

Interne notater

STATISTISK SENTRALBYRÅ

91/8

23. mai 1991

To notater fra prosjektet "MOSART og trygd".

Behov for bruttostrømsstatistikk i modellen MOSART.	1
Forprosjektrapport på "MOSART og trygd".	9

Behov for bruttostrømsstatistikk i modellen MOSART.

av

Dennis Fredriksen, Statistisk sentralbyrå.

1. Innledning.

Dette notatet presenterer MOSART og drøfter de krav denne modellen stiller til mikrodata, deriblant bruttostrømsstatistikk for arbeidsmarkedet. Navnet MOSART er et akronym for "MOdell for mikrosimulering av Skolegang og ARbeidsTilbud", og modellen er på mange måter en bruttostrømsmodell for arbeidstilbud. Avsnitt 2 og 3 gir en presentasjon av førsteutgaven av MOSART og noen av de første utvidelsene vi har planlagt å gjennomføre på modellen. En av disse utvidelsene vil gå på arbeid og trygd, og avsnitt 4 drøfter noen av de krav dette stiller til bedre datagrunnlag for modellen. Konklusjonen på dette avsnittet er at vi vil ha stor nytte av en database som er longitudinell og omfatter alle de kjennetegn vi bruker i modellen. Tilslutt gis en kort presentasjon av det dataprojektet vi selv har startet opp på emnet.

2. MOSART.

MOSART er en modell som bygger på simulering av livsløpene til en modellpopulasjon av individer. Førsteutgaven av modellen, MOSART-I, er nylig ferdig og vil dekke framskrivninger innen områdene utdanning og arbeidstilbud. Denne versjonen vil bli beskrevet nedenfor sammen med noen av de første utvidelsene av modellen. Vi har hele veien lagt stor vekt på å få en fleksibel modell som vi gradvis kan forbedre gjennom mindre delprosjekter. En figur som gir en viss oversikt over modellen finnes på side 5.

Utgangspopulasjonen.

For senest tilgjengelige år, kalt startåret¹ i modellen, har vi koblet SSB's individbaserte utdanningsstatistikk og 'kvinnefilen'. Den førstnevnte omfatter hele befolkningen ved utgangen av året med utdanningsstatus ved begynnelsen av oktober samme år. Den sistnevnte omfatter alle kvinner ved utgangen av året med mer utfyllende beskrivelse av demografiske kjennetegn som fullstendig fødselshistorie og ektefelles personnummer. Fra denne koblede filen trekkes utvalg på to prosent som brukes som utgangspopulasjoner i modellen; dvs den modellpopulasjon som MOSART starter simuleringen med. MOSART-I vil forøvrig bare omfatte personer som er 16 år og eldre. At modellen er basert på et utvalg av befolkningen, gjør at MOSART bruker relativt lite EDB-ressurser. Det igjen gjør det relativt enkelt og billig å bruke modellen, også under alternative forutsetninger om utviklingen framover. I utvalget er ektefeller trukket sammen, og modellen ivaretar denne

¹ Dette er for øyeblikket året 1987.

tilknytningen mellom to individer i modellpopulasjonen. Vi ikke fått med fullstendige forhistorier til individene i utgangspopulasjonen. Utvalget vil heller ikke være representativt for befolkningen for perioden forut for startåret.

Simuleringen.

For perioden etter startåret vil modellen simulere de videre livsløpene til individene i modellpopulasjonen. Modellen vil for hvert år som går legge nye individer til modellpopulasjonen. Disse vil være nettoinnvandrere og nye kull 15-åringer tilpasset størrelsen på utgangspopulasjonen. Simuleringen foregår ved at modellen på grunnlag av overgangssannsynligheter trekker hvilken utvikling et individ vil få fra et år til neste år. Modellen er da ikke utsagnskraftig om utviklingen for hvert enkelt individ, men kun for tilstrekkelig store grupper av befolkningen. De overgangssannsynlighetene modellen bruker vil bl.a. avhenge av kjennetegn ved individet selv, og kan ofte tolkes som en beskrivelse av individuell atferd. Et eksempel er å tolke sannsynligheten for å være i arbeidsstyrken (yrkesprosenten) som et uttrykk for arbeidsmarkedsatferd. I tillegg er det nødvendig å legge inn forutsetninger om utviklingen i den individuelle atferden, f.eks. økning (nedgang) i yrkesprosentene. Med unntak av dødelighet har vi valgt å holde alle overgangssannsynligheter konstante. Dette vil selvsagt ikke beskrive den faktiske utvikling, men vil representere en første tilnærming, og kan gi innsikt i de problemstillinger modellen tar opp. I den grad man har oppfatninger om hvordan atferden vil utvikle seg, vil det være en relativt enkelt sak å innarbeide dette i modellen.

MOSART simulerer livsløpene for alle individene tilnærmet parallelt, ved at modellen simulerer hele modellpopulasjonen for et år før den går videre til neste år. Det gjør det mulig å ha tilknytninger mellom individer i modellen, f.eks. ekteskap. Det vil også være overkommelig å bruke egenskaper ved modellpopulasjonen samlet for fjoråret som en forklaringsvariabel for modellens overgangssannsynligheter. Man kan f.eks. la differansen mellom samlet arbeidstilbud i MOSART og en framskrivning av sysselsetting, en framskrevet arbeidsledighet, påvirke det individuelle arbeidstilbudet året etter.

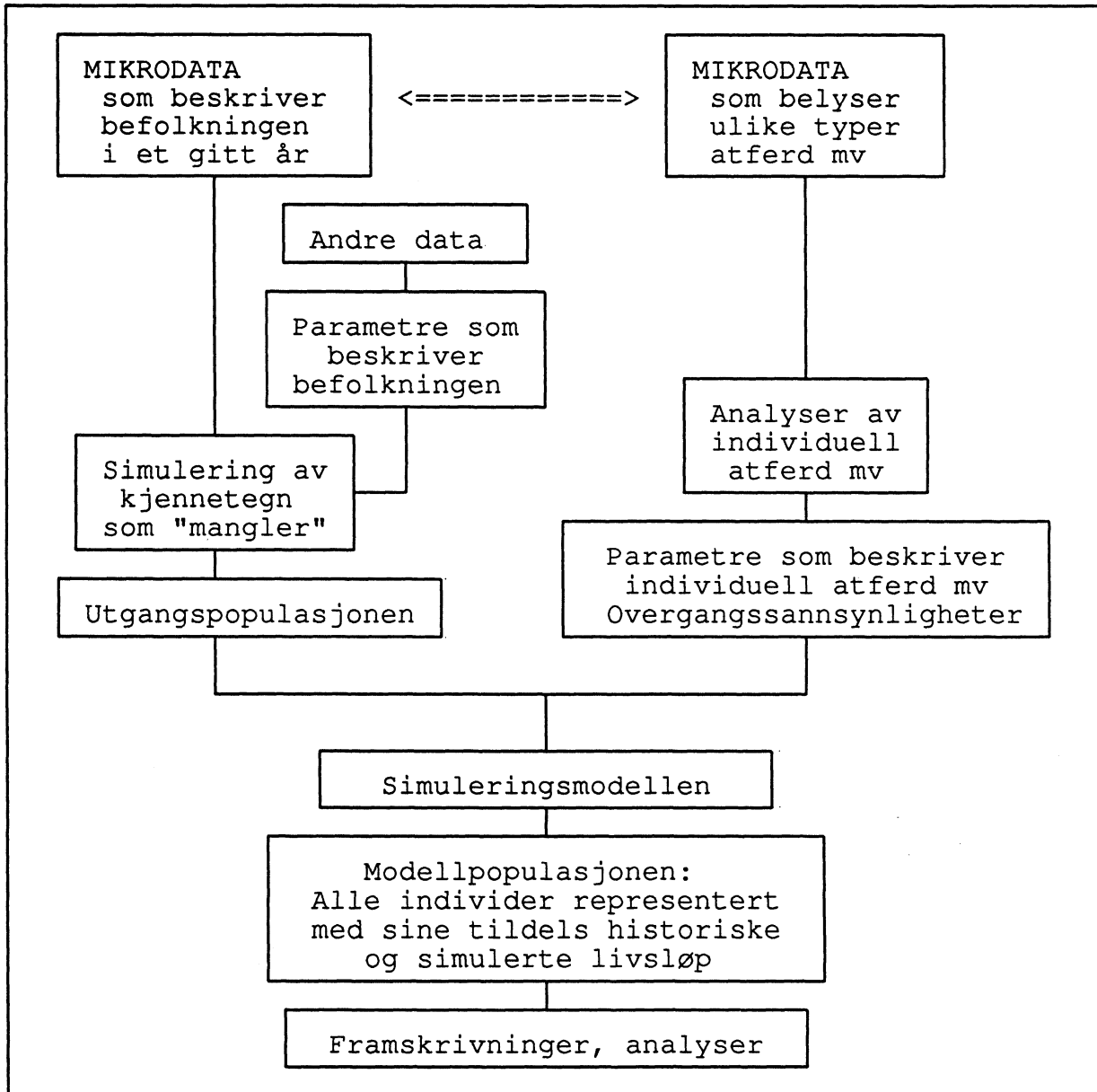
Hvert kjennetegn ved individene blir simulert separat og i bestemte rekkefølger. Det er også slik at hver gruppe av kjennetegn simuleres for hele modellpopulasjonen for et år, før man går videre til neste gruppe kjennetegn for samme år. I første gruppe simuleres netto-utvandring avhengig av kjønn, alder og ekteskapsstatus. I andre gruppe simuleres dødelighet for menn avhengig av kjønn og alder. Eventuell ektefelle gjøres til enke hvis han dør. I tredje gruppe simuleres for kvinner død, fødsler og ekteskapsbevegelser. Dødelighet simuleres også her avhengig av kjønn og alder, og ektefellen gjøres til enkemann om hun dør. Om hun føder et barn i løpet av året simuleres avhengig av hennens egen alder, antall barn hun har og varighet siden siste fødsel. Er hun gift, trekker modellen om hun skiller seg, og hvis det skjer gjøres også den faktiske mannen til fraskilt. Er hun ikke gift², simulerer modellen om hun gifter seg avhengig av hennes egen alder, om hun nylig har født ett barn og ekteskapsstatus (ugift/enke/skilt). Blir hun gift, trekkes en alder for ektefellen avhengig av hennes egen alder, og så finner modellen en tilfeldig ikke gift mann med denne alderen. I så måte er modellen kvinnedominert, men vi oppfatter det som en god forenkling av tokjønns-problemet i denne type modeller.

