle.

I denne modellen (Tabell 4.7) har vi 2 "nyttefunksjoner", én for førerkort, "UFK" og én for bil, "UBIL".

Vi har da følgende uttrykk for sannsynlighetene:

$$P_1 = \frac{1}{1 + e^{UFK}}$$

$$P_{Bil} = \frac{1}{1 + e^{-UBIL}}$$

$$P_3 = (1 - P_1) \cdot (1 - P_{Bil})$$

$$P_4 = (1 - P_1) \cdot P_{Bil}$$

Ved programmering av modellen må man altså beregne verdien på nyttefunksjonene for ulike demografiske segmenter og regne ut sannsynlighetene som ovenfor. Når vi benytter gjennomsnittsverdier for kontinuerlige variable innenfor hvert segment må nyttefunksjonene skaleres til slutt, dvs. i stedet for UFK og UBIL benytter vi:

$$UFK^* = a_1 \cdot UFK + b_1$$

$$UBIL^* = a_2 \cdot UBIL + b_2$$

Konstantleddene (b-ene) vil som regel være null, og i Tabell y er det bare skalaparametere.

Sannsynlighetene vil representere den andel av personer i demografisk segment "d" i husholdninger med én voksen person som skal ha verdiene 0 eller 1 på de respektive dummyvariable som benyttes i modeller for reisemiddelvalg.

### 4.7.2 Husholdninger med to voksne personer eller 3+ voksne personer

Her benyttes 5 nyttefunksjoner:

UFK	nyttefunksjon for førerkort
UBIL00	ikke førerkort – bil i husholdningen
UBIL10	førerkort – ikke bil i husholdningen
UBIL11	førerkort – biler i husholdningen $\geq$ antall førerkort i husholdningen
UBIL12	førerkort – biler i husholdningen $<$ antall førerkort i husholdningen

Sannsynlighetene for at en person skal tilhøre de respektive segmenter er gitt ved:

$$P_{FK} = \frac{1}{1 + e^{-UFK}}$$

$$P_{BIL0} = \frac{1}{1 + e^{-UBIL00}}$$

$$P_{BIL11} = \frac{e^{UBIL11}}{e^{UBIL10} + e^{UBIL11} + e^{UBIL12}}$$

$$P_{BIL12} = \frac{e^{UBIL12}}{e^{UBIL10} + e^{UBIL11} + e^{UBIL12}}$$

$$P_1 = (1 - P_{FK})(1 - P_{BIL0})$$

$$P_2 = (1 - P_{FK})P_{BIL0}$$

$$P_3 = P_{FK}(1 - P_{BIL11} - P_{BIL12})$$

$$P_4 = P_{FK}P_{BIL11}$$

$$P_5 = P_{FK}P_{BIL12}$$

$P_1$ - $P_5$  er altså sannsynlighetene for at en person skal tilhøre ett av de fem segmenter.

Også her blir fremgangsmåtene å beregne nyttefunksjonene for de respektive demografiske segmenter, skalering av disse med parameterne fra hhv Tabell 4.24 og Tabell 4.34 og deretter beregning av sannsynlighetene for at en person i et gitt demografisk segment skal tilhøre S1-S5.

#### 4.7.3 Kalibrering/justering av modellene

Modellene må trolig justeres/kalibreres i forbindelse med implementering. Dette går primært på to forhold:

- RVU-en er ikke nødvendigvis helt representativ når det gjelder førerkort/bilhold. Dette kan bl.a. skyldes at det alltid vil være en viss andel av personer i ulike aldersgrupper som faller utenom telefonintervju fordi de bor i ulike typer institusjoner. De personer som faller utenom vil ofte ha førerkortinnhav og biltilgang som avviker fra personer som tilhører private husholdninger. Det kan også tenkes at frafall og ubesvart i RVU generelt avviker på disse kjennetegn.
- Ved bruk av modellene i forbindelse med prognoser må vi ta hensyn til kohorteffekter som gjør at førerkortinnhav spesielt blant eldre kvinner i fremtiden kan bli vesentlig høyere enn det modellene predikerer.

For å illustrere de første punkt kan vi se på førerkortandeler fra registerdata kontra førerkortandeler i RVU.

Rekdal (2000, Vedlegg 1 og 2) gir følgende andeler for førerkort:

Tabell 4.37: Førerkortandel ifølge registerdata, prosent

	Menn		Kvinner	
	1997	1998	1997	1998
Alder				
15-24	57	56	51	50
25-34	92	91	87	86
35-49	95	94	89	87
50-64	94	93	75	75
64+	79	78	28	29
<b>Totalt</b>	<b>85</b>	<b>84</b>	<b>67</b>	<b>66</b>

TØI rapport 606/2002

Tilsvarende for RVU er vist i Tabell 4.40 hvor førerkortandel er gitt både uveiet og veiet med WESVAR.

Tabell 4.38: Førerkortandeler i RVU 1997/98

Aldersgruppe	Andel med førerkort, prosent			
	Menn		Kvinner	
	Uveiet	Veiet	Uveiet	Veiet
15-24	48.8	43.9	46.9	41.3
25-34	96.4	97.0	91.8	91.8
35-49	96.8	97.4	92.8	93.8
50-64	95.7	96.5	80.2	82.0
65+	87.4	89.0	41.4	40.2
<b>Totalt</b>	<b>85,0</b>	<b>86,8</b>	<b>76,0</b>	<b>76,4</b>

TØI rapport 606/2002

Sammenholder vi førerkortandeler i de to tabeller ser vi at registerdata gir lavere førerkortandeler enn RVU med unntak for den laveste aldersgruppe. Det siste kan bl.a. skyldes at undersøkelsen inneholder et tilleggsutvalg for Oslo hvor førerkortinnhav i de yngste aldersgrupper er lavere, men det kan også tenkes andre grunner til avviket. Når det gjelder de øvrige grupper kan forskjellen både skyldes skjevheter ved frafallet i RVU (frafallet er større blant personer uten førerkort enn blant personer med førerkort) og at RVU mangler personer i institusjoner som ikke har egen telefon eller eventuelt ikke er i stand til å svare. Disse er imidlertid med i totalbefolkningen når førerkortandeler beregnes på grunnlag av registerdata. Det er en markert forskjell på førerkortinnhavet for den eldste aldersgruppe når det gjelder menn og kvinner i Tabell 4.37. Dette skyldes nok i stor grad at kvinner gjennomgående

lever lenger og derfor – relativt sett har en større andel i de øvre aldersgrupper hvor førerkortandelen er lavest.

Dette er et fenomen som generelt vil medføre at modeller vi estimerer på RVU-data vil gi en liten tendens til å overpredikere reiseaktivitet hvis modellene anvendes ukorrigert på totalbefolkningen i ulike soner. Man bør strengt tatt korrigere for at en liten andel – som øker med alder – vil befinne seg i institusjoner eller felleshusholdninger og at disse vil ha en reiseaktivitet som avviker relativt mye fra den befolkning som er representert i RVU-ene som vi estimerer modeller på. På aggregert nivå kan dette kanskje dreie seg om 1-2 %.

Kohorteffekten vil medføre at førerkortandelen for kvinner nærmer seg en grense som trolig vil ligge litt under den for menn i aldersgruppen 35-49 år, men at

førerkortandelen for kvinner i høyere aldersgrupper om 10-20 år vil ligge vesentlig høyere enn de andeler vi finner i Tabell 4.37.

Tabell 4.39 viser en mer finfordelt aldersgruppering enn Tabell 4.38, og her har vi bare tatt med det aldersintervall hvor førerkortinnhav er aktuelt.

Justering for kohorteffekten kan gjøres på to måter. Enten kjører man modellene og justerer sannsynlighetene i etterkant med en eller annen metode eller man justerer modellenes parametere for hvert prognoseår og gjør ingen etterjustering i resultatene fra modellene. Det siste er trolig det mest ”ryddige”, men innebærer at vi må finne et opplegg for justering av parametere slik at vi får en rimelig kohorteffekt.

Tabell 4.39: Førerkortandel i ulike aldersgrupper i RVU 1997/98

Aldersgruppe	Andel i RVU av personer 18+ år		Andel med førerkort, prosent			
	Uveiet	Veiet	Menn		Kvinner	
			Uveiet	Veiet	Uveiet	Veiet
18-19	2.6	4.0	53.4	52.5	46.3	48.9
20-24	5.9	7.0	89.3	90.1	82.9	82.8
25-29	10.2	9.3	95.3	96.0	90.7	90.6
30-34	12.6	11.0	97.3	97.8	92.7	92.8
35-39	11.3	11.4	96.4	97.3	94.5	95.4
40-44	10.3	12.5	97.1	97.8	91.8	93.3
45-49	9.1	11.2	96.8	97.2	91.9	92.6
50-54	10.2	10.8	96.8	96.7	87.5	89.5
55-59	7.1	6.9	94.7	96.6	77.8	76.9
60-64	5.8	5.0	94.9	96.2	69.5	70.6
65-69	4.7	3.8	92.7	93.6	52.7	52.0
70-74	4.5	3.3	90.4	91.0	49.0	45.2
75-99	5.9	3.8	79.8	81.5	28.4	26.4

TØI rapport 606/2002

Tabell 4.40: Struktur på tabell for kalibrering og kohortfaktorer

Segment/År	2000	2005	2010	2015	2020	2025	2030
m18-19	y1	y1+0	y1+0	y1+0	y1+0	y1+0	y1+0
k18-19	y2	y2+0	y2+0	y2+0	y2+0	y2+0	y2+0
m20-24	y3	y3+0	y3+0	y3+0	y3+0	y3+0	y3+0
k20-24	y4	y4+0	y4+0	y4+0	y4+0	y4+0	y4+0
m25-34	y5	y5+0	y5+0	y5+0	y5+0	y5+0	y5+0
k25-34	y6	y6+0	y6+0	y6+0	y6+0	y6+0	y6+0
m35-64	y7	y7+0	y7+0	y7+0	y7+0	y7+0	y7+0
k35-64	y8	y8+x	y8+x	y8+x	y8+x	y8+x	y8+x
m65+	y9	y9+x	y9+x	y9+x	y9+x	y9+x	y9+x
k65+	y10	y10+x	y10+x	y10+x	y10+x	y10+x	y10+x

TØI rapport 606/2002

Tabell 4.41: Gjennomsnittlige sannsynligheter for førerkort og verdier på skalerte nyttefunksjoner for personer i husholdninger med 1 person 18+ år (uveiet)

Aldersgruppe	Menn			Kvinner		
	P(fk)-uskalert	P(fk)-skalert	U(fk)-skalert	P(fk)-uskalert	P(fk)-skalert	U(fk)-skalert
20-24	0.895	0.930	2.584	0.818	0.845	1.699
25-29	0.942	0.975	3.681	0.876	0.908	2.290
30-34	0.961	0.978	3.781	0.877	0.907	2.277
35-39	0.966	0.972	3.532	0.877	0.864	1.846
40-44	0.960	0.972	3.551	0.864	0.861	1.820
45-49	0.944	0.972	3.545	0.874	0.865	1.855
50-54	0.942	0.972	3.538	0.842	0.866	1.863
55-59	0.905	0.972	3.560	0.793	0.868	1.880
60-64	0.878	0.972	3.558	0.697	0.868	1.884
65-69	0.816	0.786	1.299	0.505	0.392	-0.440
70-74	0.754	0.787	1.304	0.431	0.388	-0.455
75+	0.685	0.786	1.301	0.311	0.381	-0.484

TØI rapport 606/2002

En justering for kohorteffekter hvor vi justerer direkte på nyttefunksjonene må ta utgangspunkt i at vi for hvert segment har beregnet en skalert nyttefunksjon (UFK\*) for førerkort:

Det vil si:  $UFK^* = a * UFK + b$

der UFK er nyttefunksjonen beregnet med gjennomsnittsverdier for de enkelte befolkningssegmenter. I de skalerte modeller er imidlertid  $b=0$  i nyttefunksjonen for førerkort.

Vi kan da tenke oss en tabell av følgende type som vist i Tabell 4.40:

y-ene for år 2000 vil være tall som adderes til de skalerte nyttefunksjoner for førerkort (UFK\*) for å kalibrere til kjent førerkortinnehav. x-ene vil være tall som adderes til de samme nyttefunksjoner for å ta hensyn til kohorteffekter.

Sammenlikner vi Tabell 4.38 og 4.39, kan det se ut som at det vil være nødvendig med en kalibrering. Vi bør imidlertid kjøre modellen ukorrigert på alle soner og summere resultatene før vi korrigerer. Forskjellen mellom de to tabeller kan jo også skyldes skjevheter ved utvalget i RVU og det har liten hensikt å korrigere før modellen er kjørt for hele landet og på en befolkning som vi først har justert litt ned for å da hensyn til at ikke alle tilhører private husholdninger. Dette er særlig viktig når det gjelder gruppen 65+.

Før den nødvendige kalibrering er gjort er det også vanskelig å fastlegge den eksakte størrelse på x-ene. *I forbindelse med programmeringen bør det imidlertid legges opp til en slik tabell for faktorer som kan*

*adderer til nyttefunksjonen. Denne tabell kan foreløpig inneholde bare nuller.*

Tabell 4.41 viser beregnede sannsynligheter og verdien på nyttefunksjonene for 12 aldersgrupper. Den skulle gi ganske god indikasjon på de korreksjoner man må gjøre i nyttefunksjonenes verdi for å få en viss justering av sannsynlighetene for førerkort.

Tilsvarende tabeller for de to andre husholdningstyper er vist i vedlegget.

## 4.8 Beregning av antall husholdninger i de tre kategorier

I NTM4 ble det operert med 31 husholdningskategorier og antall i hver kategori ble bestemt med "mininfo"<sup>5</sup> med bibetingelser knyttet til antall menn og kvinner i ulike aldersgrupper. Her hadde vi flere ukjente enn bibetingelser og denne metode var velegnet og konvergente relativt greit. Med bare tre husholdningstyper får vi flere bibetingelser enn ukjente slik at ligningssystemet er overbestemt og bare unntaksvis har en eksakt løsning.

<sup>5</sup> Ved "mininfo" regnet vi med at et ligningssystem  $C * HH = T$  skulle være tilnærmet eksakt oppfylt. Dette ville være tilfelle for "mange" vektorer HH. For å velge en av disse løste vi optimaliseringsproblemet:  $\min HH * \ln(HH/HH^*)$  st  $C * HH = T$  der  $HH^*$  er en "prior" basert på fordelingen i FoB1990, jfr Grue (1994).



Tabell 4.42: Foreløpige koeffisienter og befolkning basert på RVU

Segment	Hhold m/1 voksen	Hhold m/2 voksne	Hhold m/3+ voksne	Antall i 1000	Andel av personer 18+
m18-19	0.0023	0.0087	0.2236	62.88	0.019
k18-19	0.0018	0.0087	0.2251	67.90	0.021
m20-24	0.0178	0.0319	0.3398	113.75	0.035
k20-24	0.0228	0.0490	0.2573	111.41	0.034
m25-34	0.1075	0.2199	0.1740	307.59	0.095
k25-34	0.1134	0.2649	0.1140	350.74	0.108
m35-64	0.2096	0.6057	0.9002	945.80	0.292
k35-64	0.2556	0.5734	0.8995	929.57	0.287
m65+	0.0788	0.1311	0.1013	168.76	0.052
k65+	0.1904	0.1067	0.0840	183.32	0.057
				4241,72	1.000

TØI rapport 606/2002

Ved å benytte kvadratisk minimering på hhv relativt og absolutt avvik får vi følgende tall for husholdninger:

Tabell 4.43: Antall husholdninger i 1000, beregnet med kvadratisk minimering

HH-type	min rel. avvik	min. abs. avvik
1 voksen	235.85	203.59
2 voksne	1073.64	1138.46
3+ voksne	236.84	238.85
Hush. i alt	1546.33	1580.90
Målfunksjon	0.0243	758.2979

TØI rapport 606/2002

Tabell 4.44: Avvik på betingelser knyttet til antall personer i demografiske segmenter

Segment	Minimering av relativt avvik			Minimering av absolutt avvik		
	Rel. avvik	Abs. avvik	C*HH	Rel. avvik	Abs. avvik	C*HH
m18-19	-0.001	-0.04	62.84	0.014	0.90	63.78
k18-19	-0.071	-4.83	63.08	-0.057	-3.87	64.04
m20-24	0.046	5.17	118.92	0.065	7.35	121.10
k20-24	0.068	7.52	118.92	0.094	10.48	121.88
m25-34	-0.016	-4.93	302.66	0.020	6.21	313.79
k25-34	-0.036	-12.59	338.15	0.003	1.15	351.90
m35-64	-0.035	-32.86	912.94	0.002	1.45	947.25
k35-64	-0.044	-40.62	888.95	-0.011	-9.89	919.68
m65+	0.086	14.57	183.33	0.123	20.73	189.49
k65+	-0.022	-3.96	179.36	-0.016	-3.02	180.30

TØI rapport 606/2002



Med et overbestemt system er det kvadratisk minimering et velegnet alternativ. Problemet kan da formuleres på følgende måte:

La  $C$  være en  $10 \times 3$  matrise med koeffisienter. 10 refererer seg til de 10 demografiske segmenter og 3 til de 3 husholdningskategorier vi opererer med. Et element i matrisen  $C_{ij}$  angir gjennomsnittlig antall personer i demografisk segment "i" pr husholding av type "j".

La  $T$  være en vektor ( $10 \times 1$ ) med antall personer i hvert av de 10 demografiske segmenter. Ideelt ønsker vi å finne en vektor  $HH$  ( $3 \times 1$ ) som oppfyller lignings-systemet:

$$C * HH = T$$

Siden vi har flere ligninger enn ukjente, vil dette systemet generelt ikke ha noen løsning. Som alternativ kan vi da benytte en vektor som gir minst mulig avvik på bibetingelsene. Vi kan egentlig velge om vi vil minimere absolutt kvadratavvik eller relativt kvadratavvik.

Dette betyr å løse hhv:

$$\min_{HH} (C * HH - T)' * (C * HH - T)$$

og

$$\min_{HH} [(C * HH - T) / T]' * [(C * HH - T) / T]$$

For å illustrere metoden har jeg på grunnlag av RVU beregnet foreløpige verdier på  $C$  og  $T$ .

Vektoren  $T$  avviker relativt mye fra befolkningsstatistikk, særlig når det gjelder de yngste og eldste aldersgrupper. Til minimeringen benyttet prosedyren "Optimum" i Gauss.

Det er ikke noen umiddelbar viktig grunn som tilsier at man skal benytte det ene eller andre kriterium ved minimalisering, i hvert fall ikke så lenge man kan anta at alle bibetingelser er like presise. Hvis man har en formening om at noen bibetingelser er mer presise enn andre, enten fordi koeffisientene er mer presist bestemt eller antall i segmentene er mer presist bestemt, så kan man i realiteten minimalisere relativt avvik, men med lavere vekt på de bibetingelser som antas å være mest "upresise". Det er enkelt å modifisere GAUSS-programmet slik at man får en vekting.

Tabell 4.45: Relativ fordeling innen demografiske segmenter med minimering av relativt avvik.

Segment	Hhold	Hhold	Hhold	I alt
	m/1 voksen	m/2 voksne	m/3+ voksne	
m18-19	0.0086	0.1486	0.8427	0.9999
k18-19	0.0067	0.1481	0.8452	1.0000
m20-24	0.0353	0.2880	0.6767	1.0000
k20-24	0.0452	0.4424	0.5124	1.0000
m25-34	0.0838	0.7801	0.1362	1.0001
k25-34	0.0791	0.8411	0.0798	1.0000
m35-64	0.0541	0.7123	0.2335	0.9999
k35-64	0.0678	0.6925	0.2396	0.9999
m65+	0.1014	0.7678	0.1309	1.0001
k65+	0.2504	0.6387	0.1109	1.0000

TØI rapport 606/2002

Tabell 4.46: Relativ fordeling innen husholdningstyper med minimering av relativt avvik

Segment	Hhold	Hhold	Hhold
	m/1 voksen	m/2 voksne	m/3+ voksne
m18-19	0.0023	0.0043	0.0674
k18-19	0.0018	0.0043	0.0678
m20-24	0.0178	0.0159	0.1024
k20-24	0.0228	0.0245	0.0775
m25-34	0.1075	0.1100	0.0524
k25-34	0.1134	0.1325	0.0343
m35-64	0.2096	0.3029	0.2712
k35-64	0.2556	0.2867	0.2710
m65+	0.0788	0.0655	0.0305
k65+	0.1904	0.0534	0.0253
I alt	1.0000	1.0000	0.9998

TØI rapport 606/2002

Tabell 4.45 viser at ca 15 % av aldersgruppen 18-19 år befinner seg i hushold med 1 eller 2 voksne personer. På den annen side viser Tabell 4.43 at disse utgjør en helt ubetydelig andel av antall personer i disse husholdningstyper. Dette er også grunnen til at denne aldersgruppe er sløffet ved estimering av modeller for disse husholdningstyper.

Ved implementeringen bør vi benytte en koeffisientmatrise som gjør at vi automatisk får tatt hensyn til dette.

Tabell 4.47: Relativ fordeling innen demografiske segmenter med minimering av absolutt avvik

54307	Hhold m/1 voksen	Hhold m/2 voksne	Hhold m/3+ voksne	I alt
51781	0.0073	0.1553	0.8374	1.0000
144365	0.0057	0.1547	0.8396	1.0000
140651	0.0299	0.2999	0.6702	1.0000
349790	0.0381	0.4577	0.5042	1.0000
335573	0.0697	0.7978	0.1324	0.9999
830967	0.0656	0.8570	0.0774	1.0000
808402	0.0450	0.7280	0.2270	1.0000
285804	0.0566	0.7098	0.2336	1.0000
402149	0.0847	0.7876	0.1277	1.0000
3403789	0.2150	0.6737	0.1113	1.0000

TØI rapport 606/2002

Tabell 4.48: Relativ fordeling innen husholdningstyper med minimering av absolutt avvik

Segment	Hhold m/1 voksen	Hhold m/2 voksne	Hhold m/3+ voksne
m18-19	0.0023	0.0043	0.0674
k18-19	0.0018	0.0043	0.0678
m20-24	0.0178	0.0159	0.1024
k20-24	0.0228	0.0245	0.0775
m25-34	0.1075	0.1100	0.0524
k25-34	0.1134	0.1325	0.0343
m35-64	0.2096	0.3029	0.2712
k35-64	0.2556	0.2867	0.2710
m65+	0.0788	0.0655	0.0305
k65+	0.1904	0.0534	0.0253
I alt	1.0000	1.0000	0.9998

TØI rapport 606/2002

Den befolkning som er regnet ut på grunnlag av RVU data veiet med WVAR i Tabell 4.42 stemmer relativt dårlig med totalbefolkningen! Fra Rekdal (2000) har vi fra befolkningstatistikkene i vedlegget de tall som er oppgitt i Tabell 4.49.

Som vi ser, er det her store avvik både absolutt og relativt. De store avvik må i hovedsak skyldes frafall og skjevheter med andre årsaker i RVU. I den eldste aldersgruppen er det f eks god grunn til å forvente at personer med permanent opphold i institusjon utgjør en vesentlig høyere andel av befolkningen enn for de øvrige aldersgrupper.

SSBs husholdningsstatistikk som er basert på diverse utvalgsundersøkelser opererer med 2,05 mill. private husholdninger i 1998. Dette er opplagt for høyt tall siden befolkningen i disse husholdninger sum-

merer seg opp til den totale norske befolkning. Man har altså ikke tatt høyde for personer med permanent opphold i institusjon. På den annen side skal det mye til før man kommer så lavt i antall private hushold som Tabell 4.43, og samtidig tyder SSBs statistikk på at vi har en vesentlig høyere andel husholdninger med én voksen person enn det Tabell 4.43 indikerer.

