

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

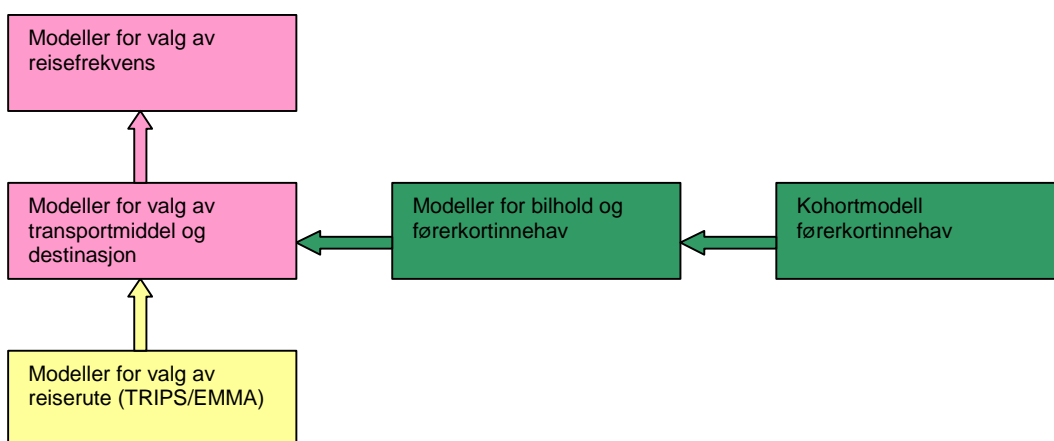
I dette dokumentet dokumenteres del B av arbeidet samt de deler av del A som foreløpig ikke er dokumentert. I dette arbeidet er det estimert modeller for valg av transportmiddel og destinasjon (multinomisk logit og strukturert multinomisk logit), for valg av reisefrekvens (binomisk vektet logit), og 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 tverrsnittsmo- dellersom i utgangspunktet gjensker det førerkortinnehav man hadde i 1997/98 når RVU ble gjennomført. Når modellene skal benyttes på fremtidsår vil ikke disse modellene ivareta kohorteffektene i førerkortinnehavet. Dataene fra kohortmodellene benyttes derfor til å kalibrere førerkortinnehavet for et gitt antall fremtidsår, slik at kohorteffektene ivaretas.

Fører kort/bilholdsmodellene ivaretar endringer i inntekt og befolkningssammensetning 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 variabler 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.

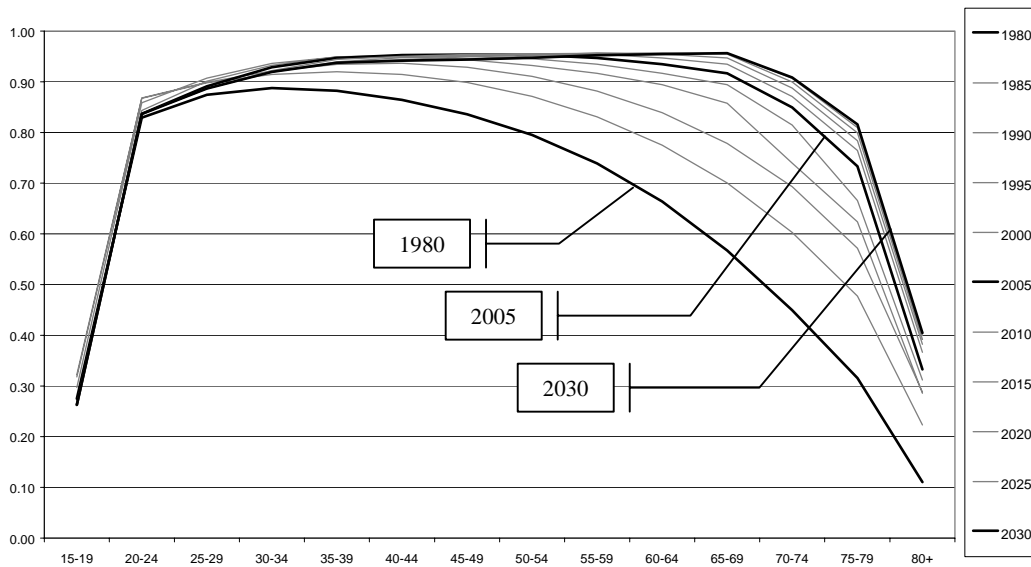
Prosedyren, som vi har kalt *kohortmodeller for førerkortinnehav*, er beskrevet i kapittel 5. Ved hjelp av denne prosedyren har vi beregnet fremtidige førerkortandeler i fem års intervaller frem til år 2030. Disse andelene skal i implementeringen av modellene for bilhold og førerkortinnehav benyttes som kalibreringsgrunnlag for å tilpasse førerkortinnehavet til tids-effekten. 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.

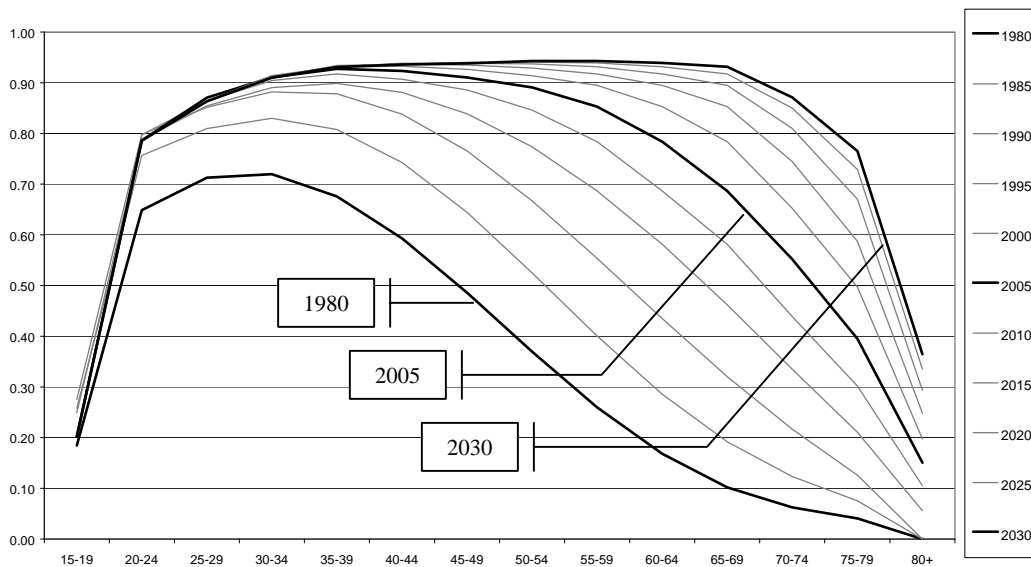
Modellene for valg av transportmåte (reisemiddel) og destinasjon (reisemål) er estimert på basis av de reisetidskomponentene og kostnadene som beregnes i nettverksmodellen EMMA, samt ulike sonedata (befolkning, arbeidsplasser med mer). Estimeringsarbeidet, slik det er dokumentert i kapittel 6, har bestått 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 kvadratrotten 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 kvadratrotten 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 (reisehensiktene). 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.