I fjerde gruppe simuleres skolegang for hvert individ. Dvs at modellen simulerer igangværende utdanninger for hvert individ, og hvilke utdanningsvalg som fattes i forhold til disse. I den grad personen (velger) å fullføre sin igangværende utdanning, vil dette føre til en økning i personens utdanningskapital, den såkalte høyeste fullførte utdanning. Utdanning er representert ved omlag 23 ulike "fagfelt" fordelt over flere nivåer. Dette representerer en sterk aggregering i forhold til den individbaserte utdanningsstatistikken, men vil likevel fange opp mange trekk ved utdanningssystemet. I siste gruppe simuleres

² Vi har også lagt inn en beskrankning at ekteskapsstatus bare kan endres en gang pr år, dvs at personer som har blitt enke/enkemann eller fraskilt, ikke kan gifte seg igjen før året etter.

arbeidstilbud for hvert individ. Hvis personen var i arbeidsstyrken ifor, trekker modellen om personen fortsatt skal være i arbeidsstyrken avhengig av kjønn. Var personen ikke i arbeidsstyrken, eller foretok en vesentlig endring på et annet område, f.eks. en fødsel for kvinner, så trekker modellen om personen skal være i arbeidsstyrken avhengig av kjønn, alder, ekteskapeleg status, om personen er under utdanning og høyeste fullførte utdanning. Disse overgangene inn i arbeidsstyrken er endogent bestemt slik at yrkesprosenten for hver slik gruppe holdes konstant lik det de er i AKU for et gitt basisår. For de som er i arbeidsstyrken trekker modellen et arbeidstilbud i timeverk pr uke avhengig av de samme kjenne-tegnene, også basert på AKU-tall.

Skisse av MOSART.



Modellpopulasjonen.

Resultatet av denne simuleringen blir en modellpopulasjon av individer med sine tildels historiske og simulerte livsforløp. Disse er beskrevet ved tre datasett for ulike typer enheter i modellen. Det første datasettet omfatter alle individer med status for hvert år fra og med

det året personen fyller 15 år³, og ut simuleringsperioden. Skulle personen dø eller utvandre i simuleringsperioden, blir personen beskrevet til og med det året personen dør eller utvandrer. Hver observasjon inneholder årstall og personens id.nummer⁴, samt kjønn, alder, ekteskapeleg status og eventuell ektefelles id.nummer, høyeste fullførte utdanning, igangværende utdanning, tilknytning til arbeidsstyrken og eventuelt arbeidstilbud i timeverk. I tillegg kommer noen indikatorer som angir endringer i ekteskapeleg status, utdanning og arbeidstilbud, og en indikator som angir om personen lever og er bosatt i landet ved utgangen av angitte år. For kvinner er det også med antall barn hun har født, samt alder på eldste og tre yngste barn.

De to andre datasettene er mindre og beskriver relasjoner mellom individene. Det andre datasettet omfatter alle ekteskap med opplysninger om når ekteskapet ble inngått, mannens og kvinnens id.nummer, og eventuelt når ekteskapet ble oppløst og årsaken til oppløsningen. Det tredje datasettet omfatter alle fødsler, med opplysninger om når fødselen fant sted og morens id.nummer. Senere vil vi få med barnets og farens id.nummer.

Bruksområder.

Den viktigste bruken av modellpopulasjonen vil være framskrivninger av befolkningen etter sentrale sosiodemografiske og sosioøkonomiske kjennetegn. Fordi modellpopulasjonen bygger på et tilfeldig utvalg kan man trekke slutninger om befolkningen på grunnlag av denne. Spissformulert vil modellpopulasjonen være en utvalgsundersøkelse av befolkningen i framtiden, gitt at befolkningen utvikler seg slik modellen forutsetter. I tillegg kan også modellen brukes til fordelingsanalyser da modellen bygger på individer som enhet.

Det andre viktige bruksområdet vil være en analyse av de overgangssannsynligheter vi selv legger inn i modellen. Disse overgangssannsynlighetene vil normalt være resultatet av (atferds)analyser hvor det ikke alltid er lett å få innsikt i hva estimatene innebærer. Modellen kan da si mye om konsekvensene av estimer fra disse atferdsanalysene, nettopp fordi den er egnet til å oppsummere utviklingen for hele befolkningen, eller grupper av befolkningen.

3. Videre arbeid på modellen.

I utviklingen av MOSART startet vi med modellen og de minimumskrav vi ville stille til den. Utfra dette bestemte vi hva som ville være et minste krav til data og atferdsanalyser til bruk i modellen. Spesielt vil de fleste overgangssannsynlighetene vi har hentet inn i modellen være basert på tidligere (atferds)analyser i SSB, og ofte ha et partielt preg i forhold til det vi kunne ønske oss. Ved å bruke allerede utførte analyser var det mulig å holde en viss framdrift i arbeidet, men samtidig vil modellen påpeke klare svakheter ved eksisterende (mikro)data og atferdsanalyser. Disse svakhetene ved modellen oppfatter vi som et arbeidsprogram for å skaffe tilveie bedre mikrodata og mer helhetlige atferdsanalyser.

Den første utvidelsen av modellen vil være å gjøre MOSART til en selvstendig demografisk modell. Dette vil skje ved at de fødsler som blir simulert i modellen, også vil bestemme befolkningstilveksten i modellen. Slik det er nå henter vi disse dataene fra SSB's framskrivningsmodell for befolkningen (BEFREG), og denne endringen vil bli vurdert på grunnlag av erfaringer med modellen. Ved endogen rekruttering vil trolig personene bli representert fra de blir født, slik at MOSART da vil beskrive hele befolkningen. Dette vil bare være første del av et større prosjekt hvor husholdning vil bli innført som enhet i modellen, og vil måtte omfatte en rekke nye atferdsanalyser på det demografiske området.

³ For innvandrere er det fra og med året før de innvandrer.

⁴ Dette er fullstendig kryptifisert i forhold til personnummeret.

Den andre utvidelsen av modellen vil være å utvide kjennetegnet for arbeidstilbud til å omfatte trygdestatus og arbeidsinntekt. Disse to egenskapene er nært knyttet til arbeidstilbudet, og vil sammen med den demografiske utvikling bestemme utbetalinger til alders- og uføretrygd fra Folketrygden. Hvis MOSART blir egnet som en trygdemodell, så vil det medføre at MOSART også må bli vesentlig bedre som en individbasert bruttostrømsmodell for arbeidstilbud. En trygdemodell vil da også kreve en mer omfattende beskrivelse av utgangspopulasjonen og bedre analyser av arbeidsmarkedsatferd enn i førsteutgaven av MOSART. Den første byggestenen i dette arbeidet vil være å skaffe tilveie mer sammensatte mikrodata, deriblant bruttostrømsstatistikk for arbeidsmarkedet. Dette databehovet vil bli drøftet i det neste kapitlet.

4. Databehov for MOSART - trygdemodellen.

Figuren på side 5 gir en oversikt over de databehov som generelt må dekkes i MOSART. Først og fremst trenger man gode mikrodata som kan belyse de atferdsmønstre man ønsker å bruke i modellen. Dernest trenger man mikrodata som beskriver befolkningen i et gitt startår hvis man skal bruke modellverktøyet som en planleggingsmodell. Det vil ofte være vanskelig å skaffe slike data som beskriver alle sider ved utgangspopulasjonen. Det er da nødvendig å syntetisere de data man mangler, og parametre for dette kan ofte hentes fra data som belyser atferdsmønstre brukt i modellen. Data til bruk i atferdsanalyser og til beskrivelse av befolkningen behøver ikke komme fra samme kilde, og det har vi benyttet i førsteutgaven av modellen. Derimot ser vi en rekke grunner til å skaffe tilveie en mer sammensatt og longitudinell database med individhistorier, som kan dekke begge områdene.

Hvorfor individhistorier i databasen.

Dette vil være viktig for utgangspopulasjonen, spesielt i forhold til en trygdemodell. Mange av de forpliktelser som Folketrygden framover vil få er allerede opptjent av kommende pensjonister. Vi ser på MOSART som en planleggingsmodell for Norge, og det vil da være viktig å få med forhistoriene, slik at disse allerede opptjente rettighetene blir beskrevet.

Tilleggstrygden i Folketrygden (ATP) vil være en tildels komplisert funksjon av arbeidsinntekten for hvert enkelt år i yrkeskarrieren. Det gjør det viktig å få vite ikke bare forventet antall yrkesaktive år, men også fordelingen av disse utover befolkningen. Vi må derfor se på i hvilken grad inn- og utstrømning på arbeidsmarkedet er knyttet til noen bestemte individer, også etter at man har kontrollert for kjennetegn som påvirker yrkesdeltakingen. Det er da nødvendig å kunne følge hvert individ over et lengre tidsrom, og aller helst over hele livsløpet så langt det er kommet.

Hvorfor en longitudinell database.

Med longitudinell database vil vi her mene at utvalget i databasen er representativ for befolkningen over et lengre tidsrom. Dvs om databasen tar hensyn til at nye personer kommer til i befolkningen (fødsler, innvandring) og at noen forsvinner ut (død, utvandring). Dette vil gjøre det mulig å bruke databasen/modellen til å se på utviklingen i en historisk periode forut for simuleringsperioden.

Av større betydning er muligheten for å se på utviklingen over tid i de atferdsparametre vi bruker i modellen. For å kunne gjøre dette må databasen være longitudinell. Et typisk eksempel kan være overgangen til uførhet. Hvis databasen utelater de som har dødd i perioden man analyserer over, vil man trolig underestimere overgangen til uførhet. Dette vil skyldes at de som blir uføretrygdet (trolig) har høyere dødelighet, og dermed i mindre grad vil være representert i dagens befolkning.

Det vil være flere momenter som taler for betydningen av å se atferden over tid. En viktig årsak er at økonomiske konjunkturer kan gi opphav til store avvik i atferd for et

enkelt år. Et eksempel kan være hvis yrkesatferden er beskrevet ved bruttostrømmer på arbeidsmarkedet. Da kan selv små endringer i beholdningen av yrkesaktive i et år, pga et konjunkturomslag, gi opphav til meget sterke endringer i strømmene på arbeidsmarkedet det året. Spesielt vil dette være tilfelle hvis disse strømmene er små i utgangspunktet, dvs at beholdningen av yrkesaktive i sterk grad består av de samme personene fra et år til det neste.