Tabell 4.49: Totalbefolkning i ulike demografiske segmenter fra befolkningsstatistikken

Segment	Befolkning 1998	Relativ fordeling bef. 18+ år
m18-19	54307	0.016
k18-19	51781	0.015
m20-24	144365	0.042
k20-24	140651	0.041
m25-34	349790	0.103
k25-34	335573	0.099
m35-64	830967	0.244
k35-64	808402	0.238
m65+	285804	0.084
k65+	402149	0.118
Personer 18+ år i alt	3403789	1.000

TØI rapport 606/2002

Tabell 4.50 er hentet fra SSBs Husholdningsstatistikk 1998. Personer refererer seg her til alle aldersgrupper. Gitt at man ikke har med personer i institusjoner, skulle man tro at særlig antall "Aleneboende" er overvurdert i denne tabell. Husholdninger m/ én person over 18 år vil omfatte "Aleneboende" og en relativt stor andel av "Mor/far m/barn 0-17 år". Riktig antall for antall personer over 18 år i husholdninger med én person 18+ år ligger trolig i intervallet 850 000 – 900 000. Antall hushold med 2 voksne personer skulle da kunne være av omtrent samme størrelsesorden, mens husholdninger med 3+ voksne kan dreie seg om 150 000 – 200 000.

Tabell 4.50: Antall husholdninger og personer

Husholdningstype	Husholdninger	Personer
Aleneboende	829 000	829 000
Par <sup>2</sup> uten barn	469 000	939 000
Par m/barn 0-6 år	264 000	1 078 000
Par m/barn 7-17 år	192 000	773 000
Mor/far m/barn 0-17 år	105 000	276 000
Hush. m/voksne barn	147 000	432 000
Flerfamiliehusholdninger	41 000	119 000
I alt	2 047 000	4 446 000

Kilde: SSB- Husholdningsstatistikk 1998.

SSB opplyser at de data som husholdningsstatistikken er basert på ikke egner seg til å beregne koeffisienter av den type vi har benyttet i Tabell 4.42. Slike koeffisientmatriser bør trolig differensieres geografisk. Metoden med kvadratisk minimering bør derfor neppe anvendes før vi har bedre data fra FoB2001. Ved implementeringen bør vi i første omgang trolig benytte en mer robust metode. En metode som trolig er mer robust er å ta utgangspunkt i hvordan antall personer (veiet med WVAR) i hvert demografisk segment i RVU prosentvis fordeler seg på de tre husholdningstypene. Dette er trolig en mer stabil fordeling i rom og tid enn den fordeling som vil ligge implisitt i en (upresis) koeffisientmatrise.

## 4.9 Referanser

Ben-Akiva, M.E. og Lermann, R. (1985), *Discrete choice analysis: Theory and application to travel demand*. The MIT Press, The Massachusetts Institute of Technology.

Grue, Berit (1994): *Prototypiske husholdninger*. TØI arbeidsdokument TRU/437/94.

Rekdal, Jens (2000): *Modeller for innehav av førerkort*. TØI arbeidsdokument PT/400/00.

## Vedlegg til kapittel 4

Tabell V4.1: Hushold med 1 voksen person. Fordeling på kjønn og alder i RVU

Aldersgruppe	Antall obs		Uveiet		Veiet med WVAR	
	Menn	Kvinner	Menn	Kvinner	Menn	Kvinner
20-24	41	50	0.046	0.046	0.048	0.043
25-29	115	112	0.128	0.128	0.132	0.087
30-34	125	138	0.139	0.139	0.124	0.112
35-39	99	79	0.110	0.110	0.122	0.081
40-44	87	70	0.097	0.097	0.100	0.069
45-49	77	64	0.086	0.086	0.087	0.056
50-54	85	96	0.095	0.095	0.098	0.084
55-59	49	93	0.055	0.055	0.054	0.078
60-64	52	97	0.058	0.058	0.058	0.077
65-69	43	89	0.048	0.048	0.043	0.074
70-74	47	117	0.052	0.052	0.050	0.085
75+	78	201	0.087	0.087	0.083	0.154
I alt	898	1206	1.001	1.001	0.999	1.000

TØI rapport 606/2002

Tabell V4.2: Hushold med 1 voksen person. Modellberegnete førerkortandeler

	Uveiet		Veiet med WVAR	
	Menn	Kvinner	Menn	Kvinner
20-24	0.895	0.818	0.895	0.823
25-29	0.942	0.876	0.940	0.877
30-34	0.961	0.877	0.964	0.875
35-39	0.966	0.877	0.967	0.876
40-44	0.960	0.864	0.963	0.861
45-49	0.944	0.874	0.944	0.888
50-54	0.942	0.842	0.939	0.840
55-59	0.905	0.793	0.905	0.772
60-64	0.878	0.697	0.873	0.686
65-69	0.816	0.505	0.813	0.496
70-74	0.754	0.431	0.742	0.444
75+	0.685	0.311	0.682	0.305

TØI rapport 606/2002

Tabell V4.3: Hushold med 2 voksne personer. Fordeling på kjønn og alder i RVU

Aldersgruppe	Antall obs		Uveiet		Veiet med WVAR	
	Menn	Kvinner	Menn	Kvinner	Menn	Kvinner
20-24	84	122	0.035	0.051	0.032	0.049
25-29	226	312	0.094	0.130	0.089	0.124
30-34	338	398	0.140	0.165	0.136	0.155
35-39	352	349	0.146	0.145	0.143	0.156
40-44	300	240	0.125	0.100	0.138	0.120
45-49	188	186	0.078	0.077	0.092	0.082
50-54	239	235	0.099	0.098	0.103	0.095
55-59	178	188	0.074	0.078	0.077	0.076
60-64	172	112	0.071	0.046	0.068	0.041
65-69	121	106	0.050	0.044	0.046	0.040
70-74	114	83	0.047	0.034	0.042	0.032
75+	96	78	0.040	0.032	0.034	0.030
I alt	2408	2409	0.999	1.000	1.000	1.000

TØI rapport 606/2002

Tabell V4.4: Hushold med 2 voksne personer. Modellberegnete førerkortandeler

Aldersgruppe	Uveiet		Veiet med WVAR	
	Menn	Kvinner	Menn	Kvinner
20-24	0.867	0.812	0.873	0.814
25-29	0.949	0.921	0.948	0.921
30-34	0.977	0.945	0.976	0.944
35-39	0.984	0.946	0.984	0.943
40-44	0.987	0.943	0.987	0.945
45-49	0.984	0.929	0.984	0.929
50-54	0.979	0.891	0.980	0.890
55-59	0.976	0.820	0.974	0.821
60-64	0.965	0.725	0.963	0.727
65-69	0.931	0.579	0.928	0.572
70-74	0.916	0.448	0.911	0.446
75+	0.867	0.280	0.868	0.273

TØI rapport 606/2002

Tabell V4.5: Hushold med 3+ voksne personer. Fordeling på kjønn og alder i RVU

Aldersgruppe	Antall obs		Uveiet		Veiet med WVAR	
	Menn	Kvinner	Menn	Kvinner	Menn	Kvinner
18-19	70	82	0.122	0.144	0.126	0.148
20-24	108	75	0.188	0.132	0.182	0.126
25-29	40	26	0.069	0.046	0.069	0.048
30-34	16	11	0.028	0.019	0.026	0.022
35-39	17	19	0.030	0.033	0.031	0.032
40-44	53	87	0.092	0.153	0.105	0.173
45-49	104	120	0.181	0.211	0.195	0.212
50-54	80	98	0.139	0.172	0.129	0.162
55-59	35	32	0.061	0.056	0.054	0.053
60-64	29	11	0.050	0.019	0.048	0.017
65-69	13	3	0.023	0.005	0.021	0.001
70-74	6	2	0.010	0.004	0.007	0.003
75+	5	3	0.009	0.005	0.007	0.004
I alt	576	569	1.002	0.999	1.000	1.001

TØI rapport 606/2002

Tabell V4.6: Hushold med 3+ voksne personer. Modellberegnete førerkortandeler

Aldersgruppe	Uveiet		Veiet med WVAR	
	Menn	Kvinner	Menn	Kvinner
18-19	0.523	0.480	0.520	0.478
20-24	0.880	0.837	0.882	0.844
25-29	0.936	0.897	0.932	0.901
30-34	0.970	0.943	0.967	0.948
35-39	0.977	0.930	0.975	0.925
40-44	0.984	0.915	0.985	0.918
45-49	0.984	0.901	0.984	0.908
50-54	0.983	0.858	0.982	0.860
55-59	0.976	0.745	0.975	0.750
60-64	0.963	0.676	0.962	0.662
65-69	0.924	0.203	0.919	0.203
70-74	0.845	0.223	0.848	0.248
75+	0.759	0.020	0.727	0.027

TØI rapport 606/2002

Tabell V4.7: Personer i husholdninger med 1 person over 18 år. Gjennomsnittlige verdier på sannsynligheter for førerkort og nyttefunksjoner for førerkort

Aldersgruppe	Uveiet					
	Menn			Kvinner		
	P(fk)-uskalert	P(fk)-skalert	U(fk)-skalert	P(fk)-uskalert	P(fk)-skalert	U(fk)-skalert
20-24	0.895	0.930	2.584	0.818	0.845	1.699
25-29	0.942	0.975	3.681	0.876	0.908	2.290
30-34	0.961	0.978	3.781	0.877	0.907	2.277
35-39	0.966	0.972	3.532	0.877	0.864	1.846
40-44	0.960	0.972	3.551	0.864	0.861	1.820
45-49	0.944	0.972	3.545	0.874	0.865	1.855
50-54	0.942	0.972	3.538	0.842	0.866	1.863
55-59	0.905	0.972	3.560	0.793	0.868	1.880
60-64	0.878	0.972	3.558	0.697	0.868	1.884
65-69	0.816	0.786	1.299	0.505	0.392	-0.440
70-74	0.754	0.787	1.304	0.431	0.388	-0.455
75+	0.685	0.786	1.301	0.311	0.381	-0.484

Aldersgruppe	Veiet					
	Menn			Kvinner		
	P(fk)-uskalert	P(fk)-skalert	U(fk)-skalert	P(fk)-uskalert	P(fk)-skalert	U(fk)-skalert
20-24	0.895	0.935	2.660	0.823	0.850	1.736
25-29	0.940	0.976	3.694	0.877	0.908	2.292
30-34	0.964	0.978	3.791	0.875	0.908	2.288
35-39	0.967	0.972	3.542	0.876	0.868	1.881
40-44	0.963	0.972	3.558	0.861	0.863	1.841
45-49	0.944	0.972	3.546	0.888	0.868	1.886
50-54	0.939	0.972	3.543	0.840	0.867	1.872
55-59	0.905	0.973	3.573	0.772	0.869	1.892
60-64	0.873	0.972	3.565	0.686	0.869	1.889
65-69	0.813	0.786	1.299	0.496	0.395	-0.425
70-74	0.742	0.787	1.305	0.444	0.389	-0.452
75+	0.682	0.787	1.306	0.305	0.384	-0.473

TØI rapport 606/2002

Tabell V4.8: Personer i husholdninger med 2 personer over 18 år. Gjennomsnittlige verdier på sannsynligheter for førerkort og nyttefunksjoner for førerkort.

Aldersgruppe	Uveiet					
	Menn			Kvinner		
	P(fk)-uskalert	P(fk)-skalert	U(fk)-skalert	P(fk)-uskalert	P(fk)-skalert	U(fk)-skalert
20-24	0.867	0.864	2.072	0.812	0.795	1.491
25-29	0.949	0.962	3.537	0.921	0.924	2.664
30-34	0.977	0.965	3.610	0.945	0.925	2.676
35-39	0.984	0.981	4.222	0.946	0.901	2.378
40-44	0.987	0.981	4.227	0.943	0.900	2.361
45-49	0.984	0.981	4.227	0.929	0.903	2.397
50-54	0.979	0.981	4.234	0.891	0.902	2.381
55-59	0.976	0.981	4.211	0.820	0.902	2.379
60-64	0.965	0.981	4.234	0.725	0.901	2.374
65-69	0.931	0.901	2.378	0.579	0.451	-0.112
70-74	0.916	0.899	2.359	0.448	0.457	-0.092
75+	0.867	0.899	2.353	0.280	0.456	-0.095
Veiet						
20-24	0.873	0.865	2.080	0.814	0.799	1.509
25-29	0.948	0.962	3.551	0.921	0.925	2.670
30-34	0.976	0.965	3.613	0.944	0.926	2.678
35-39	0.984	0.981	4.218	0.943	0.902	2.384
40-44	0.987	0.981	4.229	0.945	0.900	2.365
45-49	0.984	0.981	4.229	0.929	0.903	2.397
50-54	0.980	0.981	4.233	0.890	0.902	2.386
55-59	0.974	0.981	4.215	0.821	0.902	2.383
60-64	0.963	0.981	4.232	0.727	0.901	2.372
65-69	0.928	0.901	2.378	0.572	0.451	-0.111
70-74	0.911	0.899	2.353	0.446	0.457	-0.092
75+	0.868	0.899	2.357	0.273	0.455	-0.099

TØI rapport 606/2002

Tabell V4.9: Personer i husholdninger med 3+ personer over 18 år. Gjennomsnittlige verdier på sannsynligheter for førerkort og nyttefunksjoner for førerkort

Aldersgruppe	Uveiet					
	Menn			Kvinner		
	P(fk)uskalert	P(fk)-skalert	U(fk)-skalert	P(fk)uskalert	P(fk)-skalert	U(fk)-skalert
18-19	0.523	0.549	0.242	0.480	0.482	-0.053
20-24	0.880	0.875	2.028	0.837	0.825	1.620
25-29	0.936	0.944	3.006	0.897	0.906	2.396
30-34	0.970	0.955	3.259	0.943	0.914	2.483
35-39	0.977	0.982	4.200	0.930	0.870	2.027
40-44	0.984	0.981	4.163	0.915	0.878	2.113
45-49	0.984	0.980	4.141	0.901	0.878	2.116
50-54	0.983	0.980	4.136	0.858	0.878	2.118
55-59	0.976	0.980	4.116	0.745	0.886	2.194
60-64	0.963	0.981	4.170	0.676	0.883	2.164
65-69	0.924	0.880	2.137	0.203	0.074	-2.486
70-74	0.845	0.854	1.891	0.223	0.117	-1.985
75+	0.759	0.887	2.216	0.020	0.098	-2.206
Veiet						
18-19	0.520	0.553	0.259	0.478	0.483	-0.047
20-24	0.882	0.877	2.051	0.844	0.832	1.673
25-29	0.932	0.945	3.031	0.901	0.910	2.455
30-34	0.967	0.952	3.180	0.948	0.917	2.521
35-39	0.975	0.981	4.185	0.925	0.869	2.013
40-44	0.985	0.981	4.202	0.918	0.880	2.128
45-49	0.984	0.981	4.159	0.908	0.882	2.158
50-54	0.982	0.981	4.156	0.860	0.879	2.130
55-59	0.975	0.981	4.160	0.750	0.894	2.285
60-64	0.962	0.981	4.187	0.662	0.877	2.106
65-69	0.919	0.880	2.133	0.203	0.074	-2.486
70-74	0.848	0.853	1.882	0.248	0.124	-1.920
75+	0.727	0.886	2.193	0.027	0.104	-2.133

TØI rapport 606/2002



## 5 Kohortmodeller for førerkortinnehav

Dagens nasjonale modellsystem (NTM 4c) inneholder egentlig tre uavhengige modeller for innehav av førerkort. Det er to disaggregerte logitmodeller hvor den ene predikerer forventet antall førerkort for husholdningens "hovedperson" og partner, og den andre predikerer ytterligere førerkort i husholdningen. I tillegg er det en aggregert modell som tar seg av trendeffektene når det gjelder innehav av førerkort. Dette er en tidsrekkemodell hvor vi får tatt hensyn til kohorteffektene i førerkortinnehavet, eksemplifisert ved at dagens 50 åringer om 20 år vil ha høyere førerkortandeler enn dagens pensjonister. Denne type effekter vil nødvendigvis påvirke fremtidige reisevaner i relativt stor grad.

De to disaggregerte logitmodellene er estimert i fase 3 av modellutviklingen på data fra den nasjonale reisevaneundersøkelsen gjennomført i 1991/92, mens kohortmodellen er estimert (i 1991, dvs. i fase 2 av modellutviklingen) på tidsrekke-data for perioden 1980-1991.

Det er grunn til å tro at andelen hushold uten, og med ett, to eller flere førerkort har endret seg i perioden fra begynnelsen av 90 tallet og frem til 1997/1998. Disse andelenene endret seg betraktelig i perioden mellom de to forrige reisevaneundersøkelsene fra 84/85 til 91/92. Dessuten har gjennomsnittlig husholdsstørrelse endret seg i perioden. Større grad av sentralisering i løpet av 90 tallet, ser også ut til å ha påvirket yngre individers anskaffelse av førerkort. I Sverige er denne effekten dokumentert og tatt hensyn til bl.a. i innretningsplanleggingen i pågående planarbeid.

### 5.1 Tidsrekkemodell for utvikling i førerkortandeler

Dette er egentlig en prosedyre for fremskriving av førerkortandelene i befolkningen og i mindre grad en modell. Foreliggende prosedyre er dokumentert i Rand (1991). Resultatene fra prosedyren er ikke benyttet direkte i modellsystemet. Man har imidlertid gjennomført kalibreringer av de disaggregerte logitmodellene slik at disse treffer best mulig de førerkortandeler som predikeres i prosedyren. Følgende punkter oppsummerer prosedyren.

- Data for antall førerkort i befolkningen fordelt på kjønn og 5 grove aldersgrupper er samlet inn for 1980, 1985 og 1990. Befolkningstall fordelt på kjønn og alder med 5 års intervaller er samlet inn for samme årstall.
- Førerkortandeler for de 5 grove kjønnsfordelte aldersgrupper og for de 3 årstall er beregnet på bakgrunn av gjennomsnittsalder innenfor de grove aldersintervallene (fra de finere fordelte befolkningstallene) og befolkningsstørrelsen innenfor hvert av dem. Polynomer er tilpasset disse observasjonene under forutsetning av at ingen har førerkort når de runder 90 år. Polynomene benyttes til å beregne førerkortandeler innenfor de finere aldersgruppene.
- To typer akkvisisjonssammenhenger og en sammenheng for bortfall er forutsatt som trendeffekter<sup>6</sup>, Type 1 er antatt å gjelde for aldersgruppene 15-19 år og 20-24 år, type 2 gjelder for aldersgruppene mellom 25 og 49 år for menn og 55 år for kvinner, og type 3 gjelder fra den alderen for menn og kvinner hvor type 2 slutter.
  - Type 1:  $P_{c,t} = P_{c,t-1} + (A_{c,t} * (S - P_{c,t-1}))$
  - Type 2:  $P_{c,t} = P_{c-1,t-1} + (A_{c,t} * (S - P_{c,t-1}))$
  - Type 3:  $P_{c,t} = P_{c-1,t-1} * (1 + L_{c,t})$

Disse sammenhengene er benyttet til å predikere førerkortandeler fordelt på kjønn og 5 års aldersgrupper for et antall prognoseår

### 5.2 Nye tidsrekkemodeller

#### 5.2.1 Datatilgang tidsrekker

Opplysningsrådet for vegtrafikken (OFV) utgir årlig kjønns- og aldersfordelte tall for førerkortinnehav i Norge. Aldersgruppene er imidlertid relativt grove (18-24, 25-34, 35-49, 50-64 og 65+). Aldersfordelte tall ble første gang utgitt i 1988. For årstallene 1980 og

<sup>6</sup> c = aldersgruppe/kjønn, t = tidspunkt,  $P_{c,t}$  = førerkortandel for gruppe c på tidspunkt t,  $A_{c,t}$  = akkvisisjonsrate for gruppe c på tidspunkt t, S = metningspunkt (98% av befolkningen),  $L_{c,t}$  = tapsrate for gruppe c på tidspunkt t.

1985 har vi imidlertid opplysninger innhentet ved den forrige oppdateringen av tverrsnittmodellene (Rand, 1991). For de mellomliggende år har vi benyttet trendfunksjonen i Excel til å anslå tallene for antall førerkort.

I statistisk årbok finnes alders- og kjønnsfordelte befolkningstall. Tabellene her oppgir befolkningen fordelt på 5 års aldersgrupper. De 4 siste årbøker finnes på Statistisk Sentralbyrås (SSB) hjemmeside på Internett. For de resterende år har vi punchet tabellene tilbake til 1980.

I vedlegg 5 vises det omtalte datagrunnlaget for hhv. kvinner og menn. Her har vi også beregnet gjennomsnittsalder for OFVs grovere aldersgrupper og førerkortandeler for disse aldersgruppene.

### 5.2.2 Anskaffelse av førerkort for de yngste aldersgruppene

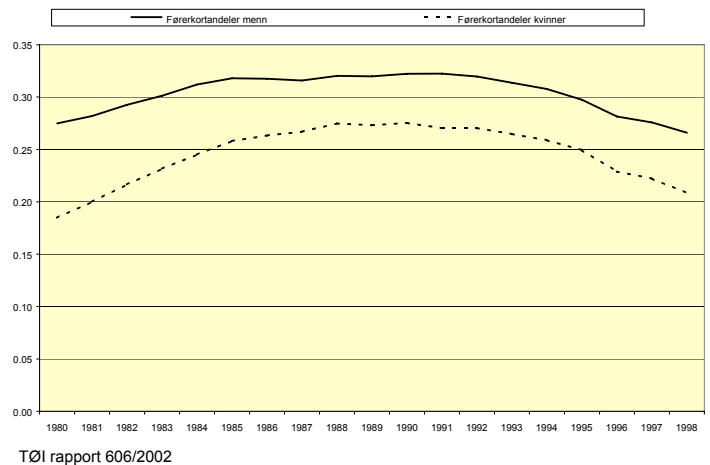
I Figur 5.1 kan vi følge utviklingen i førerkortandeler for de yngste aldersgruppene (15-19 år) fra år til år i perioden 1980-1998<sup>7</sup>. Figuren viser en jevn økning for begge kjønn frem til 1986 hvor utviklingen stagnerer. Økningen i denne perioden var sterkest blant kvinner. Etter toppen i 1992 har førerkortandelene sunket jevnt til et nivå som for menn om lag tilsvarer andelen i 1980, og for kvinner i 1982.

Spørsmålet er nå hvilke forhold i samfunnsutviklingen og betingelsene for disse aldersgruppene som har påvirket denne utviklingen. For det første har det skjedd en sentralisering i perioden. Andelen av befolkningen som er bosatt i tettbygd strøk (SSBs klassifisering) er øket fra rundt 70 % i 1980 til om lag 74 % i 1998. Det er grunn til å tro at ungdom er overrepresentert i fraflytting fra distriktene til tettstedene. Dette kan påvirke ungdoms behov for førerkort, både fordi tettstedene har et mer differensiert tilbud av varer og tjenester i korte avstander, og fordi tettbygd stør oftest har bedre kollektivtilbud enn i distriktene.

For det andre har andelen av befolkningen under videregående utdanning økt fra ca 6 % i 1980 til vel 9 % i 1998. I studietiden hvor man lever på studielån, foreldres inntekt eller inntekt fra ekstraarbeid kan det være vanskelig å skaffe midler til erverv av førerkort.

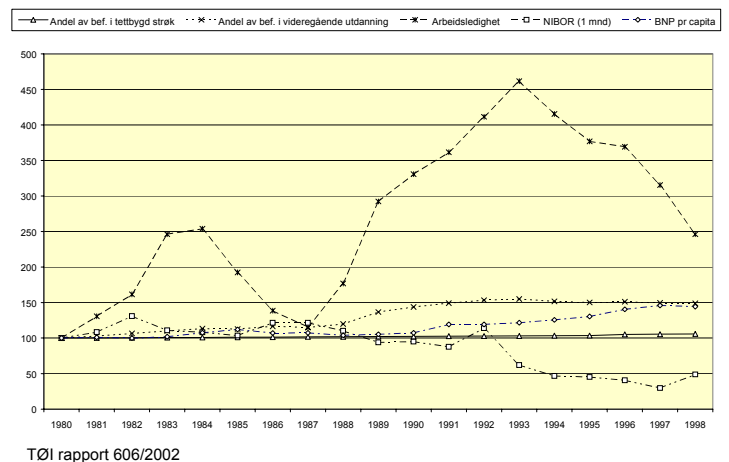
I perioden har det imidlertid også vært økonomisk vekst. Dette kan ha vært en av drivkreftene bak økningen av førerkortinnnehav i perioder med sterk

vekst. Kortsiktige statsrenter er også en indikator som gir et bilde av konjunktursvingninger.



Figur 5.1 Utvikling i førerkortandeler for de yngste aldersgruppene 15-19 år (1980-1998)

I Figur 5.2 vises utviklingen i de indikatorer som er omtalt for perioden 1980-1998.



Figur 5.2: Utvikling i noen "konjunkturindikatorer" i perioden 1980-1998

I den nasjonale transportmodellen skal tidsrekke-modellen i første rekke beskrive kohorteffektene som påvirker andelen av befolkningen med førerkort over tid. Tverrsnittmodellene skal ta for seg de resterende effekter, som skyldes økonomiske forhold, befolkningssammensetning, m.m. I Figur 5.1 har vi imidlertid sett at førerkortandelene for de yngste aldersgruppene varierer sterkt over tid noe kan knyttes til endringer i betingelsene for disse gruppene. Vi ønsker å utvikle en metodikk hvor vi kan skille ut trendeffekter som kan

<sup>7</sup> Vi har opplysninger om antall førerkort i aldersgruppen 18-24 og befolkningsmengden i aldersgruppene 15-19 og 20-24. Vi har derfor antatt at årskullfordelingen innenfor aldersgruppen 15-19 år er uniform. Det er bare årskullene 18 og 19 i denne gruppen som har førerkort. Førerkortandelene i aldersgruppen 15-19 er derfor tilpasset slik at førerkortandelene i aldersgruppen 15-24 stemmer overens med statistikken for hvert år fra 1980 til 1998.

knyttet til endringer i bakenforliggende forhold, som skal ivaretas av tverrsnittsmoellene, og kohorteffektene som skal ivaretas av tidsrekke-modellen.

For de yngste aldersgruppene kan vi for eksempel anta at utviklingen i førerkortandelene følger en enkel lineær sammenheng:

$$FA_{kt} = \alpha_k + \beta_{1,k} * x_{1t} + \beta_{2,k} * x_{2t} + \dots + \beta_{i,k} * x_{it} + e_{k,t}, e_{k,t} = \rho_k e_{k,t-1} + v_{k,t}$$

Her er  $\alpha$  og  $\beta$  parametere som angir betydningen av de tilhørende variable  $x_{it}$ . Fotskriften "k" angir kjønn, "i" angir variabelnummer og "t" angir årstall. Til sist i den lineære sammenhengen er det et stokastisk restledd  $e_{k,t}$  (residual) som vi antar er en 1. ordens autordegresjon over tidsrekken. Parameteren  $\rho$  i denne sammenhengen er ikke kjent men må estimeres, og "v<sub>k,t</sub>" er et stokastisk restledd. Sammenhengen over kan estimeres med minste kvadraters metode. Gjennomføres en slik estimering med tidsrekkeene vist i Figur 5.2 som høyre-sidevariable og førerkortandelene for hhv. menn og kvinner vist i Figur 5.1 som venstresidevariable finner vi følgende parameterverdier:

Tabell 5.1: OLS-estimerer på parametre for arbeidsledighet, andelen av befolkningen i videregående utdanning og andelen av befolkningen i tettbygd strøk

	$\beta_{3,k}$	$\beta_{2,k}$	$\beta_{1,k}$	$\alpha_k$
Variabel	Arbeidsledighet	Andel av befolkningen i videregående utdanning	Andel av befolkningen i tettbygd strøk	Konstant
Menn	-0.01	2.77	-2.45	1.86
Kvinner	-0.01	4.56	-2.97	2.05

TØI rapport 606/2002

Vi ser i Tabell 5.1 at arbeidsledigheten ser ut til å ha en negativ innvirkning på førerkortandelene i de yngste aldersgruppene, at utviklingen i befolkningsandelen i videregående utdanning har hatt en positiv innvirkning og at andelen av befolkningen i tettbygd strøk ser ut til å ha hatt en negativ innvirkning på innehavet av førerkort.

Nå er vi videre interessert i å estimere autokorrelasjonen i restleddet "e<sub>k,t</sub>". De estimerte restleddene er vist i Tabell 5.2. Autokorrelasjonen "ρ<sub>k</sub>" er beregnet som forholdet mellom de to nest siste størrelsene i tabellen. Vi kan nå anslå verdiene på "e<sub>k,t</sub>" på bakgrunn av "ρ<sub>k</sub>" og "e<sub>k,t-1</sub>". Denne prosedyren kalles "random tracking". Den skiller altså mellom de to

stokastisk komponentene "e<sub>k,t</sub>", som er autokorrelert og "v<sub>k,t</sub>" som er helt tilfeldig.

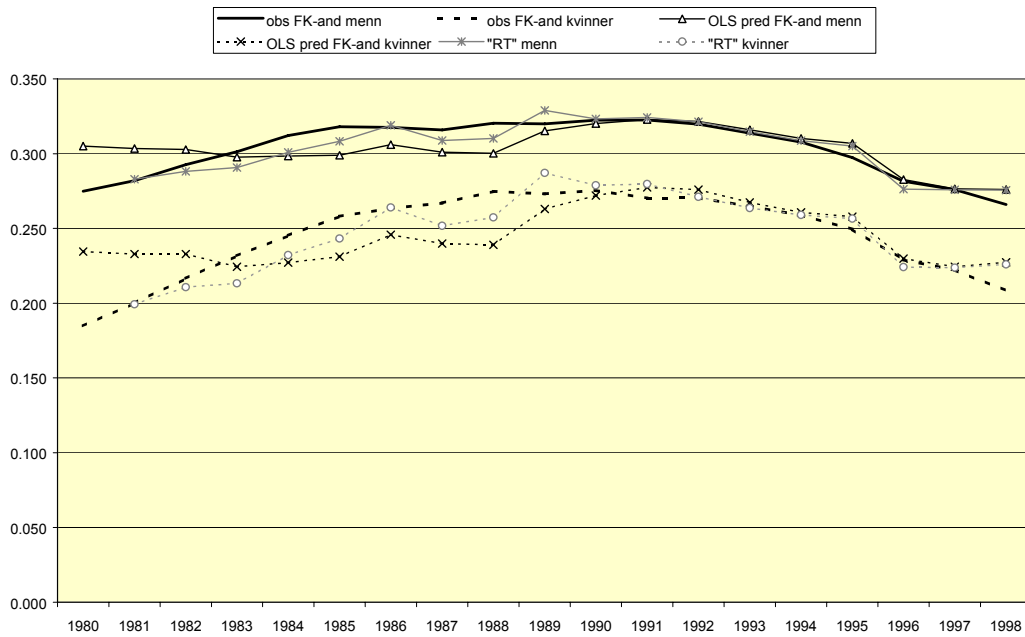
Tabell 5.2: Estimerte residualer i regresjons-sammenhengene vist i Tabell 5.1

Årstall	Estimerte residualer	
	menn	kvinner
1980	0.0302	0.0498
1981	0.0215	0.0328
1982	0.0101	0.0163
1983	-0.0037	-0.0074
1984	-0.0137	-0.0180
1985	-0.0191	-0.0270
1986	-0.0116	-0.0176
1987	-0.0148	-0.0271
1988	-0.0201	-0.0358
1989	-0.0046	-0.0102
1990	-0.0022	-0.0033
1991	0.0003	0.0070
1992	0.0018	0.0055
1993	0.0022	0.0027
1994	0.0024	0.0019
1995	0.0094	0.0084
1996	0.0011	0.0008
1997	0.0007	0.0022
1998	0.0100	0.0191
$\Sigma e_{t-1}$	-0.0020	-0.0051
$\Sigma(e_{t-1})^2$	0.0029	0.0075
$\rho_k$	-0.6777	-0.6753

TØI rapport 606/2002

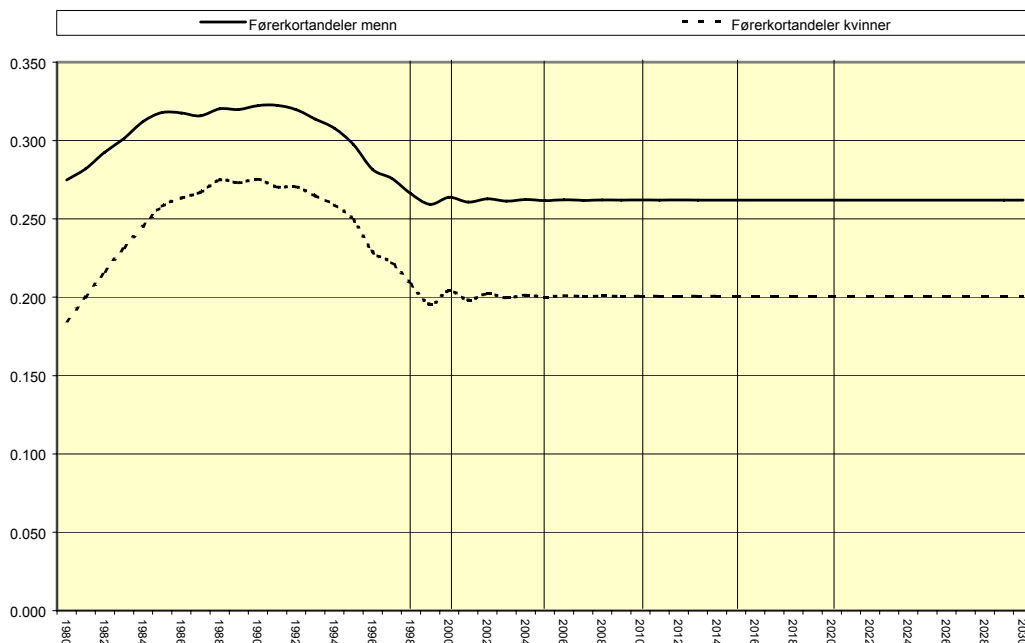
I Figur 5.3 vises observerte og OLS-predikerte, med og uten "random tracking" korleksjon (RT), førerkortandeler. Vi ser at metoden gir godt samsvar med observerte andeler, spesielt mot slutten av perioden.

Hvis vi nå, for prognoseformål med de estimerte modellene, antar at arbeidsledigheten, andelen av befolkningen i tettbygd strøk og andelen i videregående utdanning, blir konstant lik verdiene i 1998, kan vi benytte autokorrelasjonen "ρ<sub>k</sub>", til å oppdatere førerkortandelene. Vi ser av Figur 5.4 at førerkortandelene under disse forutsetningene "konvergerer" mot 0.26 for menn og 0.20 for kvinner. Konvergens oppnås som vi ser før 2010. Vi merker oss altså at førerkortandelene i aldersgruppen 15-19 for menn vil være lavere i fremtiden enn den var i 1980, mens de for kvinner vil være noe høyere enn i 1980, under forutsetning av at konjunkturindikatorerne holdes konstant.



TØI rapport 606/2002

Figur 5.3: Observerte, OLS-predikerte, og OLS "RT"-korrigerede førerkortandeler for menn og kvinner i perioden 1980-1998



TØI rapport 606/2002

Figur 5.4 Prognoser for førerkortandeler for de yngste aldersgrupper (15-19 år) ved konstant andel av befolkningen i tettbygd strøk, og i videregående utdanning.

Tabell 5.3: Utvikling i førerkortandeler for menn og kvinner i aldersgruppen 15-19 år

	1980	1985	1990	1995	2000	2005	2010	2015	2020	2025	2030
Menn	0,275	0,318	0,322	0,297	0,264	0,262	0,262	0,262	0,262	0,262	0,262
Kvinner	0,185	0,258	0,275	0,249	0,204	0,200	0,201	0,201	0,201	0,201	0,201

TØI rapport 606/2002

### 5.2.3 Utvikling i førerkortinnnehav for de resterende aldersgrupper

Data for førerkortinnnehav finnes bare for de relativt grove aldersgrupper publisert av OFV. Med utgangspunkt i de tabeller som finnes vedlegg til kapittel 5 kan man imidlertid estimere andeler for finere aldersgrupper. Dette kan gjøres ved å tilpasse polynomer til de observasjoner vi har for de grove aldersgruppene. For hvert årstall fra 1980 til 1998 har vi beregnet førerkortandelene i aldersgruppene 15-24, 25-34, 35-49, 50-64 og 65 og eldre. Når vi i tillegg både for menn og kvinner forutsetter at førerkortandelen er 0 ved 95 år, får vi for hvert år 6 ”observasjoner” av førerkortandelene. Dette er nok observasjoner til å tilpasse et 5. grads polynom på formen:

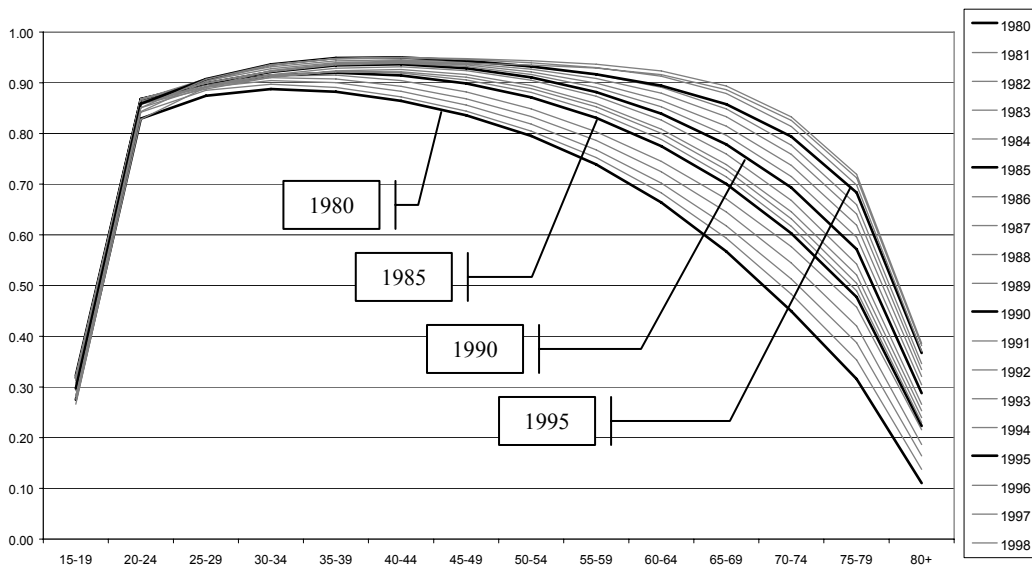
$$FA(a) = k_0 + k_1 * a + k_2 * a^2 + k_3 * a^3 + k_4 * a^4 + k_5 * a^5$$

Her er ”k”-ene parametre som må tilpasses og ”a” er alder. Dette gjøres enklest ved hjelp av lineær algebra. Vi har gjennomført dette på alle årstall i perioden fra 1980-1998. Resultatene er vist i Figur 5.5 og Figur 5.6. De uthevede kurvene representerer hele 5 års inter-

valler. Kurven for 1980 er den første uthevede kurven. Vi ser at utviklingen har vært forskjellig for menn og kvinner. Situasjonen i 1998 for kvinner ligner på den situasjonen vi kan observere for menn i 1980.

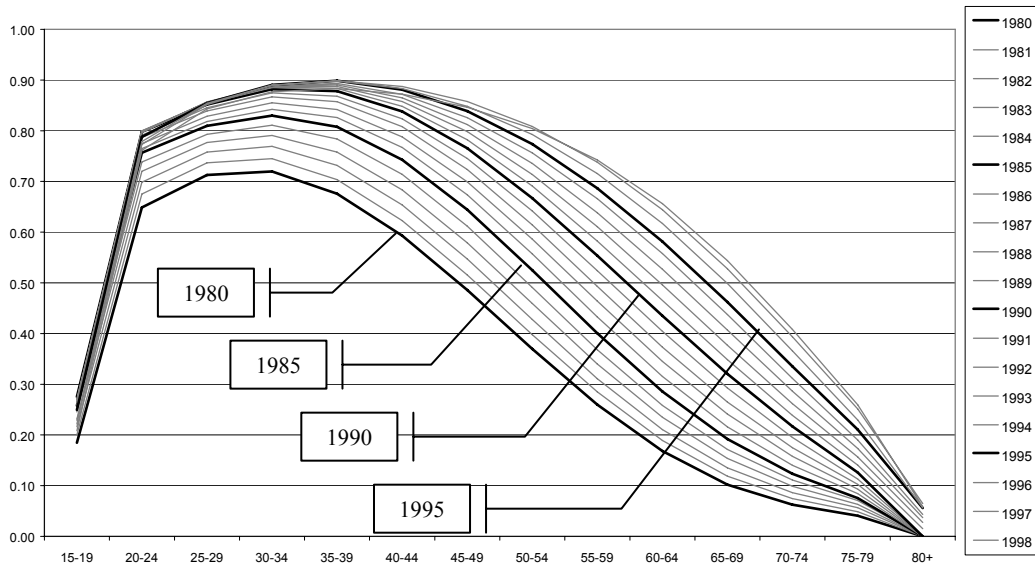
Formlene for de estimerte kurvene (én formel for hvert årstall fra 1980 til 1998 for kvinner og menn) kan nå benyttes til å beregne førerkortandeler i 5 års aldersgrupper. Ved å studere hvert 5. år kan vi da følge utviklingen i førerkortandelene for en aldersgruppe over tid.

I tabell 5.4 vises førerkortandelene for 5 års aldersgrupper beregnet med de estimerte polynomer. I tabellen kan vi følge utviklingen i en aldersgruppes førerkortandeler over tid. For eksempel er menn i aldersgruppen 15-19 i 1980 blitt 30-34 i 1995. Som vi ser når vi følger denne aldersgruppen diagonalt nedover mot høyre i tabellen har førerkortandelene for disse personene økt fra 0.37 i 1985, via 0.87 og 0.90, til 0.94 i 1995. Menn som var mellom 55 og 59 år i 1998 har kommet til aldersgruppen 70-74 i 1995. Andelen i denne kohorten med førerkort har i perioden økt fra 74 % til 78 %.



TØI rapport 606/2002

Figur 5.5: Tilpassede polynomer for førerkortandeler etter alder for menn, 1980-1998



TØI rapport 606/2002

Figur 5.6: Tilpassede polynomer for førerkortandeler etter alder for kvinner, 1980-1998

Tabell 5.4: Førerkortandeler for 5 års aldersgrupper i 5 års intervaller

Alder	Menn				Kvinner			
	1980	1985	1990	1995	1980	1985	1990	1995
15-19	0.27	0.32	0.32	0.30	0.18	0.26	0.28	0.25
20-24	0.83	0.87	0.87	0.86	0.65	0.76	0.80	0.79
25-29	0.87	0.90	0.90	0.91	0.71	0.81	0.85	0.85
30-34	0.89	0.91	0.92	0.94	0.72	0.83	0.88	0.89
35-39	0.88	0.92	0.93	0.95	0.68	0.81	0.88	0.90
40-44	0.86	0.91	0.94	0.95	0.59	0.74	0.84	0.88
45-49	0.84	0.90	0.93	0.94	0.49	0.64	0.77	0.84
50-54	0.80	0.87	0.91	0.93	0.37	0.53	0.67	0.77
55-59	0.74	0.83	0.88	0.92	0.26	0.40	0.55	0.69
60-64	0.66	0.78	0.84	0.89	0.17	0.29	0.43	0.58
65-69	0.57	0.70	0.78	0.86	0.10	0.19	0.32	0.46
70-74	0.45	0.60	0.69	0.78	0.06	0.12	0.22	0.34
75-79	0.32	0.48	0.57	0.66	0.04	0.08	0.13	0.21
80+	0.11	0.22	0.29	0.34	0.00	0.00	0.00	0.06

TØI rapport 606/2002

I tabell 5.4 vises førerkortandelene for 5 års aldersgrupper beregnet med de estimerte polynomer. I tabellen kan vi følge utviklingen i en aldersgruppes førerkortandeler over tid. For eksempel er menn i aldersgruppen 15-19 i 1980 blitt 30-34 i 1995. Som vi ser når vi følger denne aldersgruppen diagonalt nedover mot høyre i tabellen har førerkortandelene for disse personene økt fra 0.37 i 1985, via 0.87 og 0.90, til 0.94 i 1995. Menn som var mellom 55 og 59 år i 1998 har kommet til aldersgruppen 70-74 i 1995.

Andelen i denne kohorten med førerkort har i perioden økt fra 74 % til 78 %.

### 5.2.4 Prognoser for førerkortandeler

For å utarbeide prognoser for førerkortandelene i fremtidige år behøver vi estimater på tilvekst og bortfall av førerkort i de ulike aldersgrupper. Ved å studere Tabell 5.4 kan vi danne oss et bilde av hvordan dette har artet seg de siste 25 år. Hvis vi for alders-

gruppene fra 20 år og oppover, forutsetter de samme trendeffekter som før, dvs:

$$P_{c,t} = P_{c-1,t-1} + (A_{c,t} * (S - P_{c-1,t-1}))$$

$$P_{c,t} = P_{c-1,t-1} * (1 + L_{c,t})$$

kan vi beregne akkvisisjonsrater og tapsrater med følgende formler:

$$A_{c,t} = (P_{c,t} - P_{c-1,t-1}) / (S - P_{c-1,t-1})$$

$$L_{c,t} = (P_{c,t} / P_{c-1,t-1}) - 1$$

Den første sammenhengen I, sier at førerkortandelen for aldersgruppe c i tidsperiode t, er lik med førerkortandelen for den samme aldersgruppen men i tidsperioden før, hvor personene i aldersgruppen også er yngre (c-1) pluss anskaffelsen av førerkort,  $A_{c,t}$  (for den andel av aldersgruppen som ikke allerede har førerkort). Dette leddet inkluderer et metningspunkt "S", som gjenspeiler at en liten andel av befolkningen av helsemessige årsaker ikke har mulighet til å skaffe seg førerkort. Den neste sammenhengen II, trer i kraft når førerkortandelen synker for en aldersgruppe fra en periode til den neste (dvs. diagonalt nedover mot høyre i Tabell 5.4). Tapsratene  $L_{c,t}$ , skyldes at personene av ulike årsaker mister sitt førerkort.

I Tabell 5.5 vises akkvisisjons og tapsrater for menn og kvinner beregnet ut i fra datamaterialet med

formlene III og IV. De verdier man i gjennomsnitt får på  $A_{c,t}$  og  $L_{c,t}$  i det datamaterialet vi har skaffet til veie, vist i som de siste kolonner i tabellen. Som vi ser har det i perioden 1980 til 1995 vært en relativt stor tilvekst i førerkortinnhavedet også i de eldre aldersgrupper, og netto tap i førerkortandelene har oppstått først i aldersgruppene fra 75 år og over.

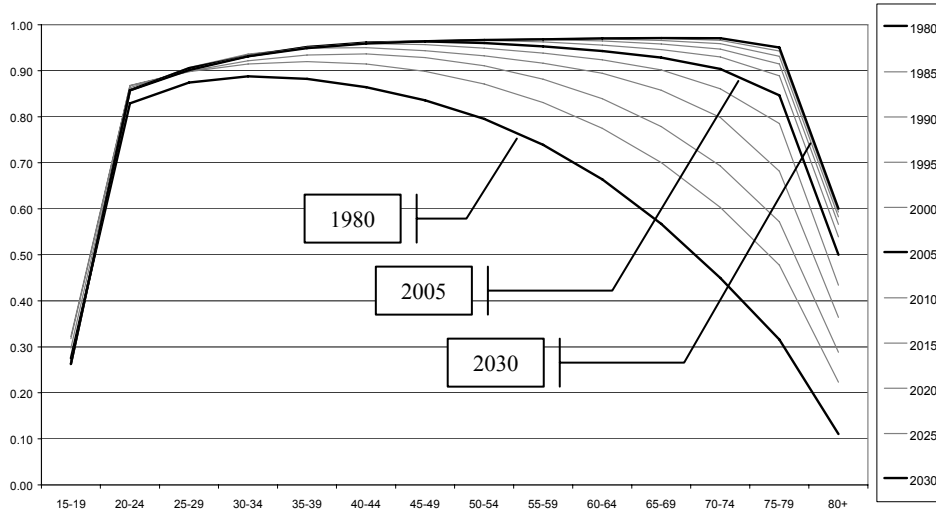
Spørsmålet vi nå bør stille er om historiske tall for tilvekst og bortfall av førerkort er egnet å benytte til prognoser for førerkortinnhavedet i fremtiden. I Tabell 5.6 gis et eksempel som kan belyse effektene av tilvekst og bortfall av førerkort. Vi tenker oss at vi observerer 100 kvinner og menn fra de er mellom 15 og 19 år og til de blir over 80 år. Vi tenker oss videre at 26 av mennene og 21 av kvinnene tar førerkort før de blir 20 år. Da blir det igjen hhv. 74 menn og 79 kvinner som ikke har tatt førerkort før de blir 20. Videre i tabellen er gjennomsnittsverdiene for kvinner og menn i Tabell 5.5 benyttet til å beregne tilvekst og bortfall. Vi ser at 60 av mennene som ikke har tatt førerkort før de er 20 år tar førerkort innen fylte 25 år. Av de 100 menn vi studerer er det da bare 14 igjen uten førerkort, mens vi ser at det er 22 kvinner igjen uten førerkort. Vi ser imidlertid at kvinnene så å si har tatt igjen mennene før fylte 35 år. I andre enden av aldersskalaen ser vi at kvinner i større grad mister førerkortet enn sine mannlige jevnaldrende.