Hvorfor en sammensatt database.

Et problem ved å bruke ulike datakilder er at de ofte har ulike definisjoner for et eller flere kjennetegn. Et eksempel kan være at MOSART er basert på registerstatistikken for utdanning, mens AKU i Norge er basert på intervju om personens utdanningsnivå. Når vi så brukte AKU som estimatorer for arbeidstilbudet i MOSART, ga forskjellene i utdanning en samlet yrkesprosent i MOSART som var 1,2 prosentpoeng lavere enn i AKU.

Det vil også være ønskelig å kunne kontrollere i hvilken grad atferden på ulike områder er avhengig av andre variabler i modellen. Dette er generelt ikke mulig med mindre man har en database som omfatter alle de variabler og enheter som er inkludert i modellen. Uten dette vil også en kritisk gjennomgang av modellens antatte uavhengigheter måtte basere seg på intuisjon, og ikke fakta.

5. Egne dataprojekter.

Vi skulle gjerne hatt en "IDA" for norske forhold, men det finnes såvidt vi vet ingen tilsvarende prosjekter i Norge på dette området. En database som omfattet hele befolkningen over en tiårs-periode med et bredt spekter av kjennetegn ville gitt et vesentlig bedre grunnlag for videreutvikling av en modell som MOSART. Derfor har vi startet et dataprojekt for modellen som vil gi noen av disse egenskapene vi etterlyser.

Vi har valgt å ta utgangspunkt i registerbasert statistikk, da vi ser dette som den eneste realistiske muligheten for å skaffe tilveie tilstrekkelige store utvalg som dekker befolkningen over et lengre tidsrom. På områder hvor denne typen statistikk står svakt vil vi supplere med opplysninger fra utvalgsundersøkelser. Spesielt vil dette gjelde forhold som arbeidstid representert ved timeverk og opplysninger om husholdningstilknytning. De fleste utvalgsundersøkelsene vil ofte ha uegnede utvalg for vårt formål, enten fordi man har trukket veid eller fordi man har utelatt deler av befolkningen som der og da er ansett som uinteressante. Et typisk eksempel kan være AKU som kun har med befolkningen fra 16 til 74 år.

Databasen for modellen skal dekke perioden 1967 og opp til og med siste tilgjengelige år med registerstatistikk. I 1967 ble Folketrygden innført i Norge, og det er fra og med dette året det er mulig og interessant å hente inn opplysninger om arbeidsinntekt. Databasen vil ta utgangspunkt i et utvalg av befolkningen på omlag 5-10 prosent. Grunnen til dette er at vi hadde store EDB-messige problemer med å koble hele kvinnefilen og utdanningsfilen. Det prosjektet vi går inn i nå vil omfatte vesentlig større datamengder, og vi ønsker å få en viss erfaring med dette, før vi eventuelt går løs på hele befolkningen.

Når databasen blir basert på et utvalg vil det gi et problem i forhold til bruk i MOSART. I modellen ønsker vi at viktige tilknytninger mellom individer skal bli ivaretatt, samtidig som hvert individ i befolkningen skal ha samme sannsynlighet for å komme med i utvalget. Dette kravet er vanskelig å imøtekomme, spesielt når man ser på befolkningen over tid. Problemet skyldes at tilknytninger mellom individer, ofte er av en slik art at alle individer

indirekte er tilknyttet alle andre individer⁵. Vi har valgt å begrense oss til ekteskap for startåret for modellen, og med få unntak utelatt alle andre tilknytninger mellom individer. En uheldig bivirkning vil bli problemene med å oppdatere databasen. Noen personer i utvalget vil med all sikkerhet komme til å gifte seg med personer som ikke befinner seg i det utvalget vi har trukket. Konsekvensen av dette er at vi må gjennomgå hele prosedyren hver gang vi skal oppdatere datagrunnlaget.

Data vil vi hente inn fra tre hovedkilder. Rikstrygdeverket, som administrerer Folketrygden i Norge, har registre med hele befolkningen på individnivå, med årlige opplysninger om pensjonsgivende inntekt og i hvilken grad de har vært alders- og uføretrygdet. Den pensjonsgivende inntekten vil være inntekter knyttet til arbeid, og denne har vi tenkt å benytte som en indikator for tilknytning til arbeidsstyrken. Fra utdanningsfilene vil det være mulig å hente ut utdanningsstatus for hvert år tilbake til begynnelsen av 70-tallet. Fra filer basert på personregisteret vil vi hente ut opplysninger om fødeår og kjønn (personnummeret), årstall for eventuell migrasjon eller dødsfall, fødsels-historier og ekteskapshistorier.

Denne databasen vil ikke dekke alle de krav MOSART stiller, arbeidstid er nevnt. Et annet problem er husholdningstilknytning mellom individer i modellpopulasjonen. Dette vil trolig måtte syntetiseres basert på estimer hentet fra andre undersøkelser. Det ville også vært ønskelig med data hentet fra skatteligningen, men det er begrenset hva som foreløpig ligger på maskinlesbare media. Det vi kunne ønsket å få inn i modellen/databasen hadde vært aggregerte årsdata for kapitalinntekt i tillegg til arbeidsinntekt, samt skatt og formue, og dermed sparing og konsum.

⁵ Kravet er at befolkningen må kunne deles opp i lukkede grupper, som er slik at hvert individ tilhører en og bare en gruppe, og at ingen individer får tilknytninger til individer utenfor sin egen gruppe. Grupper bestående av "hvert enkelt individ", eller "hele befolkningen bosatt i landet", tilfredsstill disse kravene. En annen mulighet er å ta utgangspunkt i "husholdning" eller "ekteskap", for ett bestemt tidspunkt, fortrinnsvis utgangen av startåret for modellen.

FORPROSJEKTRAPPORT PÅ "MOSART OG TRYGD":

Med hovedvekt på datagrunnlaget for
utgangspopulasjonen og atferdsanalysene.

Av Dennis Fredriksen.

Sammendrag.

Notatet plasserer prosjektet "MOSART og trygd" i forhold til modellen MOSART-1, og gir en kort beskrivelse av en utvidet modell MOSART-T som omfatter alders- og uføretrygd fra Folketrygden. En vesentlig del av dette prosjektet vil være å koble Rikstrygdeverkets register for opptjente pensjonsrettigheter mot SSB's registerdata for demografi og utdanning. Denne koblingen vil bli beskrevet større detalj, da den er viktig for framdriften i prosjektet. Arbeidet med estimering av overgangssannsynligheter og simuleringsmodellen blir beskrevet så langt dette er avklart når forprosjektrapporten avsluttes. Tilslutt drøftes forholdet til andre prosjekter rundt MOSART, og til andre prosjekter og modeller som har tilknytning til området.

Innholdsfortegnelse.

1. Innledning.	10
2. Trygdemodellen MOSART-T	10
2.1. Utgangspopulasjonen.	12
2.2. Simuleringsmodellen.	12
2.3. Modellpopulasjonen.	14
3. Datagrunnlaget for modellen	15
3.1. Trekking av utvalget	15
3.2. Innhenting av data	17
3.3. Dataene organiseres	18
3.4. Dokumentasjon	18
4. Andre sider ved modellen	19
4.1. Estimeringsarbeider.	19
4.2. Arbeid som må utføres på den nye modellen.	20
4.3. Forholdet til andre tilstøtende prosjekter.	21
Vedlegg.	23
(a) Tids- og ressursplan.	24
(b) Tabeller.	25
(c) Litteratur.	26

1. Innledning.

Høsten 1988 startet arbeidet med å utvikle en ny modell for framskrivinger av utdanning og arbeidsstyrke. Den nye modelltypen fikk navnet MOSART, et akronym for "MOdell for mikrosimulering av Skolegang og ARbeidsTilbud". Førsteutgaven av modellen som ble operativ høsten 1990, fikk navnet MOSART-1, og erstattet samtidig utdanningsmodellen MONS og arbeidsstyrkemodellen MATAUK. En foreløpig dokumentasjon av MOSART-1 finnes i Andreassen og Fredriksen (1991), og tabeller fra de første kjøringene er gjengitt i Statistisk ukehefte 91/1.

MOSART-1 tar utgangspunkt i et utvalg av befolkningen og simulerer det videre livsløpet hvert individ får. Begivenheter som simuleres er inn- og utvandring, dødelighet, fødsler, bevegelser i ekteskapelig status, skolegang og arbeidstilbud. Nye årskull av 16-åringer legges til etterhvert, og resultatet av simuleringen blir en modellpopulasjon av individer med sine livshistorier. Når modellpopulasjonen er representativ for befolkningen, vil man gitt de forutsetninger som bestemmer utviklingen i hvert enkelt livsløp, kunne gi framskrivinger av blant annet arbeidsstyrken og befolkningens utdanningsnivå. I tilknytning til MOSART-1 gjenstår en del arbeid med publisering av resultater og med simuleringer under andre forutsetninger enn de som vi har valgt som referansealternativ.

To større prosjekter som skal videreutvikle MOSART startet opp mot slutten av 1990. Det ene prosjektet omfatter utviklingen av en husholdningsmodell MOSART-H basert på samme type modellverktøy. Modellen skal holde rede på hvilke individer som bor sammen, og skal også simulere flyttinger mellom husholdninger. Videre skal tilveksten av nye årskull bestemmes av de fødsler som blir simulert i modellen, såkalt endogen rekruttering, slik at husholdningsmodellen også blir en fullstendig befolkningsmodell.

Det andre prosjektet skal utvikle en modell for alders- og uføretrygd fra Folketrygden. Her er det nær sammenheng mellom befolkningsutviklingen og individuelle yrkeshistorier på den ene siden, og alders- og uføretrygd på den andre siden. Vi tror derfor at en mikro-simuleringsmodell basert på MOSART vil være velegnet til å fange opp disse problemstillingene. De individuelle opplysninger som bestemmer størrelsen på alders- og uføretrygd, er tilgjengelig for analyseformål og vil gi prosjektet et godt datagrunnlag. Denne trygdemodellen har fått arbeidsnavnet MOSART-T, og vil erstatte alderstrygdemodellen MAFO. Trolig vil MOSART-T også bli modellen for utdanning og arbeidsstyrke, og dermed ta opp i seg de størrelser MOSART-1 framskriver idag. I den grad vi finner det hensiktsmessig vil trygdemodellen og husholdningsmodellen bli samlet i en modell på et senere tidspunkt.