Tabell 5.5: "Observerte" rater for tilvekst og bortfall av førerkort

Alder	Menn				Kvinner			
	1980-1985	1985-1990	1990-1995	Gjennomsnitt	1980-1985	1985-1990	1990-1995	Gjennomsnitt
20-24	0.84	0.83	0.82	0.83	0.72	0.75	0.73	0.73
25-29	0.46	0.28	0.35	0.36	0.49	0.43	0.31	0.41
30-34	0.38	0.29	0.46	0.38	0.44	0.42	0.30	0.39
35-39	0.35	0.30	0.47	0.37	0.34	0.32	0.17	0.28
40-44	0.33	0.28	0.35	0.32	0.22	0.18	0.03	0.14
45-49	0.30	0.22	0.16	0.22	0.13	0.09	0.00	0.08
50-54	0.25	0.15	0.07	0.15	0.08	0.07	0.04	0.06
55-59	0.19	0.10	0.09	0.12	0.05	0.06	0.06	0.06
60-64	0.15	0.05	0.13	0.11	0.04	0.06	0.07	0.05
65-69	0.12	0.02	0.13	0.09	0.03	0.05	0.05	0.04
70-74	0.09	-0.01	0.00	0.03	0.02	0.03	0.02	0.03
75-79	0.05	-0.05	-0.05	-0.02	0.01	0.02	-0.03	0.00
80+	-0.29	-0.40	-0.40	-0.36	-1.00	-1.00	-0.56	-0.85

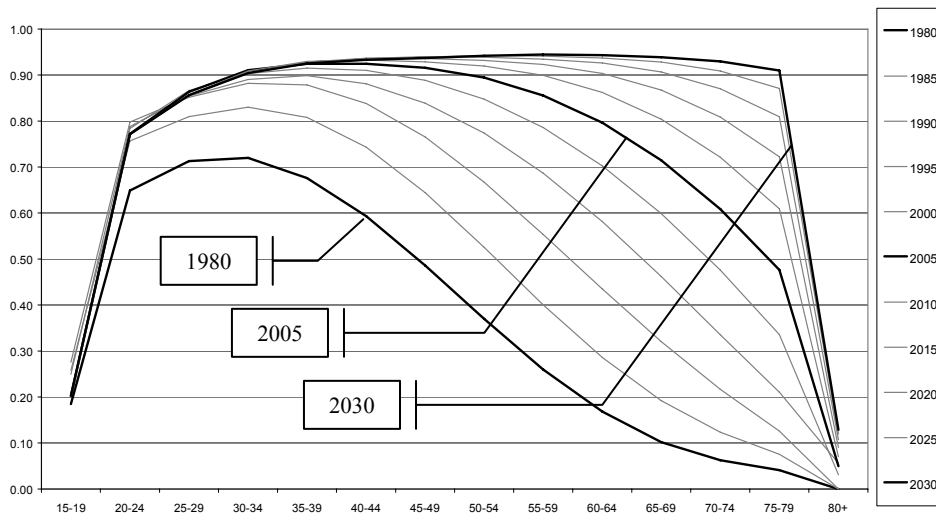
TØI rapport 606/2002





TØI rapport 606/2002

Figur 5.7: Prognoser for førerkortinnhav blant menn med "observerte" akkvisisjons- og tapsrater



TØI rapport 606/2002

Figur 5.8: Prognoser for førerkortinnhav blant kvinner med "observerte" akkvisisjons- og tapsrater

Det datamaterialet vi har til rådighet er for tynt til å kunne fastsette fremtidige akkvisisjons- og tapsrater med stor sikkerhet. Vi blir derfor nødt til å benytte en viss grad av skjønn. Tallene i Tabell 5.5 gir bare en viss pekepinn på hvordan ratene bør være. For prognoseformål velger vi å benytte ratene i Tabell 5.7. Hovedtendensene som ligger i disse ratene er at kvinner anskaffer seg førerkort senere enn menn. Ved rundt 35 år har de imidlertid langt på veg tatt igjen sine mannlige jevnaldrende. Fra 35 år og videre forutsetter vi at utviklingen blant menn og kvinner er den samme.

Vi ser tendenser til dette for aldersgruppene 50 år og eldre hvis vi sammenlikner kolonnene for menn og kvinner for 1990-1995 i Tabell 5.5. Vi forutsetter

videre en noe kraftigere tendens for tap av førerkort enn det som ligger i de observerte tallene. Bakgrunnen til dette er signaler fra trafikksikkerhetsforskere om at man fremover bør bli mer restriktive i forhold til kontroll av de eldres helsemessige forutsetninger for å kjøre bil. Selv om man fremover kanskje vil oppleve at de eldres helse vil forbedres, tror vi at kontrollen med de eldres kjøreferdigheter vil resultere i høyere tapsrater blant personer fra rundt 70 år og oppover. Det er her ingen grunn til å forvente store forskjeller blant kvinner og menn, selv om de observerte tallene viser en viss forskjell. Vi tror at disse forskjellene vil jevne seg ut, noe de observerte tallene også antyder.



Tabell 5.6: Utvikling i førerkortinnhøvet for 100 kvinner og menn ved ”observerte” akkvisisjons- og tapsrater

Aldersgruppe	Menn				Kvinner			
	Tar førerkort	Har førerkort	Har ikke førerkort	Sum	Tar førerkort	Har førerkort	Har ikke førerkort	Sum
15-19	26	26	74	100	21	21	79	100
20-24	60	86	14	100	57	78	22	100
25-29	5	91	9	100	9	86	14	100
30-34	3	94	6	100	5	92	8	100
35-39	2	96	4	100	2	94	6	100
40-44	1	98	2	100	1	95	5	100
45-49	1	98	2	100	0	95	5	100
50-54	0	98	2	100	0	95	5	100
55-59	0	99	1	100	0	96	4	100
60-64	0	99	1	100	0	96	4	100
65-69	0	99	1	100	0	96	4	100
70-74	0	99	1	100	0	96	4	100
75-79	-2	97	3	100	0	96	4	100
80+	-35	63	37	100	-80	16	84	100

TØI rapport 606/2002

Tabell 5.7: Forutsatte akkvisisjons- og tapsrater for prognoser av kvinners og menns fremtidige førerkortinnhøvet

Aldersgruppe	Forutsatte rater	Forutsatte rater
	menn	kvinner
20-24	0.80	0.75
25-29	0.35	0.40
30-34	0.35	0.40
35-39	0.30	0.30
40-44	0.10	0.10
45-49	0.05	0.05
50-54	0.05	0.05
55-59	0.05	0.05
60-64	0.00	0.00
65-69	0.00	0.00
70-74	-0.05	-0.05
75-79	-0.10	-0.10
80+	-0.50	-0.50

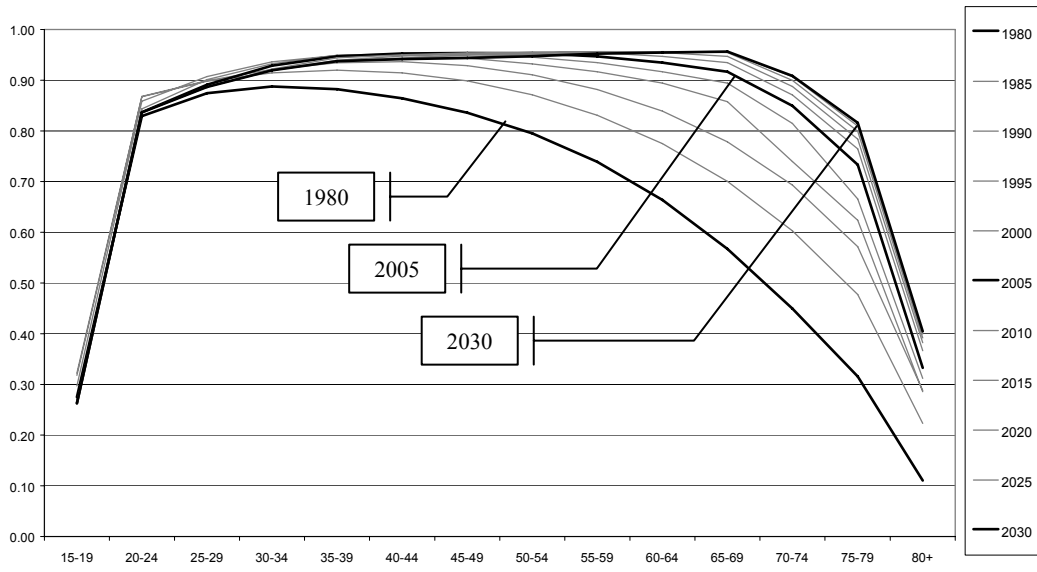
TØI rapport 606/2002

I Figur 5-9 og figur 5-10, samt i Tabell 5.8 og Tabell 5.9 vises de beregnede prognoser for utviklingen av førerkortinnhøvet i aldersgruppene fram til år 2030.

Sammenlikner vi de to figurene med de tilsvarende figurer med observerte rater, ser vi at den største forskjellen er at bortfallet blant de eldre aldersgrupper starter tidligere. Vi merker oss også at andelen blant de eldste kvinner er noe høyere, og nesten på linje med mennenes andel i år 2030.

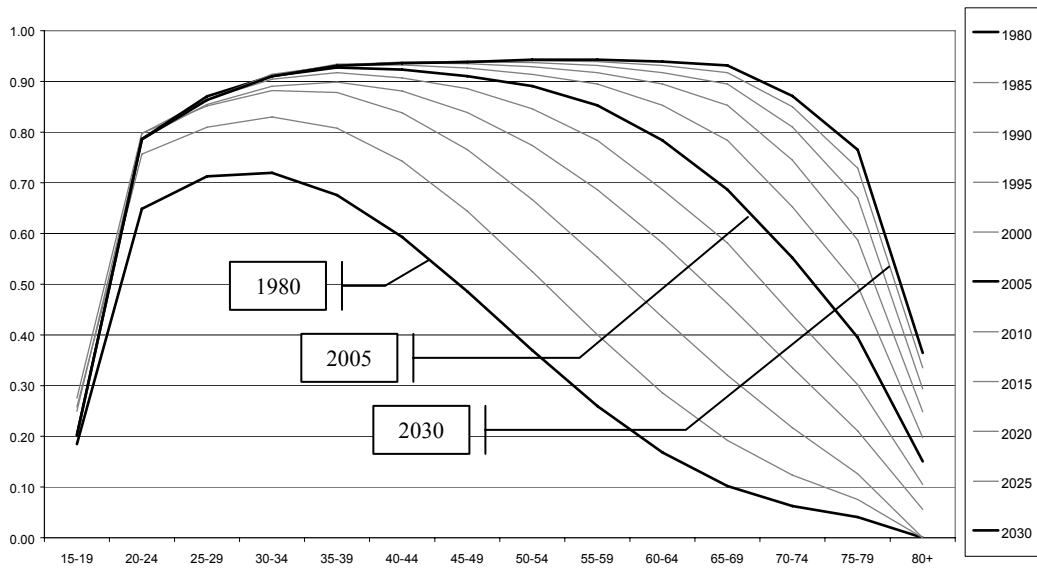
Vi vil avslutningsvis understreke at tynt data-grunnlag til etablering av rater for tilvekst og bortfall av førerkort, gjør at estimatene er usikre. Vi bør derfor følge med i hva som skjer de nærmeste årene når det gjelder endringer i innhøvet av førerkort, og oppdatere prognosene så snart nye data foreligger.

Til sist i dette avsnittet skal vi sammenlikne de tidligere prognoser for innhøvet av førerkort blant menn og kvinner med de nye. De gamle prognosene ligger til grunn for dagens nasjonale transportmodell (NTM 4c). Sammenligningene er vist i Figur 5.11 og Figur 5.12. Vi ser at de nye prognosene er noe lavere for de yngste aldersgruppene og vesentlig høyere for de eldste (over 55/60 år). For menn er prognosene for den eldste aldersgruppen (80+) noe lavere enn tidligere, mens dette er omvendt for kvinner.



TØI rapport 606/2002

Figur 5.9: Prognoser for førerkortinnnehav blant menn med forutsatte akkvisisjons- og tapsrater



TØI rapport 606/2002

Figur 5.10 Prognoser for førerkortinnnehav blant kvinner med forutsatte akkvisisjons- og tapsrater

Tabell 5.8: Observerte (1980-1995) og beregnede (2000-2030) førerkortandeler blant menn i 5 års intervaller og aldersgrupper

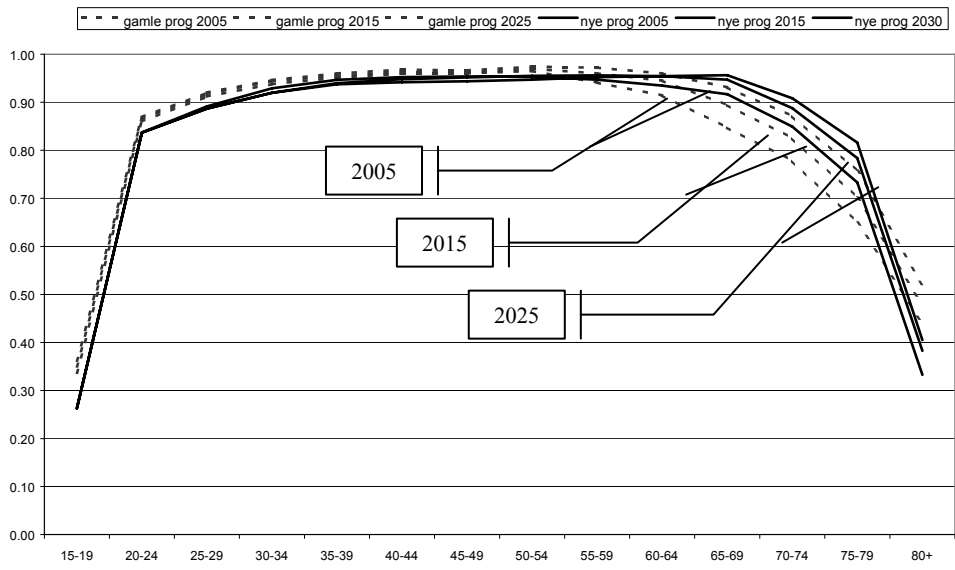
	1980	1985	1990	1995	2000	2005	2010	2015	2020	2025	2030
15-19	0.27	0.32	0.32	0.30	0.26	0.26	0.26	0.26	0.26	0.26	0.26
20-24	0.83	0.87	0.87	0.86	0.84	0.84	0.84	0.84	0.84	0.84	0.84
25-29	0.87	0.90	0.90	0.91	0.90	0.89	0.89	0.89	0.89	0.89	0.89
30-34	0.89	0.91	0.92	0.94	0.93	0.93	0.92	0.92	0.92	0.92	0.92
35-39	0.88	0.92	0.93	0.95	0.95	0.95	0.94	0.94	0.94	0.94	0.94
40-44	0.86	0.91	0.94	0.95	0.95	0.95	0.95	0.95	0.94	0.94	0.94
45-49	0.84	0.90	0.93	0.94	0.95	0.95	0.95	0.95	0.95	0.95	0.94
50-54	0.80	0.87	0.91	0.93	0.95	0.95	0.95	0.96	0.95	0.95	0.95
55-59	0.74	0.83	0.88	0.92	0.93	0.95	0.95	0.96	0.96	0.95	0.95
60-64	0.66	0.78	0.84	0.89	0.92	0.93	0.95	0.95	0.96	0.96	0.95
65-69	0.57	0.70	0.78	0.86	0.89	0.92	0.93	0.95	0.95	0.96	0.96
70-74	0.45	0.60	0.69	0.74	0.81	0.85	0.87	0.89	0.90	0.91	0.91
75-79	0.32	0.48	0.57	0.62	0.67	0.73	0.76	0.78	0.80	0.81	0.82
80+	0.11	0.22	0.29	0.29	0.31	0.33	0.37	0.38	0.39	0.40	0.40

TØI rapport 606/2002

Tabell 5.9: Observerte (1980-1995) og beregnede (2000-2030) førerkortandeler blant kvinner i 5 års intervaller og aldersgrupper

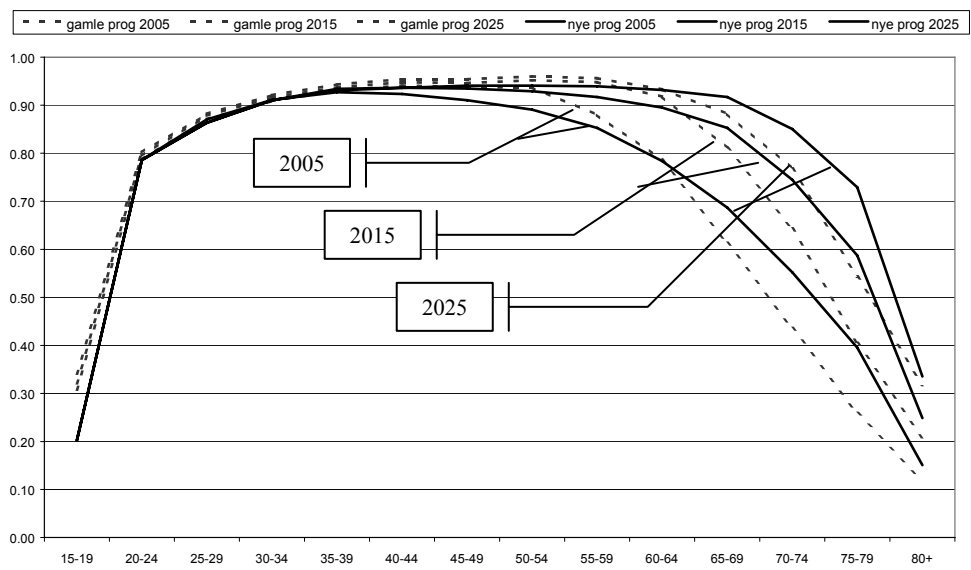
	1980	1985	1990	1995	2000	2005	2010	2015	2020	2025	2030
15-19	0.18	0.26	0.28	0.25	0.21	0.20	0.20	0.20	0.20	0.20	0.20
20-24	0.65	0.76	0.80	0.79	0.80	0.79	0.79	0.79	0.79	0.79	0.79
25-29	0.71	0.81	0.85	0.85	0.86	0.87	0.86	0.86	0.86	0.86	0.86
30-34	0.72	0.83	0.88	0.89	0.90	0.91	0.91	0.91	0.91	0.91	0.91
35-39	0.68	0.81	0.88	0.90	0.92	0.93	0.93	0.93	0.93	0.93	0.93
40-44	0.59	0.74	0.84	0.88	0.91	0.92	0.93	0.94	0.94	0.94	0.94
45-49	0.49	0.64	0.77	0.84	0.89	0.91	0.93	0.93	0.94	0.94	0.94
50-54	0.37	0.53	0.67	0.77	0.85	0.89	0.91	0.93	0.94	0.94	0.94
55-59	0.26	0.40	0.55	0.69	0.78	0.85	0.90	0.92	0.93	0.94	0.94
60-64	0.17	0.29	0.43	0.58	0.69	0.78	0.85	0.90	0.92	0.93	0.94
65-69	0.10	0.19	0.32	0.46	0.58	0.69	0.78	0.85	0.90	0.92	0.93
70-74	0.06	0.12	0.22	0.34	0.44	0.55	0.65	0.74	0.81	0.85	0.87
75-79	0.04	0.08	0.13	0.21	0.30	0.40	0.50	0.59	0.67	0.73	0.77
80+	0.00	0.00	0.00	0.06	0.11	0.15	0.20	0.25	0.29	0.34	0.36

TØI rapport 606/2002



TØI rapport 606/2002

Figur 5.11: Sammenligning av gamle og nye prognoser for menns innehav av førerkort



TØI rapport 606/2002

Figur 5.12: Sammenligning av gamle og nye prognoser for kvinners innehav av førerkort

# 6 Estimering av modeller for valg av transportmåte og destinasjon

## 6.1 Data

### 6.1.1 Behandling av reisetidskomponenter for kollektivtransport

I estimeringsarbeidet er en rekke formuleringer av variable for reisetidskomponenter testet. I dette avsnittet presenteres de formuleringer som har vist seg å passe best med datamaterialet. De reisetidskomponentene vi beregner i EMMA er:

- Kjøretid (ombordtid)
- Tilbringertid
- Total ventetid
- Første ventetid
- Antall påstigninger

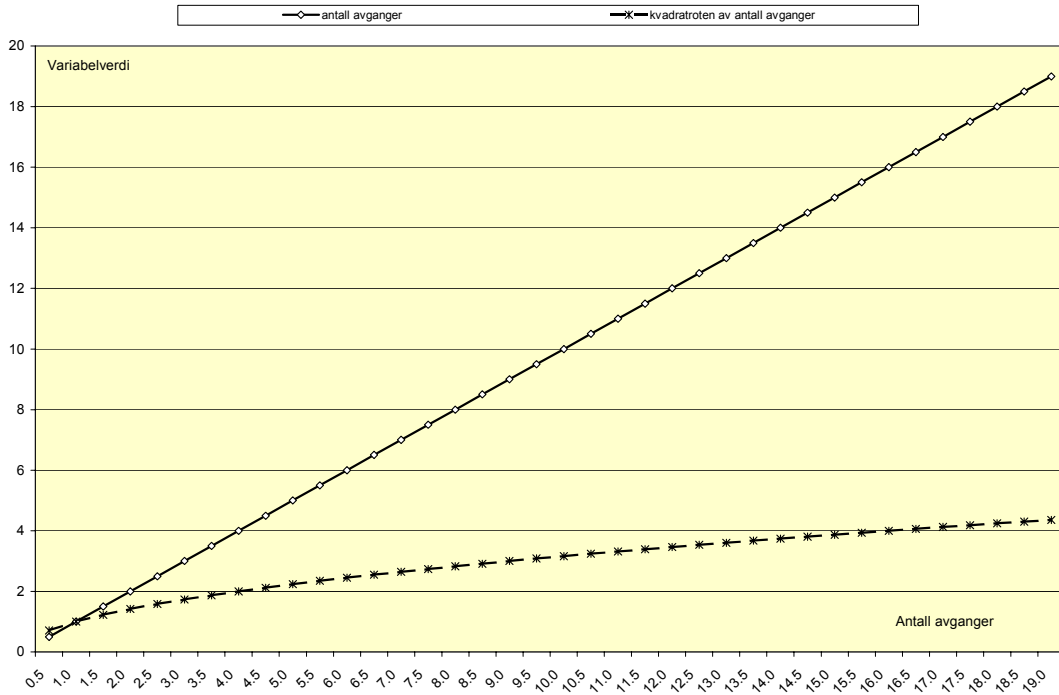
*Kjøretid* (minutter om bord transportmiddelet) inngår uten transformasjoner i alle modeller. Den *tilbringertid* som beregnes i nettverksmodellen forutsetter en gitt hastighet (j. fr tabell 3.8 over). Dette er en gjennomsnittlig tilbringerhastighet som benyttes i rutevalgsmodellen fordi den, sammen med de andre assignmentparametere, ser ut til å gi best rutevalg. Respondentene i datamaterialet har imidlertid reist på forskjellige måter til hovedtransportmiddelet og dermed med ulik hastighet. Hvis vi benytter den gjennomsnittlige tilbringertiden direkte kan vi dermed risikere å få store skjevheter i denne variabelen. Ved å konvertere tilbringertid til tilbringeravstand, unngår vi dette problemet. Det kan naturligvis være skjevheter også i denne variabel, men da mer knyttet til at valg av sted for påstigning er riktig. Variabelen for tilbringeravstand er benyttet for alle transportmåter og para-

meteren for denne variabel kan dermed sammenliknes mellom transportmåtene.

En rekke formuleringer av variable for *ventetid* eller *frekvens* er testet i estimeringsarbeidet. Det har vist seg vanskelig å skille mellom frekvens ved første påstigning og eventuelle påfølgende påstigninger underveis på reisen. Årsaken til dette er sannsynligvis knyttet til relativt få observasjoner for de kollektive transportmåtene. Vi har endt opp med en variabel-formulering som er kvadratrotten av antall avganger per døgn. Til å beregne antall avganger per døgn er total ventetid benyttet. Total ventetid er definert som tiden mellom avgangene multiplisert med en faktor som er 0.5 for båt, buss og tog, og 0.2 for fly. Faktoren angir hvor stor andel av tiden mellom avgangene som går med til venting. For de relasjoner som innebærer transportmiddelbytte underveis på reisen vil ventetidene ved hvert bytte summeres. På slike relasjoner vil totalt antall avganger være et generalisert mål på frekvensen.

Figur 6.1 viser sammenhengen mellom antall avganger og verdien på denne variabelen. Vi ser at det ved hyppige avganger er relativ stor forskjell mellom kvadratrotten av antall avganger og en lineær formulering. Ved bruk av kvadratrotten av antall avganger vil effekten av en frekvensøkning være relativt liten når frekvensen er høy fra før, og relativt stor ved lave frekvenser i utgangspunktet.

*Antall bytter* (=1 – antall påstigninger) inngår som variabel i alle modeller. I og med at ulempen ved å bytte transportmiddel underveis også inngår i variabelen for frekvens, antyder dette at det ligger en svært stor motstand mot å bytte transportmiddel underveis innbakt i reisevanene i datamaterialet.



TØI rapport 606/2002

Figur 6.1: Antall avganger og variabelverdi for benyttet frekvensvariabel

### 6.1.2 Reisekostnader

I beregningene av reisekostnader er det to forhold som er spesielt problematisk å håndtere. Det ene er behandling av størrelsen på reisefølget, og slik vi har formulert modellene knytter dette seg til bilreiser. Det er klart at hvis det er mange som skal ut å reise fremstår bilen ofte som den billigste transportmåten. Dette fordi kostnaden per person blir lavere enn om man skulle valgt andre transportmåter<sup>8</sup>. I RVU-97/98 er det bare spurt om størrelsen på reisefølget for respondenter som har reist med bil. Fordi denne endogene sammenhengen mellom størrelsen på reisefølget og transportmiddelvalget finnes, er det problematisk å benytte denne informasjon i beregningene av reisekostnadene. I RVU2001 er det spurt om størrelsen på reisefølget både for bilister og andre trafikanter. Vi har benyttet data herfra til å beregne gjennomsnittlig reisefølge etter reiseformål. Dette benyttes til å beregne reisekostnader per person for bilalternativet.

Det andre problematiske forholdet i beregningen av reisekostnader knytter seg til behandling av rabatter for kollektivreiser. De ulike kollektive transportmåtene har egne rabattregimer som i varierende grad er tilgjengelige for ulike grupper trafikanter. Også her er

det problemer knyttet til endogene sammenhenger, både i forhold til valg av transportmåte og til reise-frekvens. Alternativet for en bestemt fritidstrafikant kan for eksempel være å reise med fly med rabattert billett eller å la være å reise. De som har reist kollektivt i RVU-97/98 er spurt om billetttype for reisen. Denne type informasjon kan vi imidlertid ikke benytte fordi vi ikke vet hvilken billetttype de ville fått ved valg av en annen transportmåte. De som har reist med bil ville sannsynligvis også kunne reist med rabattert kollektivbillett, noe vi ikke har informasjon om.

Også i behandlingen av rabatter er vi derfor avhengig av å ta utgangspunkt i størrelser vi kjenner for alle trafikanter. Det dreier seg da i første rekke om informasjon om alder og beskjeftigelse.

Utgangspunktet for reisekostnadene beregnes i nettverksmodellene. For bilreiser tas det som nevnt ut to sett med matriser avhengig av om hvem som betaler for reisen (hvis arbeidsgiver betaler forutsettes bilisten kun å legge vekt på fremføringstiden i vegvalget, ellers benyttes generalisert kostnadsassignment i vegvalg-beregningene (jf. pkt. 3.3.2). I begge tilfeller beregnes reiseavstanden langs valgt reiserute (fordi det ikke er kapasitetsproblemer i vegnettet er denne entydig). I tillegg plukkes ferge- og bompengekostnader for bilfører og passasjerer opp langs vegvalget. Reisekostnaden beregnes da som

<sup>8</sup> Det finnes også ulike typer grupperrabatter for kollektive transportmåter, men slike har vi ikke kunnet ta hensyn til i estimeringene.

Kostnad bil = {kilometerkostnad + fergekostnad bil + bompenger bil + [(fergekostnad passasjer + bompenger passasjer)\*antall passasjerer]/antall personer i reisefølget}

Kilometerkostnaden er satt til kr 1.- per kilometer for private reiser og kr 2.40.- per kilometer for tjenestereiser. Det er gjennomført tester i estimeringen for å finne ut hvorvidt disse anslagene er realistiske. Hvis respondenten har en gunstig firmabilordning settes bilkostnaden til null for private reiser. Det samme gjelder dersom reisen er betalt av andre en trafikanten selv eller noen i husholdningen. Antall personer i reisefølget (etter formål) er beregnet ut i fra resultater for 1. tertial i RVU2001 (fordi vi bare har kjennskap til reisefølget for bilister i RVU97/98). Gjennomsnittsverdiene er vist i Tabell 6.1. Antall passasjerer er antall personer i reisefølget minus bilfører.

Tabell 6.1: Beregnet gjennomsnittlig reisefølge etter reiseformål, RVU2001 - resultater fra 1. tertial

Reiseformål	Beregnet gjennomsnittlig reisefølge
Arbeidsreise	1.4
Skolereise	1.4
Tjenestereise, kurs, konferanse, m.m.	2.0
Tjenestereise, salg, innkjøp, m.m.	1.7
Tjenestereise, service, konsulentbistand, m.m.	1.5
Tjenestereise, annet	1.6
Militærreise for vernepliktige	2.8
Innkjøp	2.7
Medisinske tjenester/ærend	2.1
Andre private ærend	1.9
Følge/hente andre personer	2.8
Fornøyelse, underholdning	2.8
Organisert fritidsaktivitet	2.9
Ferie og fritid	2.8
Besøk hos venner/familie	2.2
Kombinasjon tjeneste og private formål	1.9
Andre kombinasjoner	1.8
Annet	2.6
Vet ikke/ubesvart	2.3

TØI rapport 606/2002

For kollektivtransport benyttes de sammenhenger som er redegjort for i avsnitt 3.2.3 foran som utgangspunkt i beregningene av reisekostnader. For tjenestereiser benyttes disse anslagene direkte (dvs. ingen rabatter eller kostnadsavslag for tjenestereisene, men full kostnad for reisen). For private reiser er det laget et sett

med rabattindikatorer som avhenger av kjennetegn ved respondentene i RVU.

- Aldersrabatt på 50 %, omfatter trafikanter som er eldre enn 66 år, yngre enn 17 år, ufør eller annen pensjonist.
- Studierabatt på 50 %, omfatter trafikanter som er studenter eller avtjener verneplikt.
- Flyrabatt på 30 %, omfatter trafikanter som er borte lengre enn 2 netter, og ikke omfattes av de andre rabattordningene.
- Andre betaler, ingen kostnad hvis reisen er betalt av andre enn trafikanten selv eller noen i husholdningen (dvs. betalt av arbeidsgiver, oppdragsgiver, idretts- eller kulturorganisasjon, trygdekantor eller sosialkontor, eller andre organisasjoner eller foreninger).

I tillegg til billettprisene må kollektivtrafikantene betale eventuelle bompenger og fergekostnader på tilbringerreisen. Her forutsettes det at passasjer eller persontakster gjelder for reiser med tog, buss og båt, mens takster for bil gjelder for reiser med fly. Disse forutsetningene er gjort ut fra fordelingen på transportmidler for tilbringertransporten.

## 6.2 Modeller for valg av reisemiddel og destinasjon

Arbeidet med etablering av nye modeller for lange reiser i NTM5 er som kjent tredelt. Dette arbeidsdokumentet dokumenterer del B - selve estimeringsarbeidet.

Det er spesielt to forhold som skiller estimeringsarbeidet i NTM5 med tidligere estimeringsarbeid. Dette dreier seg om (1) etablering og anvendelse av opplegg for estimering med data for en finere soneinndeling enn kommunenivået, og (2) etablering og anvendelse av et opplegg for å ta høyde for en hypotese om at RVUen representerer en underrapportering av reiser i det laveste avstandsbåndet.

Avhengig av reiseformål (reisehensikt) er det definert fire delmodeller som utgangspunkt for valg av reisemiddel og destinasjon. Disse delmodellene er:

1. Tjenestereiser
2. Besøksreiser
3. Ferie- og fritidsreiser
4. Andre private reiser

I tabellene nedenfor og i vedleggene er disse referert til som hhv BUS (Business), VFR (Visit Friends), LEI (Leisure) og OTH (Other Private).

Arbeidsreiser som er betalt av arbeidsgiver regnes her som tjenestereiser, mens de resterende av disse faller inn under andre private reiser. Hvert av reiseformålene er aggregat av en finere inndeling i reiseformål fra RVUen, og dette er selvfølgelig den samme aggregeringen som benyttes i modellene for turgenerering (reisefrekvens). Ved estimeringen av hver delmodell benyttes bare de reisene som faller inn under delmodellens reiseformål.

Modellene for valg av transportmåte og reiseformål estimeres simultant og kobles deretter sammen med modellene for turgenerering gjennom logsummer som uttrykker ”summen” av nytten knyttet til transport- og destinasjonsvariable fra en gitt sone til alle destinasjoner.

Gjennom det meste av estimeringsarbeidet har det vært hensiktsmessig å estimere med data på kommune-kommune-nivå, i motsetning til NTPL-sonene (som innebærer omtrent 10 ganger så mange O/D-relasjoner og vesentlig lengre estimeringstid). Dessuten er det en modellteknisk utfordring at destinasjon bare er stedfestet til kommune i RVUen. Avsnitt 6.2.6 (med vedlegg) omhandler alternative grep som kan gjøres for å estimere på NTPL-data.

En tilsynelatende underrapportering av de korteste lange reisene (80-120 km) er bakgrunnen for å vurdere WESML-prosedyren som er beskrevet i avsnitt 6.2.5 (med vedlegg).

### 6.2.1 Modellformulering

Det finnes i utgangspunktet en mengde alternative modellformuleringer å vurdere. Kvaliteten til en gitt modell kan i første omgang analyseres ved å se på ulike statistiske egenskaper. Det vil si modellens evne til å gjenskape reisemønsteret i reisevaneundersøkelsen den er estimert på, og graden av forklaringskraft til ulike parametere.

For en gitt reisehensikt uttrykker logit-modellen de alternativene den som har reist har stått overfor - som funksjoner av ulike parametere som antas å ha betydning nettopp for valg av reisemiddel og destinasjon. Dette kan være ulike transportkvalitetsdata som reisetid, kostnad mv, samt attraktivitetsmål for de ulike destinasjonssonene, for eksempel befolkning, arbeidsplasser eller andre tilbud. Gitt modellteorien anvendes så en metode (*Maximum Likelihood*) for å beregne en vektning av disse parameterne som best gjenskaper valgene som er foretatt.

Kvaliteter ved transporttilbudet beregnes i EMMA, og bestemmes dermed av nettverket som er kodet her, samt algoritmene for beregning av rutevalg. Transportkvalitetsdata som beregnes her er reisetid, ventetid (frekvens), antall bytter, reiseavstand, billettpris, bom- og fergekostnader, tilbringertid til kollektivterminal, og antall ferger som må benyttes på reisen.

Noen av valgene vi må gjøre med tanke på formulering og estimering av modellene er knyttet til (1) valg av relevante parametere, (2) trestruktur for logit-modellen, (3) bruk av alternativspesifikke eller generiske koeffisienter for de ulike parameterne, (4) spesiell behandling av kritiske parametere som reisekostnad og frekvens, og (5) beregning/valg av gjennomsnittsverdier for bilkostnad, størrelse på reisefølge og rabatter.

For bilreise benyttes også biltilgjengelighet definert gjennom variablene ”førerkort” og ”biler pr person i husholdet”.

Når det gjelder kostnad har vi valgt å se på en segmentering i forhold til inntekt (personlig eller for husholdet) der et slikt skille viser seg å være statistisk signifikant. Segmentering gjøres ved å estimere kostnadskoeffisienter for to inntektsgrupper, en for ”lav” inntekt, og en for ”høy”.

Selve kostnadsvariabelen transformeres med funksjonen  $\ln(\text{kostnad})$ . Dette er et grep som gjøres for å behandle et estimeringsteknisk og problematisk faktum: den sterke korrelasjonen mellom reisetid og kostnad (spesielt for bilreiser). Vi skal også se på effekten av å ikke gjøre dette.

Frekvensvariabelen kan i utgangspunktet formuleres på to måter: som antall avganger pr døgn, eller som gjennomsnittlig tid mellom avganger. For den første formuleringen skal vi da forvente oss en positiv koeffisient, mens den andre formuleringen bør gi en negativ koeffisient. Det kan dessuten være hensiktsmessig å transformere avganger pr døgn med kvadratroten (se avsnitt 6.1.1). Begge formuleringene er utprøvd (sammen med andre variasjoner i modellformuleringen).

Det er gjort forsøk med innføring av to dummyvariable for bilreiser: ”firmabil” og ”årstid vinter”. For noen av modellene slår dette ut signifikant, men det vil være problematisk å ta med seg slike variable når modellene skal implementeres. Siden estimering med disse parameterne kan gi bedre modeller, er det et alternativ å estimere med dem, men ikke anvende dem videre (lås resten av koeffisientene og estimere nye konstanter i en avsluttende estimering). Avsnitt 6.2.4 (med vedlegg) gir en gjennomgang av noen ulike modeller som er estimert.



## 6.2.2 Transportkvalitetsvariable

Av transportkvalitetsvariable fra EMMA er følgende benyttet i estimeringen:

1. Reisetid (ombordtid)
2. Tilbringeravstand
3. Billettpris
4. Bom- og fergekostnader
5. "Frekvens" for kollektivtilbud
6. Antall bytter

I tillegg opereres det med kilometerkostnad for bil samt formålsavhengige gjennomsnittsverdier for reisefølge og rabatter (jf. pkt. 6.1.2).

## 6.2.3 Attraksjonsvariable

Sonedata brukes som size-variable for destinasjonsvalg. Variablene er benyttet som vist i Tabell 6.2.

## 6.2.4 Estimeringsresultater

Tabell 6.3 viser utvalgte delmodeller som kan benyttes som utgangspunkt for WESML og estimering på NTPL-nivå. Alle modellene bruker her en trestruktur med reisemiddel over destinasjon. Se beskrivelse av ulike parametere under.

Tabell 6.2: "Size-variable" benyttet i estimeringen av mode og destinasjonsmodeller

	Tjeneste	Fritid	Besøk	Andre
Befolkning			X	
Sysselsatte totalt				
Arbeidsplasser totalt				
Areal				
Hoteller				
shots. Hotellsenger	X	X		
shyt. Hytter/fritidshus		X	X	
se1. Arb pl: Jord-, skogbruk, fiske				
se2. Arb pl: Olje, berg	X			
se3. Arb pl: Industri, kraft, vannf, bygg, anl, samf	X			
se4. Arb pl: Varehandel mv				X
se5. Arb pl: Hotell og restaurant				
se6. Arb pl: Finans, forr, eiend, inter	X			
se7. Arb pl: Off adm, forsvar				X
se8. Arb pl: Undervisning				X
se9. Arb pl: Helse sos pers hus				

TØI rapport 606/2002

Tabell 6.3: Estimeringsresultater mode og destinasjon

File	BUSsx100.F12	LEIsx200.F12	VFRsx200.F12	OTHsx200.F12
Converged	TRUE	TRUE	TRUE	TRUE
Observations	1009	1500	1134	780
Final log (L)	-4136.4	-7529.4	-5733	-3395.2
D.O.F.	19	17	20	20
Rho <sup>2</sup> (0)	0.421	0.291	0.284	0.381
Rho <sup>2</sup> (c)	-4.561	-8.617	-5.396	-4.46
Prepared	15-Jan-02	22-Feb-02	22-Feb-02	22-Feb-02
Estimated	15-Jan-02	22-Feb-02	22-Feb-02	22-Feb-02
G_cost3	-0.320 (-2.0)		-0.0180 (-1.5)	-0.0433 (-3.4)
G_cost2	-0.973 (-6.0)		-0.0409 (-3.1)	-0.0878 (-6.5)
cavd5c	-0.267 (-1.2)	0.508 (2.2)	-0.233 (-1.1)	-0.0176 (-0.1)
cavd3c	-2.10 (-2.8)	-1.83 (-4.8)	-2.46 (-6.6)	-1.52 (-3.2)
cavd2c	-2.85 (-2.1)	-0.822 (-2.7)	-0.838 (-2.7)	-1.57 (-4.2)
cavd1c	-1.15 (-0.9)	-1.97 (-4.3)	-2.81 (-6.5)	-3.00 (-4.6)
G_trfer	-0.306 (-2.2)	-0.215 (-1.5)	-0.530 (-4.3)	-0.909 (-5.9)
G_freq	0.250 (5.4)	0.272 (3.9)	0.126 (2.2)	0.122 (1.6)
G_time	-0.0067 (-9.3)	-0.0010 (-2.4)	-0.0025 (-5.0)	-0.0012 (-3.0)
BB_accdis	-0.0567 (-3.9)			
G_accdis	-0.0234 (-11.0)	-0.0258 (-8.9)		-0.0147 (-4.0)
BU_Const	-5.07 (-7.3)	-4.40 (-7.3)	-3.69 (-7.0)	-3.19 (-7.5)
BO_Const	-3.25 (-4.0)	-3.75 (-5.6)	-2.61 (-4.5)	-1.88 (-4.2)
TR_Const	-3.39 (-8.6)	-3.86 (-8.1)	-2.59 (-7.6)	-3.25 (-8.5)
AI_Const	-0.520 (-1.4)	-2.99 (-5.8)	-2.60 (-5.6)	-2.15 (-4.9)
SizeMult	1.00 (*)	1.00 (*)	1.00 (*)	1.00 (*)
se6	-0.723 (-2.9)			
se3	-2.35 (-5.4)			
se2	1.17 (4.2)			
nc	0.598 (11.4)	0.862 (8.3)	0.820 (11.8)	0.846 (10.4)
G_cost		-0.0495 (-4.4)		
CA_time		-0.0062 (-19.9)	-0.0063 (-17.3)	-0.0079 (-18.2)
sbef		-3.76 (-13.0)		
shots		1.44 (14.7)		
BU_accdis			-0.0616 (-2.0)	
BO_accdis			-0.0341 (-3.9)	-0.0407 (-4.7)
TR_accdis			-0.0171 (-5.7)	
AI_accdis			-0.0229 (-8.4)	-0.0253 (-6.6)
shytt			0.647 (3.9)	
se8				2.27 (3.5)
se7				1.77 (2.8)

TØI rapport 606/2002

## Forklaring til tabell 6.3:

G	Generisk (felles for alle modes dersom ikke annet er angitt)
CA	Bil
BU	Buss
BO	Båt
TR	Tog
AI	Fly
BB	Felles for buss og båt
time	reisetid (in-vehicle time)
cost	reikostnad (billettpris/bensinpris + utgifter til bom/ferge)
freq	frekvens kollektivtilbud
trfer	antall bytter, kollektivt reisemiddel
accdis	tilbringerdistanse kollektivt (access/egress)
cost2	kostnadsparameter gjeldende for "lavere inntekt"
cost3	kostnadsparameter gjeldende for "høyere inntekt"
cavd5c	dummy for biltilgjengelighet kategori 5 (jf dok)
cavd3c	dummy for biltilgjengelighet kategori 3
cavd2c	dummy for biltilgjengelighet kategori 2
cavd1c	dummy for biltilgjengelighet kategori 1
se2	sonenhold/sizevariabel næringkategori 2: olje berg
se3	sonenhold/sizevariabel næringkategori 3: industri kraft vannf bygg anl samf
se4	sonenhold/sizevariabel næringkategori 4: varehandel mv
se6	sonenhold/sizevariabel næringkategori 6: finans forr eiend inter
se7	sonenhold/sizevariabel næringkategori 7: off adm forsvar
se8	sonenhold/sizevariabel næringkategori 8: undervisning
shytt	sonenhold/sizevariabel: hytter og fritidshus
sbef	sonenhold/sizevariabel: befolkning
shots	sonenhold/sizevariabel: hotellsenger
SizeMULT	sonenhold/sizevariabel: (the variable constrained to 1)
BUS:	shots
LEI:	shytt
VFR:	sbef
OTH:	se4
nc	logsumparameter (trestruktur mode over dest)

TØI rapport 606/2002

I tabellen angir "G\_" en generisk koeffisient (felles for alle, eller et knippe reisemidler), CA=Car, BU=buss, BO=båt, TR=tog og AI=fly. "BB\_" er en koeffisient som er felles for buss og båt. "nc" er logsumkoeffisienten, mens "se\_" angir ulike sizevariable<sup>9</sup> (se tabell 6.2).

Koeffisientene G\_cost2 og G\_cost3 representerer her henholdsvis det lave og høye inntektsintervallet. For tjenestereiser er skillet satt ved personlig bruttoinntekt på 300.000, for besøksreiser brukes brutto husholdsinntekt pr person og en grense ved 150.000,

mens andre private reiser segmenteres ved personlig bruttoinntekt på 250.000. For fritidsreiser har det vist seg at et slikt skille ikke er hensiktsmessig (signifikant).

Alle modellene over benytter en frekvensvariabel definert som kvadratroten av antall avganger pr døgn.

Dette er "utvalgte" modeller som er vurdert som de beste, men altså uten hensyn til WESML og estimering på NTPL-soner. Estimering som er gjennomført med disse metodene omtales i de to neste underkapitlene.

Ut over valgene som er representert i tabell 6.3 er det estimert med en rekke alternative modellformuleringer. Tabellene i vedlegg 8.3 dokumenterer noen

<sup>9</sup> Ved bruk av N size-variable estimeres det koeffisienter for (N-1) av disse, den første settes til 1 og er ikke vist i tabellen.

av disse modellene, med hensyn til ulik behandling av fire forhold.

1. Trestruktur. Multinomisk modell, eller strukturert modell med reisemiddel over destinasjon eller destinasjon over reisemiddel (henholdsvis  $T=0,1$  og  $2$ ).
2. Formulering av frekvensen til kollektivtilbudet. Avganger pr døgn eller tid mellom avganger ( $F=0,1$ ).
3. Dummyvariable for firmabil og årstid vinter, eller ikke ( $D=1,0$ ).
4. Kostnad transformert med logaritmen eller ikke ( $C=1,0$ ).

Styrefilene til Alogit er her nummerert ved hjelp av kodene T, F, D og C ovenfor. F eks vil estimering av en strukturert modell for fritidsreiser med  $\ln(\text{kost})$ , reisemiddel over destinasjon, kollektivfrekvens som kvadratrotten av tid mellom avganger, og uten dummyvariablene for bil ha koden LEIv1001 (filnavn LEIv1001.alo).

### 6.2.5 De korteste lange reisene: WESML

For å ta hensyn til underrapportering av de korteste lange reisene, kan det estimeres med en vektning av disse observasjonene. Et opplegg for dette er etablert og anvendt av RAND. Resultatene indikerer at dette har noe for seg, og sammen med opplegget for estimering på NTPL-nivå vil denne metoden anvendes i en siste estimeringsrunde.