2. Trygdemodellen MOSART-T.

MOSART-T vil være en modell som bygger på simulering av livsløpene til en modellpopulasjon av individer. De individuelle livsløpene vil være beskrevet ved sentrale kjennetegn innen demografi, utdanning og tilpasning på arbeidsmarkedet. Livsløpene vil være beskrevet ved historiske data til og med startåret for simuleringen (1989), og deretter ved simulerte data for perioden 1990-2050. Når modellpopulasjonen er representativ for befolkningen i Norge, vil dette danne et grunnlag for framskrivinger av alders- og uføretrygd fra Folketrygden. Utbetalinger til disse typene trygd er bestemt av befolkningsutviklingen og av trekk ved de individuelle yrkeshistoriene. Motsatt vil alders- og uføretrygd ha stor betydning for arbeidstilbudet i de eldre årsklassene.

Den modellbeskrivelsen som følger nedenfor tar utgangspunkt i det som allerede finnes på MOSART. Simuleringen av de demografiske kjennetegnene er noe omgruppert og det er forutsatt endogen rekruttering. I tillegg er overgangene i forhold til trygdesystemet og arbeidsmarkedet beskrevet slik de trolig vil bli i trygdemodellen. Utviklingen av en

husholdningsmodell kan gi ytterligere endringer i den modellbeskrivelsen som følger.

Trykdemodellen MOSART-T.

DATAGRUNNLAG

Utvalget: 10 prosent av befolkningen i Norge i perioden 1967-1989.

Data

Kvinnefilen: Fødselshistorier.

Utdanningsfilene (BHU): Utdanningsstatus for perioden (1971)-1989.

Rikstrygdeverket

Trygdefilene: Trygdestatus for de senere årene.

Pensjonspoengregistret: Pensjonsgivende inntekt, pensjonspoeng, beregnet uførepoeng (BUP) og uføregrad.

Datagrunnlag for utgangspopulasjon og atferdsanalyser.

Koblet fil med 10 prosent av befolkningen i perioden 1967-1989, med individhistorier innen ekteskap, fødsler, utdanning, pensjonsgivende inntekt og tilknytning til trygdesystemet.

Utvalg.

SIMULERINGSMODELL

For hvert år i perioden 1990-2050.

For alle personer: Simuler inn- og utvandring og dødelighet.

For kvinner: Simuler fødsler og bevegelser i ekteskapeleg status.

For alle personer (ektefeller sammen):
Simuler skolegang, trygdestatus og pensjonsgivende inntekt.

RESULTAT

Modellpopulasjon som omfatter 1-4 prosent av befolkningen i Norge i perioden 1967-2050, med livshistorier som dekker ekteskap, fødsler, utdanning, yrkesliv og trygd.

Tabellprogrammer, inklusive rutiner for beregning av alders- og uføretrygd fra Folketrygden.

2.1. Utgangspopulasjonen.

Modellen vil ta utgangspunkt i et tilfeldig utvalg av befolkningen i Norge for perioden 1967-1989. Fra registerstatistikk vil vi hente inn opplysninger for kjennetegn innen demografi, utdanning, arbeidstilbud og trygd. Den delen av utvalget som er bosatt i Norge ved utgangen av startåret, vil danne utgangspopulasjonen i simuleringsmodellen.

For hver person vil vi få med kjønn, fødeår, samt eventuelle årstall for inn- og utvandring og dødsfall. For utgangen av hvert år vil vi tilnærmet kjenne ekteskapelig status, ektefelle og reg.status, det vil si om personen lever og er bosatt i Norge. For ektepar som var gift/separert ved utgangen av 1989, vil begge ektefeller være representert i utgangspopulasjonen. Mer utfyllende husholdningsdata vil være et savn. Disse kjennetegnene kan som en nest-best løsning syntetiseres, det vil si at modellen trekker hvilke husholdninger hver persons skal tilhøre også i startåret.

For hver person vil vi få med høyeste fullførte utdanning (HFU) og igangværende utdanning (IGU) pr 1/10 for hvert år, muligens fra og med 1974. For hver person vil vi for hvert år fra og med 1967 få med opptjente pensjonspoeng og beregnede uførepoeng. Disse allerede opptjente pensjonsrettighetene vil i stor grad bestemme utbetalingene til alders- og uføretrygd i mange år framover. Samtidig vil pensjonspoeng være entydig bestemt av pensjongivende inntekt, som er arbeidsrelaterte inntekter, og vil dermed også tjene som indikatorer på yrkesdeltakingen i disse årene.

2.2. Simuleringsmodellen.

Modellen vil simulere de videre livsløpene for hvert individ i utgangspopulasjonen og for de individer som tilføres modellpopulasjonen i form av innvandring og nye generasjoner. Simuleringen foregår ved at modellen på grunnlag av såkalte overgangssannsynligheter trekker hvilken utvikling hvert individ vil få fra år til år. Modellen er da selvsagt ikke utsagnskraftig om det enkelte individ, men kun for tilstrekkelig store grupper av befolkningen. De overgangssannsynlighetene som brukes vil avhenge av kjennetegn ved individet selv, og kan ofte tolkes som en beskrivelse av atferd. Normalt vil disse sammenhengene være estimert på grunnlag av observert atferd i en gitt periode forut for startåret for simuleringen. I tillegg vil det være nødvendig å legge inn forutsetninger om utviklingen i disse atferdsparametrene. En første tilnærming er å anta samme atferd som i estimeringsperioden. Dette er selvsagt ikke realistisk, men kan likevel gi mye innsikt i de prosesser modellen tar opp.

MOSART-T simulerer livsløpene for alle individene tilnærmet parallelt, ved at modellen simulerer hele modellpopulasjonen for et år før den går videre til neste år. Det gjør det mulig å ha tilknytninger mellom individer i modellen, for eksempel ekteskap. Egenskaper ved modellpopulasjonen for året før kan også brukes som forklaringsvariabler for modellens overgangssannsynligheter, for eksempel antall personer i arbeidsstyrken. Alle kjennetegn blir simulert enkeltvis og i bestemt rekkefølge, og dette forhindrer en simultan simulering av gjensidig avhengige kjennetegn. Modellen er her begrenset til å betinge utviklingen i hvert kjennetegn med hensyn på hva som allerede har skjedd med de andre kjennetegnene. Med kalenderåret som tidsenhet vil dette utgjøre et mindre problem. Hvis man i tillegg tar hensyn til dette i estimeringsarbeidet, vil problemet være helt marginalt. Hver del av simuleringen blir også gjennomført for hele befolkningen, før modellen går videre til neste del.

I den første delen av modellen simuleres inn- og utvandring og dødelighet. Migrasjon simuleres med fast nivå på netto-innvandring, og fordeles etter kjønn, alder og ekteskapelig status. Dødelighet simuleres avhengig av kjønn og alder, og eventuell ektefelle gjøres til enke/enkemann hvis personen dør. I den andre delen av modellen simuleres fødsler og bevegelser i ekteskapelig status for kvinner. Fødsler simuleres avhengig av kvinnens alder, antall barn og yngste barns alder. Føder hun et barn, blir dette barnet opprettet som en ny

person i modellpopulasjonen. Er kvinnen gift, simulerer modellen om hun skiller seg, og hvis skilsmisse inntreffer blir også mannen skilt. Er hun ikke gift, simulerer modellen om hun gifter seg, avhengig av hennes egen alder, om hun nylig har født et barn og ekteskapelig status (ugift/skilt/enke). Blir hun gift, trekker modellen en alder for ektefellen avhengig av hennes egen alder, og finner deretter en tilfeldig valgt ikke gift mann med denne alderen. Det vil ligge som en skranke at menn og kvinner kun kan endre ekteskapelig status en gang i året. Deromot kan det inntreffe flere ulike typer begivenheter, for eksempel fødsel og inngåelse av ekteskap i samme år. Begivenhetene fødsler og bevegelser i ekteskapelig status er også kvinnedominert i modellen MOSART.

I den tredje delen av modellen simuleres overganger i forhold til utdanningssystemet, arbeidsmarkedet og trygdesystemet. Den følgende beskrivelsen er en foreløpig skisse av disse overgangene. Det endelige opplegget vil kunne bli annerledes, og trolig mer detaljert, selv om hovedtrekkene blir de samme. Ektefeller vil bli simulert sammen, slik at modellen kan ta hensyn til avhengigheter mellom deres tilpasning. Modellen vil kun ta opp overføringer fra Folketrygden. Personer som er førtidspensjonert vil i modellen være ikke (alders- og uføre)trygdede, og i den grad de ikke har bijobber, vil de også stå utenfor arbeidsstyrken.

Modellen simulerer utdanning med utgangspunkt i skolegang. Først simuleres om man skal fullføre eventuell igangværende utdanning fra skoleåret⁶ før, og om man skal begynne/fortsette å være under utdanning i det nye skoleåret. I den grad personen (velger) å fullføre sin igangværende utdanning, vil dette føre til en økning i personens utdanningskapital, den såkalte høyeste fullførte utdanning. For de som begynner/fortsetter å være under utdanning, simulerer også modellen valget av (ny) igangværende utdanning. Forklaringsvariable for utdanningsatferden er kjønn, alder, høyeste fullførte utdanning og igangværende utdanning fra skoleåret før. Utdanning er beskrevet ved fagfelt og nivå på utdanningen. Det er skilt mellom grunnskole, videregående skole og høyere utdanning. Helsefag, tekniske fag og økonomisk-administrative fag er gjennomgående skilt ut som egne fag på alle nivåer.

Overganger i forhold til trygdesystemet vil bli beskrevet ved to variable, en for trygdestatus og en for trygdegrad. I et enkelt opplegg kan trygdestatus anta verdiene alderstrygdet, uføretrygdet, samt ikke alders- og uføretrygdet. Trygdegrad vil spesielt for uføretrygdede angi uføregraden. Overgangen til alderstrygd vil som en enkel foreløpig løsning, bli simulert ved at alle som er 67 år og eldre blir satt til å være alderspensjonister. Videre vil alle de som har blitt uføretrygdet, også som en enkel foreløpig løsning, fortsette å være uføretrygdet til de blir alderspensjonister. Trygdestatistikk for 1989 viser at dette kun vil gi små avvik fra den faktiske situasjonen. Overgangen til uførhet vil derimot bli simulert, og denne overgangen vil gjøres avhengig av kjønn, alder, yrkesdeltaking og andre variable som er tilgjengelige i datamaterialet og som påvirker overgangen til uførhet. For uføretrygdede er det også nødvendig å simulere eventuell økning i uføregrad.