Hensyn til noen andre betenkeligheter som nevnes i vedlegget er allerede ivaretatt i re-estimerte modeller for de fire reisemålene.

### 6.2.6 Estimering med NTPL-soner

Som nevnt er det problematisk at RVUen bare har stedfestet destinasjon til kommune. Det er derimot mulig å anvende metoder for å likevel estimere med data på NTPL-nivå. Alternative opplegg for dette er etablert og testet i samarbeid med RAND.

## 7 Modeller for valg av reisefrekvens

### 7.1 Alternative modellformuleringer

Det er tradisjon for å benytte logitmodeller i en eller annen variant for å beskrive valget av reisefrekvens i transportmodeller. Innenfor logit tradisjonen er det hovedsakelig tre alternative muligheter:

- Binomisk logit
- Multinomisk logit
- Strukturert logit

I binomisk logit formuleres valget mellom ”å reise” eller ”ikke å reise”. Dette gjøres på tross av at folk i rapporteringsperioden kan ha reist flere ganger. I pkt. 7.2 kommer vi tilbake til hvilke grep som må gjøres for å formulere en binomisk logitmodell. I multinomisk logit formuleres valget mellom ikke å reise, å reise én, to, tre, eller flere ganger avhengig av hvor mange ganger den observasjonen som har reist flest ganger, har reist. Denne formuleringen vil reprodusere fordelingen på antall reiser eksakt, men det kan være litt vanskelig å tilordne valgene variable fordi det for eksempel ikke er sikkert at det er så mye som skiller en observasjon som har reist tre ganger, fra en observasjon som har reist fire ganger. I strukturert logit konstruerer man en trestruktur der valget på det øverste nivå står mellom ikke å reise, eller å reise minst én gang. På det neste nivå står valget mellom ikke å reise flere ganger, eller å reise minst to ganger, osv. I visse situasjoner kan en slik formulering være hensiktsmessig.

Det finnes imidlertid andre modellvarianter enn logit. Her kan vi nevne Poisson og Geometriske modeller. Spesielt Poisson modeller gir muligheter til formuleringer av valgsituasjoner kanskje på en bedre måte enn i andre modeller. Ikke minst gjelder dette kombinasjoner av frekvens og destinasjonsvalgmodeller, og når det gjelder å estimere modeller på data med ulik rapporteringsperiode (for eksempel data fra en turdagbok i kombinasjon med antall lange reiser siste måned). Etter vår oppfatning er dette såpass lovende, ikke minst med tanke på fremtidig modellutvikling, at vi parallelt med estimeringen av logitmodeller har estimert ett sett

med poissonmodeller (logitmodeller og geometriske modeller med samme formulering for å sammenlikne modellene). På et senere tidspunkt i modellutviklingen må vi ta stilling til hvilken modellvariant man skal velge.

Som beskrevet i arbeidsdokument PT/1467/2001 vil en Poisson-modell være konsistent med logitmodeller og logitformuleringer kan utledes fra Poisson-modeller. Det som skiller er at Poisson-modeller i utgangspunktet tillater at et alternativ velges flere ganger innenfor den periode modellen gjelder for.

Poisson-fordelingen har følgende form:

$$P(x = n) = e^{-\lambda} \frac{\lambda^n}{n!}$$

Når vi skal estimere innfører vi transformasjon:

$$\lambda = e^{U(\beta, Z)}$$

U er da den ”nyttefunksjon” med parametre  $\beta$  og tilhørende variable  $Z$ .

Strukturen i en Poisson-modell er imidlertid noe restriktiv fordi det er en stram sammenheng mellom sannsynlighetene for antall ganger en begivenhet opptrer.

Vi vil i en Poisson-modell nemlig ha følgende rekursive sammenheng:

$$P(x = n) = P(x = n - 1) \cdot \frac{\lambda}{n}$$

Slik sett er man mer fleksibel med en logitformulering, men fleksibiliteten avhenger noe av hvordan modellen struktureres og den økte fleksibilitet går selvsagt noe på bekostning av økt kompleksitet. Har vi en binomisk modell basert på en logitformulering av sannsynlighetene vil den også ha tilsvarende stram sammenheng. Hvis det i prinsippet og i praksis er slik at vi kan observere mer enn én reise av en bestemt type innenfor det tidsintervall vi betrakter, så vil vi i en binomisk modell ha:

$$P(x = n) = P(x = n - 1) \cdot \frac{p}{1 - p} \cdot \frac{N - n}{n}$$

En annen fordeling som også er benyttet i forbindelse med turfrekvenser er geometriske fordeling. Sannsynligheten for å foreta n reiser blir med denne fordeling:

$$P(x = n) = p^n \cdot (1 - p)$$

En ulempe ved denne fordeling som gir sannsynligheten for at man har n ”suksesser” før det man har et forsøk som feiler (eller omvendt), er at den ikke har noen tidsdimensjon. Binomisk fordeling vil eksplisitt referere seg til N uavhengige ”forsøk” og enhver Poisson-fordeling må referere seg til antall ganger en begivenhet opptrer i en gitt tidsperiode. Dette vil innebære at dersom p ligger nær én så har man ingen garanti for at forventningsverdien, ”p/(1-p)”, vil gi et realistisk resultat i forhold til den periode vi har observasjoner for.

En generalisering av den geometriske fordeling gir negativ binominal fordeling. En realistisk anvendelse av denne ville kreve at vi hadde en sekvens av dager og at man for hver dag hadde observert om det ble foretatt en lang reise eller ikke.

N må her betraktes som det maksimale antall reiser av denne type som kan foretas og p er sannsynligheten for at begivenheten ”én reise” skal forekomme en bestemt ”dag”. Skal en logitmodell være mer fleksibel må den ha form av en multinomisk modell. En fordel ved Poisson kontra binomisk er at man ikke behøver fiksure N (antall ”forsøk”) på forhånd.

Uansett er det verdt en forsøk å se på alternativer til mer tradisjonelle formuleringer. I det følgende presenteres derfor et sett med frekvensmodeller basert på Poisson-fordeling. Disse er bevisst holdt meget enkle fordi en fullstendig modell også bør inkludere en logsum som forklaringsvariabel og her vil svært mange variable bli tatt hensyn til på en indirekte måte.

Vi har estimert både poissonmodeller, geometriske modeller og logitmodeller for å beskrive valget av reisefrekvens. I alle modelltyper finner vi at stort sett de samme variable blir viktige som forklaringsfaktorer for valget av reisefrekvens. At vi til sist endte opp med å implementere logitmodeller i koden skyldes at vi for denne modelltypen (estimert i alogit 3) hadde mulighet for å

studere hvordan modellene reproduserte data-materialet. Slike prosedyrer kunne for så vidt også vært laget for de andre modelltypene som er estimert, men det ble ikke tid til dette innenfor prosjektet.

I avsnittene under dokumenteres de logitmodeller som er implementert. De endelige modellene er et resultat av en lang estimeringsprosess med prøving og feiling, testing og vurdering. Nær 300 ulike modeller er estimert i denne prosessen.

## 7.2 Valg av reisefrekvens: binomiske logitmodeller

I estimeringen av binomiske logitmodeller for valg av reisefrekvens er det benyttet en teknikk som er parallell til estimering på aggregerte data. Bakgrunnen for dette er at noen respondenter har gjennomført flere reiser i løpet av rapporteringsperioden, også flere reiser for hvert formål. Binomiske modeller kan dermed ikke benyttes direkte. Man kunne tenkt seg en multinomisk modellstruktur, der valget for eksempel står mellom ingen reiser, én reise, to reiser, osv. opp til n reiser, hvor n er antall reiser for den observasjon som har reist flest ganger. Når vi har valgt bort en slik struktur er det fordi man ikke har noen særlig relevante data som varierer mellom alternativene. Dessuten er det svært få observasjoner som har reist flere ganger, når man splitter dataene opp på reiseformål.

Tabell 7.1: Antall respondenter etter formål og antall reiser.

	All purposes	Business/work	Visit	Leisure	Other private
Ingen	5290	8132	7687	7378	7894
én reise	2489	427	916	1124	692
to reiser	675	106	110	181	99
tre reiser	181	41	16	44	26
fire reiser	69	18	5	6	16
fem reiser	20	7	1	2	6
Seks reiser	7	2	0	0	1
sju reiser	3	2	0	0	1
ti reiser	1	0	0	0	0
Sum	8735	8735	8735	8735	8735
Antall reist	3445	603	1048	1357	841
% reist	39	7	12	16	10

TØI rapport 606/2002

Tabell 7.2: Endelige estimeringsresultater for binomiske logitmodeller for valg av reisefrekvens

	Tjenestereiser	Besøkreiser	Fritidsreiser	Andre private reiser
	w8_386*	v8_141	L8_251	O8_127
Antall iterasjoner	10	9	8	9
Antall observasjoner	7870	8686	8686	8686
Likelihood med bare konstanter	-5894	-7649	-9982	-6959
Likelihood uten konstanter	-163652	-180620	-180620	-180620
Endelig Likelihood	-5544	-7571	-9716	-6802
"Rho-Squared" w.r.t. Zero	0.966	0.958	0.946	0.962
"Rho-Squared" w.r.t. Constants	0.059	0.010	0.027	0.023
Koeffisientestimer	Estimat T Ratio	Estimat T Ratio	Estimat T Ratio	Estimat T Ratio
Konstant	-10.8700 -22.3	-8.4680 -13.1	-10.3300 -21.0	-8.4150 -11.9
Mann		-0.3133 -5.3		0.3259 5.0
(malder/18)	5.1390 13.9			
(malder/18)^2	-1.0570 -13.4			
(kalder/18)	4.3880 10.8			
(kalder/18)^2	-0.9506 -9.8			
(alder/18)				0.7193 3.1
(alder/18)^2				-0.1786 -3.9
Alder20-29		0.3970 5.5		
Alder 20-49			0.1511 2.7	
Alder over 70		-0.3091 -2.5		
Enslig	-0.3053 -3.1			
Arbeider			-0.2035 -3.1	
Par uten barn			0.1108 1.9	
Vinter/sommer		0.2012 3.4		
Vinter			-0.3944 -4.4	0.1421 2.1
Høst			0.2819 3.8	
Sommer			0.7600 11.2	
(Pinnt/sqrt(Ginnt))	0.0134 14.2			
Ln(Pinnt)				0.0693 2.0
Ln(max(100,Hinnt)/personer i HH)		0.2603 4.7		
Ln(max(100,Hinnt))			0.3320 6.2	0.1683 2.5
Pinnt<200			-0.1154 -2.0	
Hinnt 300-500			-0.1198 -2.3	
Usentral1 (verdi=1)		0.2335 2.3		
Usentral2 (verdi=1 og 2)				0.6086 7.1
Nordland, Troms, Finnmark	0.6231 5.3			0.3364 2.2
Trøndelag, Nordland, Troms			0.3740 4.1	
Trøndelag, Nordland, Troms, Finnmark		0.6080 6.1		
Oslo kommune		0.3021 3.4	0.2713 3.9	
Akershus, Hedmark, Oppland		0.1608 1.9	0.1913 2.9	
Rogaland, Hordaland, Møre og Romsdal				-0.1702 -2.0
Akershus				-0.2912 -2.7
TP10 byer ekskl. Oslo		-0.2232 -2.4		
(Arbeidsplasser/bosatte) innenfor 100 km	-1.4730 -3.2			-2.1550 -4.5
Logsum fra mode/destinasjon modeller	0.1304 3.6	0.1555 2.6		0.1939 3.2

\*bare alder fra 18 til og med 79

TØI rapport 606/2002

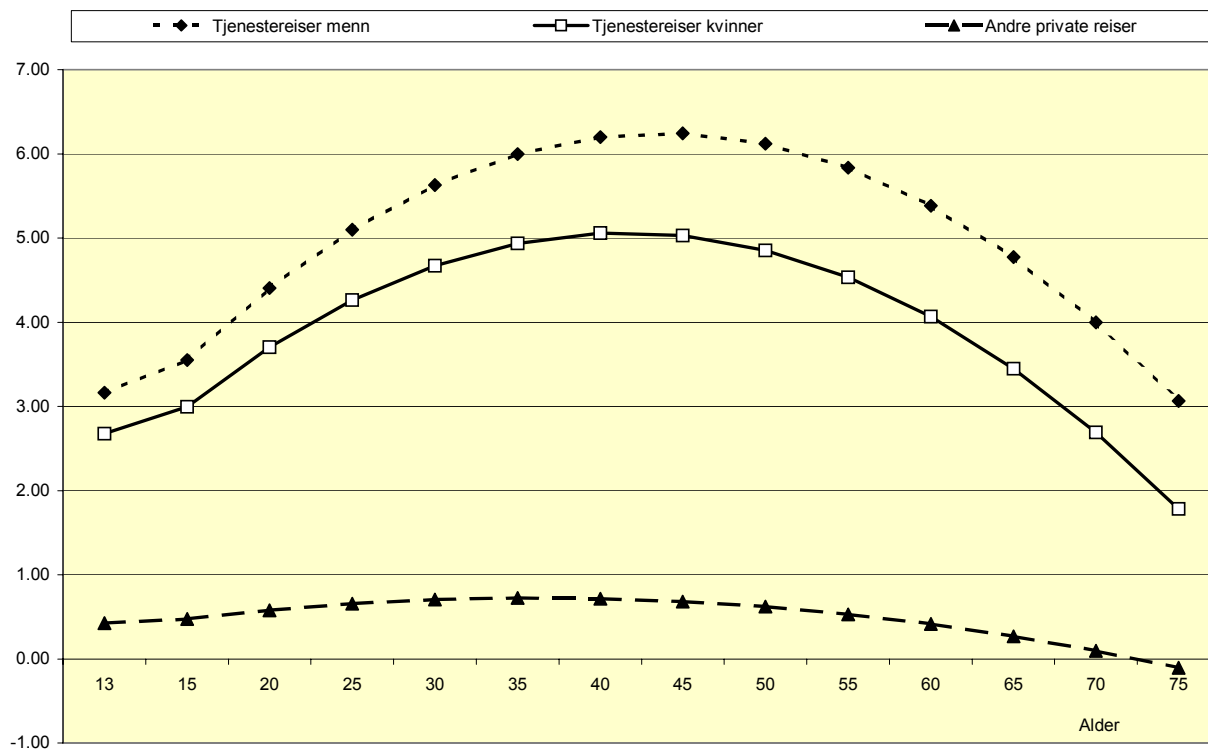
Estimering av binomiske modeller krever imidlertid et estimeringsteknisk triks. Vi tilordner observasjonene alternativet ”reist” og ”ikke reist” i den proporsjon de faktisk har tilkjennegitt i RVU-en. Rapporteringsperioden i RVU er en måned. Hvis en person for eksempel har gjennomført 3 reiser med ett bestemt reisemål i løpet av rapporteringsperioden, vektet alternativet ”reist” i estimeringen med 3/30, og alternativet ”ikke reist” får 27/30 i vekt. En person som ikke har reist i det hele tatt får 0/30 i vekt for ”reist” og 30/30 for ”ikke reist”. Dette er en normal fremgangsmåte i estimering av langdistanse frekvensmodeller av denne type.

Endelige estimeringsresultater er vist i Tabell 7.1. Koeffisientene i tabellen er inndelt i 6 bolker. I den første rammen vinner vi konstantleddene. Som vi ser er det felles for alle modeller at konstantleddene er høye i tallverdi. Dette har å gjøre med at selv om rapporteringsperioden er lang, gjennomføres lange reiser svært sjelden. Høye konstantledd reflekterer dermed at den estimerte motstanden mot å reise blir relativt høy. Den neste bolken koeffisienter har å gjøre med kjønn og alder. Deretter kommer noen segmenteringsvariable, inntekts-

variable, geografiske dummies og til sist en variabel som reflekterer forholdet mellom arbeidsplasser og antall bosatte innenfor 100 km fra bostedet. Den nederste koeffisient i tabellen representerer en logsumvariabel fra mode og destinasjonsmodellene. Koeffisientene vil bli kommentert i avsnittene under.

### 7.2.1 Modeller for arbeids/tjenestereiser

Slik modellen for arbeids- og tjenestereiser er formulert vil sannsynligheten for å foreta en slik reise øke med alder frem til et visst punkt for så å avta etter dette. Sammenhengen mellom alder og sannsynlighet for å reise er forskjellig for menn og kvinner. Dette fremgår av Figur 7.1. Figuren viser alderens betydning i nyttefunksjonen for å reise. Vi ser at menn har ett toppunkt rundt 45 år mens kvinner har en topp rundt 40 år. At kurven for menn ligger høyere enn kurven for kvinner indikerer at menn har en høyere tilbøyelighet til å gjennomføre lange arbeid/tjenestereiser enn kvinner.



TØI rapport 606/2002

Figur 7.1: Effekt av formulering av aldersvariable i modeller for tjenestereiser og andre private reiser



Videre har modellen en dummyvariabel for enslige. Hvis en person er enslig er sannsynligheten for å foreta en lang arbeids/tjenestereise lavere enn hvis man lever i et parforhold.

Inntektsvariabelen er formulert som personlig inntekt dividert med kvadratroten av gjennomsnittsinntekt (som er kr 209 000 i datamaterialet). Denne formuleringen sørger for at inntektselastisiteten avtar med økende inntekt. I tidligere versjoner av NTM har inntektseffektene vært størst for tjenestereisene. I et tverrsnittsmateriale er det kanskje slik at høy inntekt øker sannsynligheten for å foreta tjenestereiser. Sannsynligvis fanger imidlertid inntektsvariabelen en rekke effekter knyttet til stilling, næring, m.m.. Når inntekten endres over tid kan det derfor være misvisende å ha en stor inntektseffekt i modeller for arbeids og tjenestereiser. Veksten i antall lange tjenestereiser dras hovedsakelig av andre forhold enn økt inntekt isolert.

I modellen er det en geografisk dummy som gjør at sannsynligheten for å foreta lange tjenestereiser øker dersom man er bosatt i de tre nordligste fylkene. Det er konstruert en variabel som angir antall arbeidsplasser dividert med antall bosatte innenfor 100 km fra bostedssonen. Denne variabel gir en kraftig negativ effekt på sannsynligheten for å foreta lange tjenestereiser. Variabelen reduserer altså sannsynligheten for å reise hvis det er mange arbeidsplasser per bosatt i nærområdet. Tanken er da at sannsynligheten for å foreta korte tjenestereiser er større. Til sist gir logsummen fra mode/destinasjonsmodellene en liten effekt på sannsynligheten for å foreta lange tjenestereiser. Transporttilbud og tilgjengelige destinasjoner (lenger unna enn 100 km) attraktivitet spiller dermed en viss rolle i genereringen av tjenestereiser.

### 7.2.2 Modeller for besøksreiser

Ifølge modellen har menn isolert sett en lavere sannsynlighet for å foreta lange besøkreiser enn kvinner. Aldersgruppen fra 20 til 29 år ser ut til å være mest aktiv når det gjelder dette reiseformålet, mens de eldste blant oss (over 70) er mindre aktive. Besøkreiser ser ut til å ha en topp på sommeren og om vinteren. Inntektsvariabelen i modellen er formulert som logaritmen til husholdningens samlede inntekt dividert med antall personer i husholdet. Inntektselastisiteten kan dermed leses direkte ut av tabellen, 0,26, men vil avhenge av forholdet mellom antall personer i husholdet med inntekt arbeid og totalt antall personer i husholdet.

Det er en del geografiske dummyvariable i modellen, og alle har etter det vi kan se korrekt fortegn og er rimelig i forhold til hverandre. Hvis man bor i svært usentrale strøk øker sannsynligheten for å foreta en lang besøkreise. Er man bosatt i Trøndelagsfylkene eller fylkene nordover, øker også sannsynligheten for å foreta lange besøkreiser. Bosatte i Oslo kommune foretar også flere lange besøkreiser, noe som bl.a. kan knyttes til store innslag med innflyttere fra andre deler av Norge som reiser på besøk hjem til venner og familie. Bosatte i TP-10 byene utenom Oslo foretar færre lange besøkreiser enn andre.

Også i modellen for besøkreiser spiller logsummen fra mode/destinasjonsmodellene inn med begrenset betydning.

### 7.2.3 Modeller for fritidsreiser

Slik modellen er estimert ser aldersgruppen fra 20-49 år ut til å være de mest aktive når det gjelder å gjennomføre lange fritidsreiser. Vi finner ingen forskjeller mellom kjønnene for dette reiseformålet. Arbeidere (ufaglærte og faglærte) gjennomfører imidlertid slike reiser i mindre grad enn andre, mens par uten barn har en mer aktiv adferd. Lange fritidsreiser gjennomføres i mindre grad om vinteren, men helst om høsten, og naturligvis i størst grad om sommeren.

Det er tre inntektsvariable med i modellen, én kontinuerlig, formulert som logaritmen av husholdsinntekt, og to dummyvariable, som demper reiseaktiviteten hvis personlig inntekt er lavere enn kr 200.000.- per år eller husholdinntekten ligger mellom kr 300.000 og 500.000 per år. Det er tre geografiske dummyvariable som alle øker reiseaktiviteten avhengig av hvor man er bosatt. Logsumvariabelen ble ikke signifikant i denne modellen og er derfor ikke tatt med.

### 7.2.4 Modeller for andre private reiser

Andre private reiser (arbeidsreiser betalt selv, skolereiser, innkjøpsreiser, medisinske reiser, militærreiser, følge/hente reiser, andre private reiser) gjennomføres i litt større grad av menn enn av kvinner. Det er en svak sammenheng mellom reiseaktivitet med dette formålet og alder. Sammenhengen fremgår av Figur 7.1, og vi ser at det er et svakt toppunkt rundt 35 år. Det er en svak tendens til at slike reiser gjennomføres om vinteren.

Både husholdningens og personlig inntekt er med som forklaringsvariable og formulert som

logaritmen av inntekt. Elastisiteten fremgår dermed av tabellen som summen av de to koeffisienter.

Der er fire geografiske dummyvariable med i modellen. Den første øker sannsynligheten for å gjennomføre lange private reiser hvis man er bosatt i usentrale områder. Er man bosatt i Nord-Norge økes sannsynligheten ytterligere, mens den dempes hvis man er bosatt på vestlandet eller i Akershus. I denne modellen ble også variabelen som angir ”aktiviteten” innenfor 100 km fra bostedet tatt med, og den blir som vi ser svært betydningsfull som en faktor som demper denne type reiser. Logsummen fra mode/destinasjonsmodellen viser at transporttilbud og tilgjengelighet til attraktive destinasjoner lenger unna enn 10 mil også spiller en rolle for denne type reiser.

### 7.3 Skalering og segmenttabeller

En del av de koeffisienter som er vist i tabellen representerer variable som vi ikke vil ha tilgjengelig når modellene skal implementeres og benyttes på sonedata. I en siste estimeringsrunde må disse variablene fjernes fra modellene. Hvis vi kaller disse koeffisienter og variablene for b-koeffisienter og y-variable, og de øvrige for a-koeffisienter og x-variable kan nyttefunksjonen for alternativet ”reist” skrives:

$$1. U(\text{reist}) = k + \sum_i a_i x_i + \sum_j b_j y_j$$

I den endelige estimeringen vil man da estimere modeller av typen

$$2. U(\text{reist}) = K^* + A(\sum_i \alpha_i x_i + \sum_j \beta_j y_j)$$

Her er konstantleddet k i 1. erstattet med et nytt konstantledd  $K^*$ . I stedet for å benytte y-variablene benyttes såkalte segmentverdier for disse, og koeffisientestimatene låses til  $\beta_i$ , som er verdiene estimert i 1. Koeffisientestimatene for x variablene fra 1. er låst til  $\alpha_i$ , og det estimeres en skala-koeffisient A som vil ivareta effektene av den aggregeringen som i prinsippet skjer.

I tabellene under fremgår det hvilke variable hvor vi må benytte såkalte segmentverdier (y-variable). I prinsippet er dette alle variable som ikke er geografiske dummyvariable i tillegg til de to siste variablene i Tabell 7.2. Når det gjelder inntekt har vi en sonevariabel som representerer gjennomsnittlig inntekt for personer over 17 år i sonene. Denne variabel benyttes som en indeks for å gi geografisk variasjon i inntektene.

Segmentvariablene beregnes ut i fra data-materialet i RVU. Med unntak for inntekt vil de beholdes som dette for alle soner i modellen. Slik sett kan man se på denne type variable, sammen med den estimerte parameter, som konstantledd som varierer mellom segmenter men ikke mellom sonene utover det som finnes av variasjon i befolkningsstrukturen mellom sonene.

Tabell 7.3: Segmentvariable i modell for arbeids/tjenestereiser (w8\_386)

Kjønn	Alder	Antall	alder/18	(alder/18)^2	enslig	Pinnt/(sqrt(209))
M	-17	319	0.84	0.71	0.00	2.25
K	-17	308	0.84	0.71	0.01	1.96
M	18-19	101	1.02	1.05	0.02	3.49
K	18-19	107	1.03	1.05	0.04	2.12
M	20-24	231	1.23	1.53	0.16	8.41
K	20-24	250	1.23	1.52	0.16	5.32
M	25-34	850	1.65	2.76	0.26	17.36
K	25-34	991	1.66	2.78	0.15	11.50
M	35-49	1276	2.31	5.38	0.17	21.29
K	35-49	1216	2.31	5.41	0.10	12.92
M	50-64	912	3.11	9.76	0.18	20.58
K	50-64	955	3.10	9.65	0.28	12.21
M	65+	526	4.05	16.51	0.32	11.20
K	65+	693	4.11	17.07	0.59	6.28
M	alle	4215				
K	alle	4520				
alle	alle	8735				

TØI rapport 606/2002

Tabell 7.4: Segmentvariable i modell for besøkreiser (v8\_141)

Kjønn	Alder	Antall	alder 20-29	alder>70	vinter & sommer	ln(Hinnt/P)
M	-17	319	0.00	0.00	0.55	1.48
K	-17	308	0.00	0.00	0.50	1.47
M	18-19	101	0.00	0.00	0.56	1.53
K	18-19	107	0.00	0.00	0.40	1.51
M	20-24	231	1.00	0.00	0.54	1.57
K	20-24	250	1.00	0.00	0.52	1.55
M	25-34	850	0.45	0.00	0.49	1.61
K	25-34	991	0.45	0.00	0.51	1.57
M	35-49	1276	0.00	0.00	0.51	1.61
K	35-49	1216	0.00	0.00	0.53	1.58
M	50-64	912	0.00	0.00	0.49	1.66
K	50-64	955	0.00	0.00	0.48	1.64
M	65+	526	0.00	0.67	0.50	1.58
K	65+	693	0.00	0.71	0.51	1.55
M	alle	4215				
K	alle	4520				
alle	alle	8735				

TØI rapport 606/2002

Tabell 7.5: Segmentvariable i modell for fritidsreiser (18\_251)

Kjønn	Alder	Antall	arbeider	alder 20-49	par-u/barn	vinter	host	sommer	ln(Hinnt)	pint<2'	Hint3'-5'
M	-17	319	0.20	0.00	0.01	0.27	0.22	0.28	5.77	0.99	0.49
K	-17	308	0.23	0.00	0.00	0.23	0.22	0.26	5.72	1.00	0.50
M	18-19	101	0.49	0.00	0.01	0.22	0.19	0.35	5.84	0.98	0.36
K	18-19	107	0.30	0.00	0.08	0.16	0.31	0.24	5.75	0.98	0.46
M	20-24	231	0.52	1.00	0.14	0.29	0.23	0.25	5.74	0.77	0.26
K	20-24	250	0.35	1.00	0.20	0.24	0.25	0.28	5.56	0.95	0.38
M	25-34	850	0.37	1.00	0.20	0.27	0.27	0.22	5.83	0.21	0.46
K	25-34	991	0.20	1.00	0.17	0.27	0.24	0.24	5.79	0.59	0.43
M	35-49	1277	0.30	1.00	0.11	0.26	0.25	0.26	6.03	0.11	0.43
K	35-49	1216	0.19	1.00	0.15	0.24	0.24	0.29	5.94	0.49	0.45
M	50-64	912	0.22	0.00	0.55	0.23	0.27	0.26	5.96	0.18	0.43
K	50-64	955	0.14	0.00	0.48	0.24	0.25	0.24	5.73	0.55	0.36
M	65+	526	0.03	0.00	0.61	0.24	0.25	0.26	5.36	0.68	0.20
K	65+	693	0.02	0.00	0.35	0.22	0.22	0.28	5.02	0.91	0.06
M	alle	4215									
K	alle	4520									
alle	alle	8735									

TØI rapport 606/2002

Tabell 7.6: Segmentvariable i modell for andre private reiser (o8\_127)

Kjønn	Alder	Antall	(alder/18)	(alder/18)^2	vinter	ln(Pinnt)	ln(Hinnt)
M	-17	319	0.84	0.71	0.27	1.37	5.77
K	-17	308	0.84	0.71	0.23	1.10	5.72
M	18-19	101	1.02	1.05	0.22	2.94	5.84
K	18-19	107	1.03	1.05	0.16	2.68	5.75
M	20-24	231	1.23	1.53	0.29	4.38	5.74
K	20-24	250	1.23	1.52	0.24	3.94	5.56
M	25-34	850	1.65	2.76	0.27	5.36	5.83
K	25-34	991	1.66	2.78	0.27	4.80	5.79
M	35-49	1276	2.31	5.38	0.26	5.62	6.03
K	35-49	1216	2.31	5.41	0.24	4.96	5.94
M	50-64	912	3.11	9.76	0.23	5.57	5.96
K	50-64	955	3.10	9.65	0.24	4.90	5.73
M	65+	526	4.05	16.51	0.24	4.95	5.36
K	65+	693	4.11	17.07	0.22	4.41	5.02
M	alle	4215					
K	alle	4520					
alle	alle	8735					

TØI rapport 606/2002

Tabell 7.7: Estimerte konstantledd og skalaparametre i frekvensmodellene

	Tjenestereiser W8_386		Besøkreiser v8_141		Fritidsreiser L8_251		Andre private reiser O8_127	
	Estimate	T Ratio	Estimate	T Ratio	Estimate	T Ratio	Estimate	T Ratio
Nytt konstantledd	-10.330	-42.200	-7.567	-35.900	-11.120	-23.400	-8.295	-46.700
Skalaparameter	0.913	20.900	0.993	10.700	1.177	12.900	0.965	16.800

TØI rapport 606/2002

Tabell 7.7 viser estimatene på K\* (nytt konstantledd) og A (skalaparameter) når vi legger inn aggregerte segmentvariable isteden for de respondentspesifikke variablene og samtidig låser alle parameterverdier slik de er vist i Tabell 7.2

Konstantleddene blir litt større i tallverdi, mens skalaparametrene ligger svært nær 1. Dette innebærer at aggregeringen i realiteten har liten betydning for modellene.

## Vedlegg til kapittel 7

### Valg av reisefrekvens: poissonmodeller

Hvis vi selekterer bort personer med uoppgitt alder og som mangler oppgitt eller beregnet inntekt for IO og/eller husstanden har vi 8690 observasjoner.

For de fire reiseformål:

- Tjeneste- og arbeidsreiser
- Ferie- og fritidsreiser
- Besøksreiser
- Andre reiser

er det i RVU da registrert følgende antall reiser (tur/retur):

Tabell V7.1: Lange reiser etter formål

Reiseformål	Antall reiser
Ferie- og fritid	1646
Tjeneste- og arbeidsreiser	900
Besøksreiser	1196
Andre reiser	1074
Sum	4815

TØI rapport 606/2002

Et spørsmål man kan stille er om det for IO er noen sammenheng mellom reiser med ulike formål. Hvis det f.eks. er slik at personer som har mange reiser med et bestemt reiseformål da har få reiser med et annet reiseformål, så bør egentlig frekvensmodeller estimeres simultant slik at man får tatt hensyn til slik korrelasjon.

En enkel korrelasjonstabell tyder på at det ikke er påfallende sterk sammenheng mellom antall reiser med ulike formål. Det er en meget svak negativ korrelasjon mellom antall reiser med ulike formål bortsett fra mellom 2 og 3 hvor korrelasjonen er meget svakt positiv og praktisk talt null.

Med så vidt lav korrelasjon skulle det ikke være grunn til å regne med at det er en gevinst av betydning forbundet med å estimere på alle reiseformål simultant. Ved å generere reiser uavhengig for hvert reiseformål vil den (overveiende) negative korrelasjon imidlertid medføre at vi får en svak overprediksjon av totalt antall reiser fordi vi ikke

tar hensyn til den korrelasjon som tross alt finnes på individnivå.

Tabell V7.2: Korrelasjonskoeffisienter, antall reiser med ulike formål

	1	2	3	4
Ferie- og fritid	1	-0,0215	-0,0512	-0,0593
Tjeneste- og arb.reiser	-0,0215	1	0,0005	-0,0218
Besøksreiser	-0,0512	0,0005	1	-0,0375
Andre reiser	-0,0593	-0,0218	-0,0375	1

TØI rapport 606/2002

### Fritidsreiser

I den modell som er vist i Tabell V7.3 har vi inkludert kjønn og alder på IO, husholdningens inntekt og andel av befolkningen i bostedskommunen som bor i tettbygde strøk

Parametrene inngår lineært i en ”nyttefunksjon” (U) og parameteren i Poisson-fordelingen er  $\exp(U+\ln(30))$  når det dreier seg om en 30-dagers periode og  $\exp(U)$  når det er tale om ÅDT.

Som det fremgår av Tabell V7.3, blir alle parametrene signifikante, modellen reproducerer antall reiser i datamaterialet som summen av forventningsverdiene  $\exp(U+\ln(30))$ . Summerer vi de beregnede sannsynligheter for 0,1,2,3 og 4+ reiser for enkeltobservasjoner ser vi at fordelingen også er ganske brukbar, men med en underprediksjon for 2 eller flere reiser og for 0 reiser. Strukturen på en Poisson-fordeling er slik (jf. den rekursive sammenheng ovenfor) at det er umulig å gjenskape en fordeling som i utgangspunktet ikke er Poisson og dette gjelder her for fordelingen av observert antall reiser.

Vi har i materialet kanskje 8 personer som har foretatt en lang weekend reise hver helg i løpet av de siste 30 dager. Siden det med de variable vi har, er nesten umulig å skille ut disse personer vil modellen ikke bli særlig treffsikker mhp. å plukke ut disse personer med meget høy reisefrekvens.

Tabell V7.3: Frekvensmodell for fritidsreiser

Parameters	Vnames	Estimates	Std. err.	Est./s.e.
leis_poi.pgm				
min hhinnt=1				
Mean log-likelihood		-0.522919		
Number of cases		8690		
P01	konst	-5.9949	0.1311	-45.717
P02	m_ald/18	0.221	0.1044	2.116
P03	k_ald/18	0.3507	0.0867	4.044
P04	(m_ald/18)^2	-0.0729	0.0255	-2.856
P05	(k_ald/18)^2	-0.0794	0.02	-3.968
P06	hhinnt 100'	0.0349	0.0057	6.112
P07	and.tett	0.6981	0.1132	6.165
Number of iterations	35			
Minutes to convergence	0.12			
observert og forventet ant. Reiser	1646	1646		
4=4+ reiser				
ant.reiser	obs.	modell		
0	7378	7202.5		
1	1124	1341.2		
2	181	135.4		
3	44	10.1		
4	8	0.9		

TØI rapport 606/2002

Ifølge modellen vil reisehyppigheten for lange fritidsreiser for menn øke frem til 38 år for deretter å avta. Kvinner når maksimum reisehyppighet senere (44 år). Inntektselastisiteten i denne modell er lik parameteren for inntekt multiplisert med inntekten, dvs. elastisiteten øker med inntekten. For en inntekt på 300 000 kr (3) blir elastisiteten nokså nøyaktig 0,1. Høy andel bosatte i tettbygde strøk i bostedskommunen øker sannsynligheten for lange fritidsreiser.

I tillegg til de variable i modellen ovenfor er det også forsøkt med dummies for husholdningstyper. Dette gav ikke signifikante parametre. Trekker man inn førerkortinnehav og biltilgang får disse variable positiv effekt, men siden de er korrelert med alder, kjønn og inntekt får parametrene for disse variable dårligere t-verdier og blir til dels ikke-signifikante. Med de forklaringsvariable som er tilgjengelige når det gjelder person- og husholdningskarakteristika er det ikke lett å oppnå vesentlig forbedring av likelihood-verdien.

En opplagt kandidat når det gjelder å forbedre likelihood-verdien er sesong-dummies. Ved konstruksjon av sesongdummies er intervjumåned "lagget" med 1 og vi har 3 sesong-dummies:

vinter	des,jan,feb (intervjumåned jan, feb og mars)
vår	mars, april,mai (intervjumåned april, mai og juni)
sommer	juni,juli,august (intervjumåned juli, august og september)

Modell med sesongdummies er vist i Tabell V7.4.

Alle sesong-dummies blir signifikante og likelihood-verdien reduseres. De øvrige parametre påvirkes noe, men ikke i betydelig grad. Vi ser også at fordelings "hale" blir litt bedre. Parametrene for hhv menn og kvinner når det gjelder alder og alder kvadrert blir nå forholdsvis like (ikke signifikant forskjellige) og kan muligens slås sammen i denne modell. Maksimum reisefrekvens blir ved 38,9 år for menn og 43,0 år for kvinner.

Påvirkes reisefrekvens av andre geografiske forhold enn de som er angitt ved parameteren 8? Etter å ha kjørt med dummies for alle fylker ble følgende gruppert sammen:  
 g1 = Oslo, Akershus, Aust-Agder, Sogn&Fj, Sør-Trøndelag, Nord-Trøndelag.  
 g2= Hedmark og Oppland. Finnmark ble beholdt som eget fylke.

Tabell V7.4: Frekvensmodell for fritidsreiser med sesong-dummies

leis_poi.pgm				
Mean log-likelihood	-0.502232			
Number of cases	8690			
Parameters	Vnames	Estimates	Std. err.	Est./s.e.
P01	konst	-6.5629	0.1937	-33.884
P02	m_ald/18	0.4872	0.1463	3.329
P03	k_ald/18	0.5379	0.1435	3.748
P04	(m_ald/18)^2	-0.1135	0.0317	-3.582
P05	(k_ald/18)^2	-0.1133	0.03	-3.782
P06	hhinnt 100'	0.0319	0.0067	4.785
P07	and.tett	0.7496	0.1144	6.553
P08	vinter	-0.2532	0.0829	-3.052
P09	vår	0.2453	0.0817	3.003
P10	sommer	0.8992	0.0717	12.541
Number of iterations	52			
Minutes to convergence	0.205			
observert og forventet ant. reiser	1646	1646		
<b>4=4+ reiser</b>				
ant.reiser	obs.	modell		
0	7378	7233.6		
1	1124	1286		
2	181	152.9		
3	44	15.7		
4	8	1.7		

TØI rapport 606/2002

Igjen ser vi at likelihood-verdien forbedres og både g1 og g2 får presist bestemt parametere.

Det kan altså være grunn til å differensiere mellom fylker når det gjelder turgenerering. Vi får også en marginal forbedring når det gjelder å gjenskape fordelingen i data. P07 fanger tydeligvis opp effekten av større byer fordi en dummy for de 4 største byer ikke blir signifikant. Korrigert for de øvrige variable blir alder for maksimum reise-frekvens nå 37,7 år for menn og 42,6 år for kvinner.

Ved praktisk anvendelse i en modell er det et spørsmål om man skal benytte sesong-dummies. Det bør vel egentlig bare gjøres hvis man ellers i modellen har tilsvarende sesong-dummies. Dette kan f eks være aktuelt når det gjelder reisemiddelvalg.

I samplet er det 84,46 % som ikke har foretatt en lang reise i løpet av siste 30 dager. Hvis modellen i Tabell 3 anvendes på et døgn finner vi at 99,37 % ikke foretar en lang reise. 0,63 % foretar en lang reise, mens antall som foretar to eller flere lange reiser er neglisjerbart. I gjennomsnitt foretas det 0,0063 lange ferie- og fritidsreiser pr person pr

døgn. Dette stemmer med de data modellen er estimert på. På aggregert nivå er det neppe tvil om at de estimerte modeller vil fungere relativt bra hvis det ikke er systematiske skjevheter i de data som modellen er estimert på.

En forklaringsvariabel som kanskje kunne hatt betydning er antall utenlandsreiser som er foretatt. Disse kan "substituere" lange innenlandsreiser når det gjelder ferie og fritid.

I modellformuleringen har vi benyttet alder/18 og  $(\text{alder}/18)^2$  som variable for hhv menn og kvinner som er 20 år eller eldre. Dette er kontinuerlige variable som det ikke er hensiktsmessig å benytte i forbindelse med modellimplementering. I stedet bør disse variable erstattes av gjennomsnittsverdier innenfor de ulike aldersintervall. F eks hvis det er tale om en mann i aldersgruppen 30-34 år så kan  $(\text{ald}/18)$  erstattes av 32,5/18. For aldersgruppen 70-74 bør man kanskje benytte 72/18 siden gjennomsnittet her vil ligge lavere enn medianen i intervallet.

Som vi ser her og også vil se i de øvrige modeller er det i praksis ikke mulig å reprodusere observert fordeling på antall reiser ved hjelp av en

Poisson-modell. Dette skyldes primært at det er noen få personer i RVU som foretar relativt mange reiser. Hvem dette er, er trolig ganske tilfeldig, noe som gjør at man ikke – i praksis – kan regne med at det finnes modelltyper som vil være spesielt ”gode” på dette punkt. Konsekvensen av å presse frekvensmodellen inn i en Poisson-formulering er trolig at man får riktig prediksjon av antall reiser, men at vi ikke får tatt hensyn til at visse personer har en meget høy reisefrekvens i forhold til andre med samme bakgrunnsvariable. I praksis er dette neppe noe problem når det gjelder modell-anvendelser.

Poisson-modellen gjør at vi kan beregne en ”nyttefunksjon”,  $U$ , og deretter forventet antall reiser som  $e^U$  og det er det vi trenger innenfor det enkelt markedssegment.

I Tabell V7.6 er modellen utvidet med logsum fra reisemiddelvalg og destinasjonsvalg slik denne er beregnet av RAND. Det viser seg imidlertid at logsummen ikke blir signifikant forskjellig fra null hvis vi ikke sløyfer dummies for fylker. Tabell V7.6 må derfor sammenliknes med Tabell V7.4 som ikke har slike dummies. Som vi ser blir parameteren for logsum da signifikant forskjellig fra null, men har relativt lav tallverdi. For observasjoner hvor beregnet logsum mangler har vi benyttet gjennomsnittlig logsum.

Likelihood-verdien bedres ikke så mye som når vi har med fylkesdummies, men denne modell-spesifikasjon med signifikant logsumparameter er trolig å foretrekke til tross for dette.

Tabell V7.5: Frekvensmodell med sesong-og fylkesdummies 1)

leis_poi.pgm				
Mean log-likelihood	-0.498699			
Number of cases	8690			
Parameters	Vnames	Estimates	Std. err.	Est./s.e.
P01	konst	6.6321	0.222	-29.877
P02	m_ald/18	0.4666	0.1647	2.834
P03	k_ald/18	0.5263	0.1659	3.173
P04	(m_ald/18)^2	-0.1125	0.0346	-3.252
P05	(k_ald/18)^2	-0.1116	0.0345	-3.238
P06	hhinnt 100'	0.0287	0.0071	4.064
P07	and.tett	0.6312	0.1265	4.988
P08	vinter	-0.268	0.0841	-3.187
P09	vår	0.252	0.0818	3.081
P10	sommer	0.908	0.0721	12.595
P11	g1	0.3482	0.0547	6.363
P12	g2	0.5254	0.1063	4.941
P13	Finnmark	-0.5782	0.3572	-1.619
Number of iterations	75			
Minutes to convergence	0.36983			
observert og forventet ant. reiser	1646	1646		
4=4+ reiser				
ant.reiser	obs.	modell		
0	7338	7239.3		
1	1120	1276.5		
2	180	155.3		
3	44	16.8		
4	8	2.0		

TØI rapport 606/2002



Tabell V7.6: Frekvensmodell for ferie- og fritidsreiser med logsum

Parameters	Vnames	Estimates	Std. err.	Est./s.e.
P01	konst	-7.1846	0.3074	-23.369
P02	m_ald/18	0.4511	0.147	3.068
P03	k_ald/18	0.5397	0.1436	3.758
P04	(m_ald/18)^2	-0.1058	0.0318	-3.328
P05	(k_ald/18)^2	-0.1144	0.03	-3.817
P06	hhinnt 100'	0.0319	0.0066	4.859
P07	and.tett	0.7187	0.1148	6.263
P08	vinter	-0.2309	0.0834	-2.769
P09	vår	0.2546	0.0818	3.114
P10	sommer	0.907	0.0718	12.636
P11	logsum	0.072	0.0276	2.61
Number of iterations	54			
Minutes to convergence	0.24533			
observert og forventet ant. reiser	1646	1646		
4=4+ reiser				
ant.reiser	obs.	modell		
0	7338	7234.3		
1	1120	1285		
2	180	153.1		
3	44	15.9		
4	8	1.7		

TØI rapport 606/2002

## Tjeneste- og arbeidsreiser

Mens alle IO i RVU i prinsippet kan foreta en lang ferie eller fritidsreise er arbeids- og tjenestereiser ”forbeholdt” yrkesaktive. Dette innebærer at vi har valget mellom å estimere en modell på alle IO eller en modell for bare yrkesaktive (etter en eller annen definisjon). Ved å estimere en modell på alle IO vil vi implisitt ta hensyn til at yrkesaktivitet varierer med kjønn og alder og vi trenger ikke en metode eller modell for å fordele personer mellom yrkesaktive og ikke-yrkesaktive i forkant. Isolert sett er dette en fordel. Hvis frekvensmodellen skal gjelde bare for yrkesaktive må vi først bestemme oss for hvilket kjennetegn som skal legges til grunn for å definere yrkesaktivitet og ved implementering må vi ha en modell som fordeler etter samme definisjon.

Nedenfor presenteres en modell hvor alle IO er med og 3 modeller hvor det er benyttet ulike definisjoner av yrkesaktivitet:

Mer enn 30 timer pr uke (kode 1 og 2 i spm 2)  
 Heltid og deltid dvs. kode 1, 2 og 3 i spm 2  
 Hovedbeskjeftigelse yrkesaktiv, dvs. kode 1 i spm 12

I Tabell V7.7 som omfatter alle IO ser vi at det er 900 reiser i data og dette reproduseres som summen av beregnede forventet antall reiser i modellen. Likelihood og t-verdier blir meget bra, men noe av det som ”forklares” er altså yrkesaktivitet.

Inntekt er et ”problem” i denne modell er i likhet med den modell som ligger i NTM. Vi fanger her opp variasjoner i et tverrsnitt i befolkningen, men for denne type reiser kan det reises tvil om relevansen av dette hvis vi skal se på inntektsendringer over tid.

Når inntekt inngår lineært i nyttefunksjonen som her er det også slik at inntektselastisiteten øker med inntekten. Her er den f.eks 0,34 ved kr 300 000 og 0,57 ved kr 500 000. I og for seg har vi ikke grunnlag for å si at dette er urimelig. Alternativet til inntekt er personlige kjennetegn som yrke og utdanning. Menn vil i flg estimerte parametre ha maksimum reisefrekvens ved 37,6 år og tilsvarende maksimum for kvinner nås ved 43,7 år, dvs. forbausende nær tilsvarende for ferie og fritidsreiser.