Ved simulering av arbeidsstyrkestatus har vi et valg mellom to løsninger. Det mest ambisiøse vil være å simulere en konsumenttilpasning som simultant gir arbeidstilbud og timelønn. Slike modeller finnes i Statistisk sentralbyrå, hvor de første versjonene ble utviklet under navnet GATO. En slik modell vil kreve opplysninger om formue og kapitalinntekter, og dette er kjennetegn som (ennå) mangler i utgangspopulasjonen. De arbeidstilbudsmodellene som hittil er utviklet vil også ha en del svakheter ved bruk i MOSART. De krever trolig for mye regnekraft og det vil også medføre et vesentlig dataarbeide for at de skal omfatte hele befolkningen (selvstendige, uføretrygdede). I denne versjonen av trygdemodellen vil vi derfor inntil videre satse på det andre alternativet.

Det andre alternativet består i å simulere pensjonsgivende inntekt direkte uten å gå veien om en fullstendig spesifisert atferdsmodell med timelønn og arbeidstilbud i timer. Vi vil da ta utgangspunkt i de pensjonspoenghistorier som Rikstrygdeverkets data gir. Her inngår kun produktet av arbeidstimer og timelønn, men til gjengjeld vil datagrunnlaget være meget godt

⁶) Utdanningsstatus pr 1/10 året før, tilnærmet skoleåret før.

når det gjelder livshistorier. Tilknytning til arbeidsstyrken vil i dette opplegget være definert som å ha en pensjonsgivende inntekt over en viss grense. I den grad det faller heldig ut, vil denne inntektsgrensen bli valgt slik at modellen treffer Arbeidskraftundersøkelsene (AKU) i startåret for simuleringen. Trygdestatistikk og AKU for 1988 viser at dette bør være mulig. I tillegg må inntektsgrensen for å være i arbeidsstyrken gjøres avhengig av tid for å fange opp vekst i reallønnen. Arbeidstilbud i timer vil da ikke bli simulert, men kan beregnes ved en enkel ettermodell i et regneark.

Pensjonsgivende inntekt (PGI) vil bli simulert i en bruttostrømsmodell basert på funksjonsuttrykk hvor man eksogent kan angi nivået på både arbeidsstyrken og reallønnen. Da estimeringsopplegget for PGI langt fra er klart, vil det følgende bare være en foreløpig skisse av modellen. Først vil modellen simulere om man har PGI, og her vil det inngå (en) eksogen periodevariabel som påvirker sannsynlighetene for å ha PGI. I estimeringen vil denne eksogene variabelen inngå som en serie dummyvariable for hvert år. Derneft vil modellen simulere størrelsen på PGI, og veksten i denne vil være bestemt av en eksogent gitt reallønnsvekst og et individuelt tillegg som for befolkningen samlet har forventningsverdi (tilnærmet) lik null. Referansebanen i dette opplegget vil være å hente en "troverdige" reallønnsbane fra en makroøkonomisk modell og sette periodevariabelen til et observert gjennomsnittsnivå for 1980-tallet.

Formålet med denne "troverdige" reallønnsbanen er ikke å inkludere makroøkonomiske virkninger i MOSART, men å få en rimelig referansebane. Med makromodeller hvor arbeidsstyrken og reallønnen er endogent bestemt, måtte begge disse ha vært bestemt av makromodellen for å sikre full konsistens. Dette innskrenker MOSART såvidt mye, at det kun vil være aktuelt i alternative simuleringer.

Sannsynlighetene for å ha PGI og størrelsen på PGI vil også avhenge av individuelle forklaringsvariable, og datamaterialet gir gode valgmuligheter. Viktige variable som kan nevnes er fjorårets yrkesdeltaking, kjønn, alder, ekteskapeleg status/ektefelle, om man er under utdanning, utdanningsnivå og trygdestatus. I tillegg vil vi prøve å trekke inn tidligere trekk ved yrkeshistorien med utgangspunkt i de pensjonspoenghistorier vi har hentet inn for de samme individer. For kvinner vil det også være aktuelt å bruke trekk ved fødsels-historien, for eksempel yngste barns alder og antall barn.

En slik bruttostrømsmodell kan gjøre det vanskelig å forstå hva som bestemmer veksten i arbeidsstyrken i modellen. En tilnærming vil da være å dekomponere veksten med utgangspunkt i konstante yrkesprosenter etter kjønn, alder og andre interessante variable. Det vil da framgå hvor mye som skyldes utviklingen i befolkningens størrelse og sammensetning etter kjønn, alder og de andre variablene. I tillegg vil det gjenstå noe som ikke kan forklares med endringer i befolkningens størrelse og sammensetning, og dette vil være knyttet til bruttostrømsmodellen. Trolig vil nivået på denne uforklarte delen være av moderat størrelse for befolkningen samlet.

2.3. Modellpopulasjonen.

Denne simuleringen skal gi en modellpopulasjon som er representativ for befolkningen for perioden 1967-2050. For de fleste kjennetegn vil det da være mulig å ta ut historiske tids-serier som er direkte sammenlignbare med de utviklingsbanene simuleringmodellen gir. Viktige størrelser som vil beregnes er blant annet tallet på alders- og uføretrygdede og nivået på utbetalingene til alders- og uføretrygd fra Folketrygden. Andre overføringer fra Folketrygden som er mulige å beregne, er for eksempel etterlattetrygd. Siden modellen bygger på et representativt utvalg, kan MOSART brukes til provenyberegninger av alders- og uføretrygd, gitt forutsetningene om reallønnsvekst og arbeidsstyrke.

Når alders- og uføretrygd først blir beregnet etter simuleringen, betyr det at alders- og uføretrygd bare i begrenset grad kan påvirke den individuelle tilpasningen. Dette valget er gjort, og skyldes at beregningene av disse størrelsene avhenger av svært mange kjennetegn fra hele livsløpet. Av den grunn er det for kostbart å beregne trygdeutbetalinger fortløpende

i simuleringen. Derimot skal det være enkelt å ta utgangspunkt i samme modellpopulasjon (simulering), og herfra beregne virkningene av å endre beregningsreglene for alders- og uføretrygd. Dette vil trolig bli en viktig måte å bruke modellen på, som både sparer regnekraft og reduserer den modellgenererte usikkerheten. Nesten alle de variable som bestemmer tilleggspensjonen i dag, samt mange andre variable som kan være aktuelle, inngår i datagrunnlaget. Modellen er derfor i stand til å analysere mange typer av endringer i beregningsreglene for alders- og uføretrygd.

3. Datagrunnlaget for modellen.

En vesentlig del av prosjektet vil være å koble Rikstrygdeverkets registre for opptjente pensjonsrettigheter mot SSB's registerdata for demografi og utdanning. De opptjente pensjonsrettighetene foreligger som pensjonsgivende inntekt, pensjonspoeng og beregnet uførepoeng, for hvert år fra og med 1967. Disse opplysningene er det faktiske grunnlaget for beregning av trygden for hver enkelt alders- og uføretrygdet. Det er derfor god grunn til å tro at dette registeret vil gi individhistorier med god datakvalitet.

Dette datagrunnlaget skal brukes både som utgangspopulasjon i simuleringmodellen og ved estimering av viktige overgangssannsynligheter knyttet til overganger på arbeidsmarkedet og til alders- og uføretrygd. Utfra de erfaringer vi har gjort med MOSART-1 bør gjennomføringen av slike dataarbeider tillegges stor vekt i planleggingen. Ved filkoblinger dukker det fort opp problemer som kan gi betydelige forsinkelser. Arbeidet er blant annet av disse grunner tenkt oppdelt i fire-fem deler, som nok kommer til å gå en del over i hverandre.

- (a) Forprosjekt.
- (b) Trekking av utvalg.
- (c) Innhenting av data.
- (d) Organisering av data.
- (e) Dokumentasjon.

Punkt (a) er en del av denne forprosjektrapporten, og skal kartlegge hvilke muligheter vi har, samt skissere en farbar vei fram til et ferdig dokumentert datagrunnlag. En stort problem med MOSART-1 har vært behandling av store datamengder, og da spesielt sortering av store datasett. Vi vil derfor først trekke det utvalg vi vil ha med i datagrunnlaget (b), og så samle inn registerdata fra andre datakilder om disse personene (c), som ikke allerede framkommer ved trekkingen av utvalget. Tilslutt vil det være viktig å gjøre datagrunnlaget tilgjengelig for atferdsanalyser og som basis for utgangspopulasjoner (d-e).

3.1. Trekking av utvalget.

Krav.

Hver enkelt del av utvalget skal foreligge som adskilte datasett hvor hver person er oppført med personnummer. Dette samles tilslutt i et stort datasett, hvor hver person er representert som en observasjon med personnummer og en indikatorvariabel for hvilken del(er) av utvalget denne personen kommer fra. Visse opplysninger bør allerede her ivaretas, deriblant ektefelles personnummer.

Første del av utvalget skal være et x-% tilfeldig utvalg av befolkningen pr 1/1-1990, uveid forhåndsstratifisert⁷ etter nødvendige variable; kjønn, ekteskapelig status og muligens alder. Ektefeller må trekkes sammen. Klarer vi å ivareta andre personbånd, for eksempel foreldre-barn, vil det være en fordel. Dette utvalget vil danne grunnlaget for utgangspopulasjoner i MOSART-T. Jeg har også den oppfatning at vi bør legge lite arbeid i å rette opp mindre

⁷ Det er viktig at hvert individ i utvalget kan tildeles samme vekt ved oppblåsning av tallene.

skjevheter, da forhåndsstratifiseringen kan ivareta mye av dette.

Det er grunn til å tro at uføretrygdede har en høyere dødelighet enn sine jevnaldrende. Hvis dette er riktig vil antall personer som ble uføretrygdet i perioden forut for startåret være underrepresentert i dagens befolkning. Man ville dermed underestimere overgangen til uførhet, men dette unngås ved å inkludere personer i utvalget som ikke lenger lever. Andre del av utvalget vil derfor være et tilsvarende $x\%$ tilfeldig utvalg av ikke-bosatte pr 1/1-1990, uveid forhåndsstratifisert etter årsak og årstall for at personen er ikke-bosatt. Dette vil omfatte personer som har utvandret eller dødd i perioden 1967-1989. Noen personbånd vil vi ikke forsøke å få inn her. Dette utvalget vil sammen med det forrige utvalget danne primærutvalget i datagrunnlaget. Disse personene vil blant annet være grunnlaget for estimeringsarbeidene.