Tabell V7.7: Arbeids- og tjenestereiser, alle IO

work_poi.pgm				
Mean log-likelihood		-0.331603		
Number of cases		8690		
Parameters	Vnames	Estimates	Std. err.	Est./s.e.
P01	konst	-9.364	0.314	-29.817
P02	m_ald/18	3.557	0.3214	11.066
P03	k_ald/18	3.2937	0.2735	12.045
P04	m_ald/18)^2	-0.8503	0.0879	-9.678
P05	k_ald/18)^2	-0.6785	0.0613	-11.066
P06	pinnt 100'	0.1136	0.0063	18.155
P07	ensl m/barn	-0.356	0.2059	-1.729
P08	f18,19&20	0.5591	0.103	5.428
P09	f15	-0.4975	0.1202	-4.139
Number of iterations	65			
Minutes to convergence	0.2435			
observert og forventet ant. reiser	900	900		
4=4+ reiser				
ant.reiser	obs.	modell		
0	8086	7866.5		
1	427	758.6		
2	106	59.2		
3	41	4.1		
4	30	1.7		

TØI rapport 606/2002

Tabell V7.8: Arbeids- og tjenestereiser, arbeidstid  $\geq 30t/u$ 

work_poi.pgm				
Mean log-likelihood		-0.500079		
Number of cases		4997		
Parameters	Vnames	Estimates	Std. err.	Est./s.e.
P01	konst	-7.3868	0.3951	-18.695
P02	m_ald/18	1.5999	0.3968	4.032
P03	k_ald/18	1.6595	0.3493	4.751
P04	m_ald/18)^2	-0.3903	0.1021	-3.822
P05	k_ald/18)^2	-0.3392	0.0759	-4.466
P06	pinnt 100'	0.139	0.0093	14.909
P07	ensl m/barn	-0.1752	0.2117	-0.827
P08	f18,19&20	0.6377	0.1042	6.121
P09	f15	-0.4141	0.1214	-3.41
Number of iterations	62			
Minutes to convergence	0.12817			
observert og forventet ant. reiser	851	851		
4=4+ reiser				
ant.reiser	obs.	modell		
0	4433	4236.9		
1	391	685.9		
2	104	66.5		
3	41	5.5		
4	28	2.2		

TØI rapport 606/2002

Tabell V7.8 er estimert på personer som oppgir minimum 30 timers arbeidstid pr uke, dvs. tilnærmet fulltid og mer.

Vi ser at vi med denne avgrensning "mister" 3693 observasjoner og 49 reiser.

Overensstemmelsen mellom observert og modellprodusert antall reiser blir bedre fordi den modellproduserte fordeling får en "tykkere" hale. t-verdiene blir jevnt over dårligere bortsett fra for "pinnt". Maksimum reisefrekvens for menn blir nå ved 36,9 år og for kvinner ved 44,0 år, dvs. små endringer i forhold til foregående modell. P07 er ikke signifikant forskjellige fra null. Dette må vel forstås dit hen at når denne kategori først arbeider "fulltid" så er ikke reisefrekvensen signifikant forskjellig fra andre.

Tabell V7.9 omfatter dem som har oppgitt arbeidstid i spm 2, dvs. kode 1-3. Ved å ta med

personer med mindre enn 30 t pr uke økes antall observasjoner med 1058 og antall reiser med 29. t-verdier og likelihood-verdi blir forbedret, med parameteren for inntekt reduseres. Vi "mangler" fortsatt 20 reiser.

I Tabell V7.10 har vi selektert de IO som har oppgitt "yrkesaktiv, inntektsgivende arbeid" som hovedbeskjeftigelse.

I forhold til Tabell V7.9 mister vi nå 511 observasjoner, men bare én reise. Inntekt får nå en høyere parameter som er større og som har bedre t-verdi. De to siste modeller gir lavere alder for maksimal reisefrekvens for menn, men for kvinner blir denne alder lite påvirket. Modellen som er vist i Tabell V7.10 gir klart det høyeste estimat på parameteren for inntekt.

Tabell V7.9: Arbeids- og tjenestereiser, "yrkesaktive"

work_poi.pgm				
Mean log-likelihood	-0.445956			
Number of cases	6055			
Parameters	Vnames	Estimates	Std. err.	Est./s.e.
P01	konst	-8.4996	0.3851	-22.073
P02	m_ald/18	2.6863	0.3781	7.105
P03	k_ald/18	2.6002	0.3325	7.821
P04	m_ald/18)^2	-0.6253	0.0984	-6.356
P05	k_ald/18)^2	-0.522	0.0732	-7.128
P06	pinnt 100'	0.0987	0.0072	13.754
P07	ensl m/barn	-0.2571	0.209	-1.23
P08	f18,19&20	0.5942	0.1032	5.757
P09	f15	-0.4745	0.1211	-3.92
Number of iterations	64			
Minutes to convergence	0.16567			
observed og forventet ant. reiser	880		880	
4=4+ reiser				
ant.reiser	obs.	modell		
0	5466	5255.8		
1	414	728.9		
2	105	64		
3	41	4.8		
4	29	1.5		

TØI rapport 606/2002

Tabell V7.10: Arbeids- og tjenestereiser, hovedbeskjeftigelse yrkesaktiv

work_poi.pgm				
Mean log-likelihood	-0.476308			
Number of cases	5544			
Parameters	Vnames	Estimates	Std. err.	Est./s.e.
P01	konst	-7.069	0.382	-18.504
P02	m_ald/18	1.413	0.3832	3.687
P03	k_ald/18	1.3213	0.3376	3.914
P04	m_ald/18)^2	-0.3746	0.0982	-3.814
P05	k_ald/18)^2	-0.2755	0.0734	-3.751
P06	pinnt 100'	0.1495	0.0085	17.539
P07	ensl m/barn	-0.2745	0.2128	-1.29
P08	f18,19&20	0.6011	0.1031	5.831
P09	f15	-0.4466	0.1212	-3.685
Minutes to convergence	0.14467			
observert og forventet ant. reiser	879	879		
4=4+ reiser				
ant.reiser	obs.	modell		
0	4959	4755		
1	410	718.1		
2	104	63.8		
3	41	4.8		
4	30	2.3		

TØI rapport 606/2002

Tabell V7.11: Arbeids- og tjenestereiser, alle IO. Med sesong dummies

work_poi.pgm				
Mean log-likelihood	-0.327795			
Number of cases	8690			
Parameters	Vnames	Estimates	Std. err.	Est./s.e.
P01	konst	-8.778	0.3079	-28.505
P02	m_ald/18	3.5143	0.3121	11.259
P03	k_ald/18	3.1253	0.2643	11.826
P04	m_ald/18)^2	-0.8477	0.0865	-9.794
P05	k_ald/18)^2	-0.6475	0.0596	-10.873
P06	vinter	-0.7165	0.0918	-7.807
P07	vår	-0.3828	0.0925	-4.137
P08	sommer	-0.5026	0.0926	-5.427
P09	pinnt 100'	0.1134	0.0062	18.278
P10	ensl m/barn	-0.3913	0.2064	-1.896
P11	f18,19&20	0.5348	0.103	5.191
P12	f15	-0.477	0.1203	-3.966
Number of iterations	77			
Minutes to convergence	0.3405			
observert og forventet ant. reiser	900	900		
4=4+ reiser				
ant.reiser	obs.	modell		
0	8086	7872		
1	427	749.4		
2	106	61.6		
3	41	4.9		
4	30	2.2		

TØI rapport 606/2002

I Tabell V7.11 er modellen kjørt med "sesongdummies" for vår sommer og vinter og på hele materialet. Dette bedrer likelihooden og sesongdummies får parametere med bra t-verdier. Det gir imidlertid moderate utslag på andre parametere i forhold til Tabell 7. Sesongdummies blir altså høyst signifikante for denne type reiser og tyder på at reisefrekvensen er desidert størst om høsten og lavest om vinteren.

Tjeneste- og arbeidsreiser har altså relativt store sesongvariasjoner, med topp i høstmånedene og bunn i vintermånedene.

Logsummer beregnet av RAND omfatter 7602 observasjoner. For kunne benytte alle IO har jeg på dem som mangler lagt inn gjennomsnittlig logsum. Estimering med logsum i tillegg til de variable som er inkludert i Tabell V7.12 gir da følgende resultat.

Som vi ser, gir logsummen en vesentlig forbedring av likelihoodverdien og logsumparameteren får en meget høy t-verdier og en relativt høy absoluttverdi. Siden logsummen påvirkes av en rekke variable, inkludert inntekt, må vi forvente at andre parametre også påvirkes. Her ser det først og fremst ut til å gjelde inntekt og parameteren for Møre og Romsdal fylke. For Nord-Norge blir imidlertid parameteren høyere i tallverdi og får høyere t-verdi. Vi ser også at "halen" på den reproduserte fordeling blir bedre i overensstemmelse med observert fordeling, selv om det fortsatt mangler mye på at modellen "fanger opp" dem med 3 eller flere reiser.

Tabell V7.12: Arbeids- og tjenestereiser, alle IO. Med sesong dummies og logsum

work_poi.pgm				
Mean log-likelihood		-0.321397		
Number of cases		8690		
Parameters	Vnames	Estimates	Std. err.	Est./s.e.
P01	konst	-9.8164	0.2887	-34.002
P02	m_ald/18	2.9722	0.2904	10.235
P03	k_ald/18	2.6317	0.2383	11.045
P04	m_ald/18)^2	-0.7335	0.0834	-8.79
P05	k_ald/18)^2	-0.5517	0.0544	-10.135
P06	vinter	-0.7094	0.0916	-7.745
P07	vår	-0.4077	0.0926	-4.402
P08	sommer	-0.5538	0.0928	-5.967
P09	pinnt 100'	0.0869	0.0074	11.738
P10	ensl m/barn	-0.3219	0.206	-1.563
P11	f18,19&20	0.9837	0.1106	8.897
P12	f15	-0.243	0.1222	-1.989
P13	logsum	0.3847	0.0355	10.846
Number of iterations	117			
Minutes to convergence	0.56117			
obsvert og forventet ant. reiser	900	900		
4=4+ reiser				
ant.reiser	obs.	modell		
0	8086	7881.5		
1	427	730.5		
2	106	68.4		
3	41	7.3		
4	30	2.3		

TØI rapport 606/2002

## Besøksreiser

I likehet med ferie- og fritid er det ikke grunn til å selektere personer som kan foreta slike reiser. Etter en del forsøk med ulike variable ser de ut til at modellen som er vist i Tabell V7.13 er ganske brukbar.

I denne modellen har vi trukket inn yrkesaktivitet (over 30 t arbeidstid pr uke). Isolert sett er det en ulempe fordi det innebærer at man må ha en segmentering etter yrkesaktivitet. Hvis man ikke har med yrkesaktivitet som variabel er det vanskelig å få en noenlunde presis parameter for inntekt. Når det gjelder husstandstype vil dette uansett bli benyttet i forbindelse med segmentering.

I modellen som er vist i Tabell V7.14 har vi sløyet både yrkesaktivitet og inntekt som variable. Til gjengjeld har vi tatt med en dummy for personer over 75 år, en dummy for de tre nordligste fylker, en dummy for TP10-byer 5-10 og mindre

byer. Skal inntekt ha noen en virkning i denne modell må dette skje indirekte via logsummen.

Den modellen blir altså marginalt bedre når det gjelder likelihoodverdi, men har en forklaringsvariabel ekstra. Dette synes å være det reiseformål hvor det er vanskeligst å finne "gode" og naturlige forklaringsvariable.

Tabell V7.15 viser samme modell som over men med logsum i variabel som tillegg til de øvrige. Logsumparameteren er ikke signifikant forskjellig fra null på 5-prosent nivå, men er av rimelig størrelse. Også her er det benyttet gjennomsnittlig logsum for de observasjoner hvor logsummen mangler. Generelt ser det ut som at logsum (og inntekt) betyr relativt lite for besøksreiser.

Tabell V7.13: Modell for besøksreiser

visi_poi.pgm				
modelltype =poisson				
Mean log-likelihood		-0.418764		
Number of cases		8690		
Inverse of computed Hessian				
Parameters	Vnames	Estimates	Std. err.	Est./s.e.
P01	konstant	-5.179	0.1627	-31.84
P02	(ald/18)^2	-0.0506	0.0073	-6.976
P03	hinnt 100'	0.0445	0.0247	1.798
P04	enslig u/b	0.4507	0.1593	2.829
P05	par m/barn	-0.2831	0.0742	-3.816
P06	kvinne	0.2938	0.1545	1.902
P07	under 20 år	-0.5609	0.1278	-4.389
P08	yrkakt helt	-0.2136	0.0677	-3.154
P09	FM,TRO,S-T	0.33	0.0982	3.359
P10	M&R	-0.2757	0.093	-2.966
Number of iterations	63			
Minutes to convergence	0.2425			
observert og forventet ant. reiser	1196	1196		
4=4+ reiser				
ant.reiser	obs.	modell		
0	7651	7579.1		
1	910	1030.3		
2	108	76.3		
3	15	4.1		
4	6	0.2		

TØI rapport 606/2002

Tabell V7.14: Besøksreiser, uten yrkesaktivitet og inntekt som forklaringsvariable

visi_poi.pgm				
modelltype =poisson				
Mean log-likelihood	-0.418708			
Number of cases	8690			
Parameters	Vnames	Estimates	Std. err.	Est./s.e.
P01	konstant	-5.2395	0.1535	-34.138
P02	(ald/18)^2	-0.0383	0.0079	-4.842
P03	kvinne	0.2839	0.1528	1.858
P04	u20	-0.4083	0.1197	-3.413
P05	ensl u/barn	0.4498	0.1568	2.869
P06	par m/barn	-0.3011	0.0727	-4.141
P07	f18-19-20	0.227	0.094	2.416
P08	M&R	-0.2803	0.0964	-2.907
P09	o75	-0.3105	0.2183	-1.422
P10	TP10-5-10	-0.4039	0.1219	-3.313
P11	mindre by	-0.1405	0.0825	-1.702
Number of iterations	69			
Minutes to convergence	0.30033			
observert og forventet ant. reiser 4=4+ reiser	1196	1196		
ant.reiser	obs.	modell		
0	7651	7578.9		
1	910	1030.6		
2	108	76.2		
3	15	4.1		
4	6	0.2		

TØI rapport 606/2002

Tabell V7.15: Besøksreiser med logsum

visi_poi.pgm				
modelltype =poisson				
Mean log-likelihood	-0.4186			
Number of cases	8690			
Parameters	Vnames	Estimates	Std. err.	Est./s.e.
P01	konstant	-5.8804	0.4926	-11.939
P02	(ald/18)^2	-0.0376	0.0080	-4.723
P03	kvinne	0.2771	0.1532	1.809
P04	u20	-0.3885	0.1207	-3.220
P05	ensl u/barn	0.4389	0.1573	2.790
P06	par m/barn	-0.2925	0.0730	-4.007
P07	f18-19-20	0.2949	0.1059	2.784
P08	M&R	-0.2404	0.1007	-2.388
P09	o75	-0.2830	0.2194	-1.290
P10	TP10-5-10	-0.4176	0.1223	-3.414
P11	mindre by	-0.1507	0.0829	-1.819
P12	logsum	0.0556	0.0406	1.371
Number of iterations	82			
Minutes to convergence	0.36067			
observert og forventet ant. reiser 4=4+ reiser	1196	1196		
ant.reiser	obs.	modell		
0	7651	7579.0		
1	910	1030.4		
2	108	76.3		
3	15	4.1		
4	6	0.2		

TØI rapport 606/2002

## Andre private reiser

Tabell V7.16 gir signifikante parametre og bra likelihoodverdi. Også her har vi problemer med å predikere fordelings hale, men i likhet med Tjeneste- og arbeidsreiser inngår her også arbeidsreiser, noe som gjør at vi kan ha noen få personer med mange reiser og disse vil ikke modellen kunne identifisere.

Denne modellen vil også være grei og håndtere når det gjelder segmentering. Den implisitte inntektselastisitet ved en husholdsinntekt på kr

300 000 er her ca 0,2. I Tabell V7.17 har vi i tillegg med logsummen som variabel. Dette gir en betydelig bedring av likelihooden og logsumparameteren blir høy i absoluttverdi og får høy t-verdi. Samtidig blir imidlertid P02 og P06 ikke-signifikante og P06 får ”feil” fortegn.

Modellen i Tabell V7.17 ble derfor re-estimert uten mald/19 og hint som variable. Resultatene er vist i Tabell V7.18.

Tabell V7.16: Modell for andre private reiser

oth_poi.pgm				
Mean log-likelihood		-0.392563		
Number of cases		8690		
Parameters	Vnames	Estimates	Std. err.	Est./s.e.
P01	konstant	-5.4680	0.1433	-38.165
P02	mald/18	0.4655	0.1357	3.430
P03	kald/18	0.7512	0.1093	6.870
P04	(mald/18)^2	-0.1448	0.0349	-4.153
P05	(kald/18)^2	-0.1941	0.0263	-7.387
P06	hinnt 100'	0.0649	0.0205	3.169
P07	and.tett	-0.7316	0.1170	-6.255
P08	fylke 2	-0.3949	0.0994	-3.973
P09	fylke 7	0.5940	0.1325	4.482
P10	fy 9,10 &11	-0.2921	0.1203	-2.427
P11	fy 17,18&19	0.4045	0.0989	4.090
Number of iterations	70			
Minutes to convergence	0.31317			
observert og forventet ant. reiser	1074	1074		
4=4+ reiser				
ant.reiser	obs.	modell		
0	7850	7693.1		
1	691	924.4		
2	99	68.2		
3	26	4.1		
4	24	0.3		



Tabell V7.17: Andre private reiser med logsum

oth_poi.pgm				
Mean log-likelihood		-0.377796		
Number of cases		8690		
Parameters	Vnames	Estimates	Std. err.	Est./s.e.
P01	konstant	-8.8548	0.2550	-34.724
P02	mald/18	0.1213	0.1370	0.886
P03	kald/18	0.4101	0.1102	3.723
P04	(mald/18)^2	-0.0606	0.0350	-1.733
P05	(kald/18)^2	-0.1162	0.0267	-4.346
P06	hinnt 100'	-0.0047	0.0295	-0.159
P07	and.tett	-0.8797	0.1175	-7.486
P08	fylke 2	-0.4524	0.0993	-4.555
P09	fylke 7	0.3196	0.1334	2.396
P10	fy 9,10 &11	-0.3735	0.1205	-3.099
P11	fy 17,18&19	0.7422	0.1003	7.398
P12	logsum	0.4066	0.0248	16.365
Number of iterations	70			
Minutes to convergence	0.315			
observert og forventet ant. reiser	1074	1074		
4=4+ reiser				
ant.reiser	obs.	modell		
0	7850	7712.3		
1	691	890.2		
2	99	79.4		
3	26	7.3		
4	24	0.8		

Tabell V7.18: Andre private reiser med logsum (2)

oth_poi.pgm				
Mean log-likelihood		-0.377842		
Number of cases		8690		
Parameters	Vnames	Estimates	Std. err.	Est./s.e.
P01	konstant	-8.8308	0.2519	-35.061
P02	kald/18	0.3542	0.0891	3.974
P03	(mald/18)^2	-0.0315	0.0115	-2.732
P04	(kald/18)^2	-0.1058	0.0237	-4.454
P05	and.tett	-0.8810	0.1162	-7.584
P06	fylke 2	-0.4547	0.0991	-4.587
P07	fylke 7	0.3163	0.1334	2.371
P08	fy 9,10 &11	-0.3759	0.1204	-3.123
P09	fy 17,18&19	0.7464	0.1002	7.449
P10	logsum	0.4105	0.0239	17.176
Number of iterations		60		
Minutes to convergence		0.23167		
observert og forventet ant. reiser		1074	1074	
<b>4=4+ reiser</b>				
ant.reiser	obs.	modell		
0	7850	7712.4		
1	691	890.1		
2	99	79.4		
3	26	7.3		
4	24	0.8		

Det litt påfallende er at vi får to modeller hvor logsumparameteren blir relativt høy i tallverdi og får en meget høy t-verdi, mens to modeller gir lav tallverdi på parameteren og relativt lav t-verdi.

### Sammenlikning av alternative sannsynlighetsmodeller

Som nevnt innledningsvis er det ikke bare Poisson-modeller og logit modeller som kan benyttes i forbindelse med reisefrekvenser. For sammenliknings skyld er en modell for hvert reiseformål også estimert som binomisk modell med logit-formulering av sannsynlighetene og geometrisk fordeling med logitsannsynligheter. Parameterene i seg selv vil selvsagt få forskjellig verdi med i forhold til Poisson, med det som direkte kan sammenliknes er likelihoodverdi og modellen evne til å reprodusere fordelingen når det gjelder antall reiser.

Av de tre modelltyper er geometrisk fordeling best mhp. likelihood og reprodusert fordeling, men denne fordelings forventningsverdier reproduserer

ikke observert antall reiser. Forskjellen er imidlertid marginal. Når det gjelder Poisson kontra binomisk så virker er Poisson marginalt bedre både på likelihood og reproduksjon av observert fordeling.

Som tidligere nevnt refererer geometrisk fordeling ikke til en bestemt periode eller til et bestemt antall "forsøk" og den har den ulempe at observert antall reiser ikke reproduseres av modellen. Dette kommer enda tydeligere fram for neste reiseformål.

Tabell V7.20 viser det samme fenomen: Poisson er marginalt bedre enn binomisk logit, mens geometrisk scorer høyest på likelihoodverdi og reprodusert fordeling. Her ser vi imidlertid det som ble nevnt tidligere, geometrisk fordeling kan for en lenger periode gi en vesentlig overprediksjon av antall reiser. Dette skyldes at for én eller noen få observasjoner vil predikert sannsynlighet ligge meget nær én, og dette gir en meget stor forventningsverdi.

Tabell V7.19: Ferie- og fritidsreiser

	Observert	Poisson	Binomisk	Geometrisk
Likelihood/obs		-0,502232	-0,502868	-0,497378
Ingen reiser	7338	7233,6	7228,5	7364,6
1 reise	1120	1286,0	1294,8	1078,0
2 reiser	180	152,9	150,6	192,3
3 reiser	44	15,7	14,7	41,0
4+ reiser	8	1,7	1,4	14,1
Predikerte reiser	1646	1646,0	1646,0	1648,2

Tabell V7.20: Tjeneste- og arbeidsreiser

	Observert	Poisson	Binomisk	Geometrisk
Likelihood/obs		-0,321397	-0,322913	-0,297098
Ingen reiser	8086	7881,5	7879,1	7948,6
1 reise	427	730,5	734,5	618,3
2 reiser	106	68,4	67,4	88,8
3 reiser	41	7,3	6,9	18,5
4+ reiser	30	2,3	2,1	15,8
Predikerte reiser	900	900,0	900,0	453994,9

Tabell V7.21: Besøksreiser

	Observert	Poisson	Binomisk	Geometrisk
Likelihood/obs		-0,417724	-0,418134	-0,413549
Ingen reiser	7652	7581,1	7578,5	7651,0
1 reise	910	1028,5	1033,3	905,8
2 reiser	108	77,0	75,2	115,6
3 reiser	15	4,2	3,9	15,9
4+ reiser	6	0,2	0,2	2,8
Predikerte reiser	1196	1196,0	1196,0	1196,1

Tabell V7.22: Andre reiser

	Observert	Poisson	Binomisk	Geometrisk
Likelihood/obs		-0,377842	-0,378902	-0,364892
Ingen reiser	7580	7712,4	7228,5	7781,7
1 reise	691	890,1	894,8	776,9
2 reiser	99	79,4	78,1	105,9
3 reiser	26	7,3	6,8	19,3
4+ reiser	24	0,8	0,7	6,2
Predikerte reiser	1074	1074,0	1074,1	1073,8

I Tabell V7.21 har vi samme rangering, men her gir den geometriske fordeling også praktisk talt korrekt prediksjon.

Tabell V7.22 viser det samme mønster som de øvrige og også her predikerer den geometriske fordeling praktisk talt eksakt antall reiser.

For samtlige sammenlikninger gjelder det at parametrene blir noe forskjellige, men t-verdier for parametrene er gjennomgående av samme størrelsesorden.

Selv om den geometriske fordeling jevnt over gir best resultat bortsett fra på predikert antall reiser

er Poisson-fordelingen kanskje å foretrekke fordi den har en grei tolkning og er konsistent med underliggende modeller av typen logit når det gjelder destinasjons- og reisemiddelvalg. Den geometriske fordeling kan tydeligvis gi helt

urimelige resultater under visse omstendigheter. Dette henger trolig sammen med at den ikke refererer seg til en bestemt tidsperiode eller et gitt antall ”forsøk” som de to andre modeller.