For personer i primærutvalget vil vi forsøke å få med personnummeret for tilknyttede personer, deriblant tidligere ektefeller. Dette kan være viktig i forhold til pensjonsberegning for enker, og i forhold til estimering på ektefellers simultane tilpasning til utdanning og arbeidsmarked. Andre tilknyttede individer har vi foreløpig ikke tenkt å ha med.

Forslag.

Løsningen vil bli å ta utgangspunkt i statusfilen for befolkningen pr 1/1-1990, og eventuelt supplere denne med opplysninger som mangler fra andre kilder. Spesielt gjelder dette ektefelle av separerte kvinner som trolig må bestilles fra det Sentrale personregisteret.

(1) Statusfilen for befolkningen leses inn på fem sas-datasett som omfatter følgende:

IP1: Ikke gifte personer med personnummer, kjønn, fødedato, reg.status, dato for reg.status, ekteskapeleg status, dato for ekteskapeleg status.

GM1: Gifte menn med personnummer, fødedato, reg.status, dato for reg.status, ekteskapeleg status, dato for ekteskapeleg status og ektefelles personnummer.

GK1: Gifte kvinner med personnummer, fødedato, reg.status, dato for reg.status, ekteskapeleg status, dato for ekteskapeleg status og ektefelles personnummer.

SFEIL1: Et utdrag av GK1 som omfatter alle kvinner som mangler ektefelles personnummer.

SFEIL2: Alle personer hvor ekteskapeleg status mangler med opplysninger om personnummer, fødedato, reg.status, dato for reg.status, ekteskapeleg status, dato for ekteskapeleg status og ektefelles personnummer.

I tillegg telles opp alle personer hvor reg.status er 'ukurant'.

Fra alle filene tas det ut tabeller etter alder, kjønn, reg.status og ekteskapeleg status for å kontrollere at innlesningen har gått riktig. Tabellen fra SFEIL1 brukes til å vurdere om vi trenger å hente inn ektefelles personnummer fra personregisteret e.l..

(2) Et utdrag (personnumre) fra SFEIL2 sendes det Sentrale personregister med ønske om opplysninger om ekteskapeleg status, dato for ekteskapeleg status og ektefelles personnummer pr 1/1-1990. Dette kobles mot GK1, og de kvinner som da har ektefelles personnummer legges ut på datasettet GK2.

(3) GK2 sorteres på mannens personnummer og kobles mot datasettet GM1, og variablene som beholdes er kvinnens og mannens personnummer, reg.status, dato for reg.status, ekteskapeleg status og dato for ekteskapeleg status, samt en feilkode som angir inkonsistenser. Resultatet legges ut på to datasett som omfatter:

EP1: Ektepar som ifølge både GK2 er gift med hverandre, og hvor begge ektefeller er bosatt i Norge eller begge ektefeller er bosatt i utlandet.

EFEIL1: Resterende observasjoner.

EFEIL tabelleres for å kontrollere at mengden feil ikke er for stor eller konsentrert i visse

grupper av befolkningen.

(4) Datasettet IP1 og EP1 splittes på fire datasett som vil utgjøre trekkegrunnlaget for utvalget i datagrunnlaget. Disse filene vil kun inneholde personer som er 16 år og eldre.

BEP1: Ektepar hvor begge ektefeller er bosatt i Norge.

IEP1: Ektepar hvor begge ektefeller er bosatt i utlandet.

BIP1: Ikke gifte personer bosatt i Norge.

IIP1: Ikke gifte personer som er døde eller utvandret.

Fra disse fire datasettene tas det ut tabeller etter de kjennetegn det er aktuelt å forhåndsstratifisere etter, og disse tabellene sammenlignes med offisiell befolkningsstatistikk. Fra disse fire datasettene trekkes så fire delutvalg på 10 prosent av befolkningen. Disse delutvalgene samles tilslutt i et stort datasett med en observasjon for hver person. Dette vil omfatte primærutvalget. Personnummeret for tidligere ektefeller og for barn vil bli vurdert tatt med på grunnlag av erfaringer med trekkingen av primærutvalget og utfra signaler fra mulige brukere.

Personer som kommer med i trekkgrunnlaget.

Reg.status/ Ekteskapelig status	Skal med i trekkgrunnlaget Eventuelle bibetingelser
Ukjent	Nei
Død	Ja
Utvandret Ekt.stat ukjent Ikke gift Gift, EPNR ukjent Gift, EPNR kjent	Nei Ja Nei Ja: Hvis ektefellen gift og utvandret.
Bosatt i Norge Ekt.stat ukjent Ikke-gift Gift, EPNR ukjent Gift, EPNR kjent	Nei Ja Nei Ja: hvis ektefellen gift og bosatt i Norge.
EPNR: Ektefelles personnummer.	

3.2. Innhenting av data.

Avsnittet beskriver hvilke datakilder og data som er aktuelle for denne koblingen. Det kunne vært ønskelig med flere data, deriblant fra skattebåndene. Det kunne også vært ønskelig å samkjøre utvalget med andre undersøkelser, spesielt Folke- og bolig tellingen. Dette er ikke aktuelt i denne omgang, dels fordi det ligger utenfor vår søknad om å koble registre, dels fordi det vil medføre forsinkelser og komplikasjoner vi ikke ønsker i en utviklingsfase. Derimot kan dette bli aktuelt ved oppdateringer av datagrunnlaget, spesielt hvis husholdningsmodellen og trygdemodellen skal samles i en modell.

Fra kvinnefilen vil hente inn opplysninger om fødeår for hvert barn av kvinner i primærutvalget. Ved endogen rekruttering kan det også være viktig å kjenne reg.status for disse barna, spesielt de som er under 16 år.

Fra utdanningsfilene (BHU) vil vi hente inn opplysninger om høyeste fullførte utdanning

og igangværende utdanning pr 1/10 for perioden 1985-1989. I den grad det er hensiktsmessig vil opplysninger om ekteskapeleg status, reg.status og bostedskommune bli hentet fra utdanningsfilene. I 1985 ble utdanningskoden endret, men i den grad det finnes BHU-filer for årene før, vil vi forsøke å koble lengre bakover i tid. Opplysninger som om personen er under utdanning kan også ha interesse, selv om vi ikke skulle klare å omkode fagfeltet. Andre utdanningsopplysninger vil bli vurdert.

Fra Rikstrygdeverkets pensjonspoengregister vil vi hente ut pensjonsgivende inntekt, pensjonspoeng, beregnet uførepoeng (BUP), uføregrad og skattekommune for hvert år i perioden 1967-1989. I tillegg har Statistisk sentralbyrå også årlige statusfiler med alle som har mottatt stønader fra Folketrygden. Herfra vil vi hente noen utfyllende opplysninger for årene (1985)-1989. Dette gjelder opplysninger om alders- og uføretrygd som er nødvendige i estimeringsarbeidet og for å få en utgangsbestand av alderstrygdede. I tillegg kan vi her finne opplysninger som er tilstrekkelige for å beregne tilleggspensjoner for eksisterende enker og enkemenn. For personer som blir enker/enkemenn i simuleringen vil vi uansett kjenne ektefellen.

3.3. Dataene organiseres.

Krav.

Datagrunnlaget må være tilgjengelig for så ulike formål som utgangspopulasjoner i MOSART, grunnlag for atferdsanalyser og i mulig statistikkproduksjon. Det er viktig at simuleringsmodellen klarer å lese inn forhistorien for individer i utgangspopulasjonen, samt for ikke-bosatte, på en slik måte at modellpopulasjonen kan brukes til å ta ut historisk statistikk for perioden 1967-1989, i tillegg til framskrivingsperioden 1990-2050. Det er lite trolig at de koblede filene vil bli oppdatert; vi vil starte på nytt ved hver oppdatering. Estimerte individuelle parametre må allikevel kunne legges tilbake til de samme individer for bruk i simuleringsmodellen. Det må også være mulig å tilføre syntetiske husholdningsdata. Det er viktig at hensynet til personvern ivaretas på en tilfredsstillende måte, samtidig som viktige personbånd ivaretas på de endelige filene. På grunn av tidligere ektefeller, er det nødvendig med et annet system for anonymisering enn det som er brukt i MOSART-1.

Forslag.

Denne delen av dataarbeidet vil neppe gi de store problemene, og forslaget vil derfor bare bli løselig skissert. Basis i datagrunnlaget vil være et sas-datasett med person*år som observasjonsenhet, med data for alle størrelser som endrer seg fortløpende. Rundt denne vil det være datasett som fanger opp begivenheter som ikke er hensiktsmessige å lagre på denne måten. Dette omfatter mer utfyllende opplysninger om ekteskap, fødsler med mere. Organiseringen av output fra MOSART kan brukes som et utgangspunkt.

3.4. Dokumentasjon.

Krav.

Dokumentasjonen av filene skal være så god at en person som kan noe SAS og Comparex/JCL, skal kunne bruke denne som et oppslagsverk ved bruk av datagrunnlaget. Med litt mer kunnskaper skal den også tjene som oppslagsverk ved oppdatering av datagrunnlaget. Det vi har sett på kobling/livshistoriefil for MOSART-1 vil mer enn tilfredstille kravene. Det er imidlertid viktig at dokumentasjonen blir "publisert" rimelig raskt, for at det skal være mulig å distribuere det til eksterne brukere og for å kunne referere til det i andre sammenhenger.

Forslag.

Noe konkret forslag er ikke nødvendig, men det er siktemål å få det ut som et Internnotat eller lignende.

4. Andre sider ved modellen.

4.1. Estimeringsarbeider.

I MOSART-1 har vi i stor grad benyttet råmater⁸ som estimerer for overgangssannsynlighetene. Dette har vært mulig fordi vi har hatt tilgang på datamaterialer med et stort antall observasjoner og fordi "mengden" forklaringsvariable har vært begrenset. I utvidelsene av MOSART er vi i større grad avhengige av å gjøre overgangssannsynlighetene til funksjoner av forklaringsvariablene, istedenfor å ha de liggende som store n-dimensjonale tabeller. "Mengden" av forklaringsvariable øker, og dette øker presset i retning av bedre utnyttelse av data og mer kompakt lagring av informasjonen. Det gjør det også mulig å trekke inn periodevariable for å estimere forskjeller i overgangssannsynlighetene mellom ulike perioder (år). Dette vil være et nyttig verktøy når vi skal endre overgangssannsynlighetene i alternative simuleringer. Av disse grunner vil vi legge stor vekt på å bruke mer formell økonometri enn til nå, helst med utgangspunkt i økonomisk teori. I MOSART-1 har programspråket SAS vært mye brukt, og SAS gir også gode muligheter for å estimere relasjoner med utgangspunkt i blant annet logitmodellen.

Datagrunnlaget beskrevet i avsnitt (3) vil dekke de fleste estimeringsbehovene. Spesielt kan fremheves at det vil gi opptil 23 år lange individhistorier med mange kjennetegn. 16-åringer og personer som er 70 år og eldre, vil ikke få pensjongivende inntekt selv om de skulle ha lønnsinntekt. Dette er imidlertid et lite problem, da yrkesdeltakingen allikevel er lav og lønnsinntekten ikke får noe å si for opptjeningen av pensjonsrettigheter i disse aldersgruppene. En løsning kan være å ekstrapolere de overgangssannsynligheter vi estimerer. Dahl (1988) har undersøkt yrkesdeltakingen blant personer over aldersgrensen, og en annen mulighet er å benytte resultater herfra til et enkelt opplegg. Inntektsundersøkelsene gir en tilsvarende mulighet. En annen svakhet er at timeverk helt mangler i datagrunnlaget. AKU kan her supplere med data til en (enkel) ettermodell som beregner timeverk med utgangspunkt i den arbeidsstyrke MOSART-T gir.

Uføretrygd.

Overgangen til uførhet ville mest naturlig blitt estimert som en hasard-rate modell med kontinuerlig tid hvor alder måler tid under risiko. Imidlertid er innslaget av tidsavhengige kovariater sterkt, deriblant nødvendige periodevariable som kan fange opp liberalisering av (praktiseringen av) trygdereglene og variasjoner i blant annet arbeidsledighet. I tillegg vil data og input til simuleringsmodellen være diskret i tid. Derfor vil overgangen til uførhet bli estimert med utgangspunkt i en enkel logitmodell. Siktemålet vil være å få en god føyning mellom data og de høyresidige variablene som blir valgt. Logitmodellen inngår som en standard del av programspråket SAS. Utgangspunktet vil være relasjonen:

$$p_{it} = \text{EXP}(X_{it}\beta + z_t\gamma) / (1 + \text{EXP}(X_{it}\beta + z_t\gamma))$$

Hvor:

- p_{it} angir sannsynligheten for at person (i) skal bli uføretrygdet i løpet av år (t).
- X_{it} er en vektor (forklarings)variable som beskriver person (i) i år (t).
- β er en tilhørende vektor med parametre.
- z_t er en vektor dummyvariable med en variabel for hvert år unntatt ett.
- γ er en tilhørende vektor med parametre.

Risikomengden vil omfatte alle personer i alderen 16 til 66 år som ikke allerede er uføretrygdet. Høyresidige variable vil i første omgang være kjønn, alder, fjorårets yrkesdeltaking,

⁸ Observerte andeler i hver enkelt gruppe av befolkningen som gjør en overgang.

utdanningsnivå, periodeeffekter og muligens uobservert heterogenitet. Menns og kvinners overgang til uførhet vil bli estimert uavhengig, slik at alle kjønnsforskjeller kommer fram. Alder vil brukt som en kontinuerlig variabel, og overganger til uførhet i 1989 tyder på at det vil være mulig å få en god føyning med en enkel transformasjon av alder.

Kjeldstad (1990) har analysert overganger til uførhet med utgangspunkt i Rikstrygdeverkets pensjonspoengregister. Her påvises klar sammenheng mellom overgang til uførhet og faktorene stabil yrkesdeltaking og/eller lav yrkesinntekt. Imidlertid er de begrepene som er brukt her lite operasjonelle i MOSART. Yrkesdeltaking vil derfor i første omgang bare være representert ved om man hadde pensjongivende inntekt året før. Utdanning vil være representert ved dummyvariable for relativt grove utdanningsgrupper, i hovedsak etter nivå på høyeste fullførte utdanning. Utdanning er tatt med som variabel fordi MOSART viser en klar vekst i utdanningsnivået. Periodevariable vil være representert ved en serie (tids)-dummies, slik at "nivået" på overganger til uførhet blir estimert for hvert enkelt år. Denne tidsserien vil danne utgangspunktet for å angi nivået på overganger til uførhet i simuleringsperioden. Dette vil ikke gjøre modellen til en prognose i statistisk forstand, men vil allikevel være en stor framgang i forhold til det å estimere overgangene i et enkelt år.

Risikoen for å bli uføretrygdet er trolig mer ujevnt fordelt i befolkningen enn det våre forklaringsvariable klarer å fange opp (uobservert heterogenitet). Etterhvert som en større andel blir uføretrygdet, vil dette påvirke gjennomsnittlige risiki for å bli uføretrygdet blant de som ennå ikke er blitt det. Kan slike effekter påvises, vil det ha interesse både for simuleringsmodellen og i seg selv. Kravdal et al. (1989) har analysert uobservert heterogenitet ved skilsmisse på grunnlag av registerdata i Statistisk sentralbyrå. Velges et tilsvarende opplegg vil estimeringen bli mer krevende økonometrisk og EDB-teknisk, og forutsetter ekstern veiledning.

Uføregrad vil bli simulert med utgangspunkt i en enkel overgangsmatrise med ny uføregrad som resultat, og med kjønn, grove aldersgrupper og fjorårets uføregrad som forklaringsvariable. Disse overgangene er trolig av mindre betydning, da omlag 75 prosent av alle uføretrygdede er 100 prosent uføre.

Pensjongivende inntekt.

Simuleringen av pensjongivende inntekt vil nødvendigvis måtte bli mer komplisert enn overgangen til uføretrygd. Dette skyldes at pensjongivende inntekt er en kontinuerlig variabel med mange ulike innslag av diskrete sprang, for eksempel vesentlige endringer i arbeidstid fra ett år til det neste. Pensjongivende inntekt vil også i større grad "velges", og økonomisk teori og skatteparametre vil derfor være interessante å få med i modellen. Det er da lite hensiktsmessig å forsøke å spesifisere et estimeringsopplegg før man har arbeidet mer med problemstillingen.

Arbeidet vil starte med å liste ut en del individhistorier for å finne mulige mønstre. Dette vil etterhvert gjøres mer systematisk gjennom tabelluttak, spesielt for å finne faktorer som påvirker endringen i pensjongivende inntekt og hvilken betydning observerbar individuell heterogenitet har (datamaterialet gir opptil 22 overganger for hvert individ). Parallelt vil vi utvikle et økonometrisk opplegg, og en mulighet er en kombinasjon hvor en logitmodell fanger opp overganger ut av og inn i arbeidsstyrken, og vanlig lineær regresjon fanger opp veksten i pensjongivende inntekt for de som er i arbeidsstyrken. Valget av variable vil i stor grad følge modellbeskrivelsen i avsnitt (2.2), men det er viktig at de variable som velges kan gjøres operasjonelle i simuleringsmodellen. Dette begrenser mengden av forklaringsvariable (internminne) og modellens kompleksitet (regnekraft).

4.2. Arbeid som må utføres på den nye modellen.

Det må innarbeides nye kjennetegn for trygdestatus, uføregrad, pensjongivende inntekt og muligens noen variable som fanger opp observerbar individuell heterogenitet. Simuleringen av arbeidstilbud og trygdestatus må endres. Disse punktene burde være

relativt greie å få gjennomført, og krever i liten grad nytenkning på programmeringssiden. Videre må en ny utgangspopulasjon leses inn, og en ny livshistoriefil må også lages. Her er det viktig å ta vare på livshistoriene fra utgangspopulasjonen. I tillegg må dataene organiseres slik at det blir enkelt å beregne alders- og uføretrygd. Også dette burde være en relativt enkel oppgave.

Det som blir vesentlig nytt er rutiner for beregning av alders- og uføretrygd. Problemet er at disse reglene er kompliserte og omfatter opplysninger fra hele livsløpet. I tillegg kommer at pensjonen samordnes mellom ektefeller, og at enker/enkemenn kan ha krav på tilleggspensjon på grunnlag av tidligere ektefelle. Det er viktig at disse beregningsrutinene er effektive og enkle å redigere, fordi mye av modellens bruk vil ligge i å endre på beregningsreglene. Arbeidet med denne delen av modellen må ikke undervurderes, selv om det er mulig å trekke på tidligere erfaringer og programvare fra modellene ODIN og MAFO/MIFO. Sentrale brukere som Finansdepartementet har gitt uttrykk for ønsker om at denne delen av modellen også skal være tilgjengelig for brukere av modellen. De store datamengdene kan gjøre dette vanskelig, men vil bli vurdert på et senere tidspunkt.

4.3. Forholdet til andre tilstøtende prosjekter.

Arbeidet med en trygdemodell MOSART-T vil ha berøringspunkter med en rekke andre prosjekter og modeller, og dette avsnittet drøfter noen av disse forholdene. Det er viktig at disse prosjektene er rimelig koordinert, slik at dobbeltarbeide unngås i estimering og modellbygging. Det er også viktig for brukere av modellen at tallene i størst mulig grad er konsistente. Det er en lite givende jobb å kommentere "tekniske avvik", og det vil også øke bruksverdien hvis de ulike modellverktøyene bygger på samme forutsetninger om utviklingen framover.

MOSART - en eller flere modeller:

Selv om det er et klart siktemål å samle de ulike MOSART-modellene i en modell på et senere tidspunkt, kan det likevel oppstå problemer hvis de skal forenes i en modell. Dette gjelder slike forhold som bruk av regnekraft og internminne, og det er neppe rom for noen ekstreme utvidelser i en bestemt retning. Valget av utgangspopulasjon kan også by på problemer, men dette bør kun føre til mindre (stokastiske) forskjeller i den grad trygdemodellen og husholdningsmodellen blir basert på ulike datakilder. Derimot kan det oppstå problemer ved valg av endogen og eksogen rekruttering, og dette problemet drøftes her mer utførlig.

De som blir pensjonister i 2050 er allerede født, og innvandring vil være eksogent bestemt i modellen. I så måte er trygdemodellen i stor grad uavhengig av de andre befolkningsmodellene. Derimot vil det være en stor fordel om trygdemodellen også blir arbeidsstyrkemodellen. Det vil gjøre at flest mulig av de forutsetninger som påvirker modellens framskrivning av alders- og uføretrygd er konsistent med andre planleggingsverktøyer. I den grad fødsler påvirker (kvinner) tilpasning på arbeidsmarkedet, vil opptjening av pensjonspoeng være påvirket av de forutsetninger som bestemmer befolkningstilveksten. I så måte kan det være ønskelig med endogen rekruttering for å sikre konsistens, spesielt ved simuleringer med ulike fruktbarhetsalternativer.

Andre modeller og prosjekter.

BEFREG er en modell som framskriver befolkningen etter alder og kjønn på kommune-nivå, og det er lagt stor vekt på regionale aspekter. Dette er en godt innarbeidet og sentral modell i offentlig forvaltning. BEFREG er den offisielle befolkningsframskrivningen, og det er derfor viktig at MOSART er konsistent med denne. Det kan gjøres ved å holde på eksogen rekruttering i MOSART-T, ved å legge inn skranker på antall fødsler i MOSART eller ved at BEFREG avleder fødselsrater fra MOSART. De to siste alternativene gjør at trygdemodellen og husholdningsmodellen fortsatt kan samles i en modell på et senere tidspunkt.

Fruktbarhets- og yrkesundersøkelsen (FoY) er en omfattende intervjuundersøkelse av noen kohorter kvinner og menn. Det er lagt vekt på å få fram livshistoriedata for samliv, barn, utdanning og yrkesdeltaking. Det er planlagt analyseprosjekter som kan være relevante for MOSART, og muligheter for å benytte overgangssannsynligheter herfra vil bli vurdert. Dette gjelder spesielt demografiske begivenheter, men også andre overganger er aktuelle.

MSG er en makroøkonomisk likevektsmodell rettet inn mot langsiktige analyser. Arbeidstilbudet og reallønnen blir endogent bestemt i en av de nyeste versjonene av modellen. Befolkningen er eksogent framskrevet i fjorten ulike husholdningstyper, og dette er input som på et senere tidspunkt vil bli hentet fra MOSART(-H). I den grad trygdemodellen skal hente vekst i reallønner fra en makromodell vil MSG være den mest aktuelle. MODAG er makroøkonomisk modell med kortere tidshorisont hvor også ulikevekter inngår. Arbeidsstyrke, sysselsetting og reallønn blir endogent bestemt. Befolkningen er framskrevet etter kjønn og alder, men også utdanning kan bli inkludert på et senere tidspunkt. MODAG inneholder også relasjoner for uføretrygd, og en viss koordinering kan være ønskelig her.

LOTTE er en statisk mikrosimuleringsmodell basert på inntektsundersøkelsene, med en detaljert beskrivelse av skattereglene. Modellen brukes til å gi (presise) skatteanslag for budsjettarbeidet i Finansdepartementet og Stortinget. MOSART er i større grad rettet inn mot langsiktige endringer i sammenheng med befolkningsutviklingen. Det vil derfor være liten grad av overlapp mellom LOTTE og MOSART-T, selv om metode og problemstillinger kan ligne hverandre. ODIN er en modell som beregner virkningen av skatter og overføringer på ulike typehushold over flere år. Dette kan blant annet brukes til å kontrollere skatte- og overføringsreglene. Heller ikke her vil det være noe overlapp, men erfaringer og programvare fra ODIN kan være nyttig ved programmeringen av MOSART-T.

MAFO var tidligere alderstrygdmodellen i SSB. Av ulike grunner har den ikke blitt oppdatert, og vil bli erstattet av MOSART-T når denne fullføres. Rikstrygdeverket utvikler også en trygdemodell, og vi har kontakt med denne prosjektgruppen. Selv om modellene har sterke felles trekk, var det enighet om at begge prosjektene burde fortsette. Rikstrygdeverket har bruk for egne budsjettmodeller og egen beredskap, mens MOSART-T inngår som en naturlig utvidelse av en større mikrosimuleringsmodell. Et annet større dataprojekt med utgangspunkt i Rikstrygdeverkets data er også underveis, og dette omhandler klientstrømmer i trygdesystemet, forkortet KIRUT. Formålet er å analysere hvordan personer går fra yrkesaktivitet til å bli uføretrygdet, og hvilken rolle ulike ordninger i trygdesystemet spiller i denne prosessen. KIRUT kan ikke erstatte det dataarbeidet som inngår i dette prosjektet. Derimot kan analyseresultater være av interesse, i den grad de ikke er for detaljerte i forhold til trygdesystemet.

Seksjon for mikroøkonometri (GØK) arbeider med modeller for arbeidstilbud med utgangspunkt i mikroøkonomisk teori. Slik disse modellene foreligger i dag er de lite egnet til bruk i MOSART, og grunnen til dette er som tidligere nevnt at de bruker for mye regnekraft og at de ikke omfatter hele befolkningen. GØK har imidlertid som mål å utvikle en mikromodell som simultant simulerer (alle) husholdningenes tilpasning over livsløpet med hensyn til konsum, arbeid, utdanning og familie. Hvis dette prosjektet realiseres, vil det være av stor interesse å bruke dette i MOSART. Inntil disse planene blir mer konkrete, vil det være for usikkert å basere trygdemodellen på dette. De angitte estimeringsarbeidene for trygdemodellen vil derfor fortsette inntil videre.

Vedlegg.

(a) Tids- og ressursplan.

PROSJEKT/EMNE	PÅBEGYNT/ AVSLUTTET	RESSURS BRUK	HVEM	KOMMENTARER
FORPROSJEKT- RAPPORTEN	1/1-1991 13/5-1991	Måneds- verk Ca 2-3	SPA	Inklusive prosjektsøknader m.v.. Trykkes umiddelbart som Internotat.
DATAGRUNNLAGET		Ialt: 3		
- Trekking av utvalget	9/4-1991 1/6-1991	0,5	SPA	Hjelp fra GEM e.a. er ønskelig for å komme gjennom administrativt arbeid som lese- tillatelser og filidenter, samt problemer med JCL/SAS.
- Innlesning av filer	1/6-1991 30/8-1991	1	SPA	
- Organisering av data	9/4-1991 30/8-1991	1	SPA	
- Dokumentasjon	9/4-1991 15/9-1991	0,5	SPA	Internnotat Hjelp fra GEM e.a. for systemdokumentasjon vil være meget ønskelig.
ESTIMERINGS- ARBEIDER				
- Uføretrygd/ alderstrygd (første versjon)	1/9-1991 15/10-1991	1-2	SPA	Dokumenteres i eget notat e.l., for å opp- summere erfaringer med datagrunnlaget og estimeringsmetoder, samt rapportere eventuelle empiriske resultater.
- Pensjonsgivende inntekt (første versjon)	15/9-1991 15/12-1991	2	SPA	Dokumentasjonen inngår i Rapporten på trygdemodellen.
MOSART-T				
- Simulerings- modellen	1/1-1992 ??	??	GEM	
- Tabell- programmet	1/1-1992 ??	??	GEM	Egen dokumentasjon, Internnotat e.l., viktig ved mulig "salg" av modellpopulasjonen/tabellprogrammet.
- Dokumentasjon	15/9-1991 ??	??	SPA/ GEM	Rapport med modellbeskrivelse, bruksmåter og resultater.

(b) Tabeller.**Tilgang og avgang av uføretrygdede 1987-1989.**

	1987	1988	1989
Tilgang	35 203	31 516	31 456
Avgang	21 994	21 813	22 098
- Døde	3 307	3 527	3 257
- Alderspensjon	17 018	16 643	17 592
- Annen årsak	1 002	267	487
- Uoppgitt	667	1 376	744

Kilde: Trygdestatistisk årbok 1990.

Befolkningen etter kjønn, alder og viktige kjennetegn i forhold til Folketrygden.

Alder	Andel alders- og uføretrygdede				Andel med pensjonsgivende inntekt			
	Beholdning		Overgang		I alt		Større enn 1G	
	Menn	Kvinner	Menn	Kvinner	Menn	Kvinner	Menn	Kvinner
16-19	0.3	0.2	0.1	0.1	89.6	88.2	41.3	30.4
20-24	0.6	0.6	0.1	0.1	96.8	92.9	77.3	71.2
25-29	1.2	1.3	0.2	0.2	97.7	87.7	90.8	73.4
30-34	2.0	2.4	0.3	0.4	97.3	85.9	93.2	73.0
35-39	2.9	4.1	0.4	0.7	96.9	87.4	93.5	77.1
40-44	4.1	6.4	0.5	1.1	96.5	88.2	93.3	80.0
45-49	5.9	10.0	0.8	1.6	95.2	86.5	91.8	78.6
50-54	10.7	16.6	1.8	2.9	92.5	80.1	88.3	70.7
55-59	18.6	24.5	3.3	3.9	87.8	69.8	82.2	58.1
60-64	32.1	33.0	6.6	5.2	78.6	55.3	69.4	41.7
65-66	45.7	39.0	10.7	6.1	67.5	41.9	54.6	27.8
67	91.9	95.5			60.0	35.4	43.1	19.2
68	94.9	97.2			44.2	24.8	17.9	7.0
69	95.5	97.7			38.5	19.6	14.0	4.8

Kilde: Trygdestatistisk årbok 1990, alders- og uføretrygdede gjelder 1989, mens pensjonsgivende inntekt gjelder 1988 og omfatter heller ikke 16-åringer.

Andel i arbeidsstyrken i 1988 etter alder og kjønn.

Alder	Kjønn	
	Menn	Kvinner
16-19	53,4	53,1
20-24	82,8	72,6
25-29	91,0	73,6
30-39	95,6	80,3
40-54	92,9	81,6
55-66	71,0	49,9
67-74	16,8	7,5

Kilde: AKU.

(c) Litteratur.

Andreassen, Leif og Dennis Fredriksen (1991): "MOSART - En mikrosimuleringsmodell for utdanning og arbeidsstyrke", Økonomiske analyser 91/2, Statistisk sentralbyrå.

Allison, Paul D. (1984): "Event history analysis - regression for longitudinal event data", Sage university paper 46.

Dahl, Grete (1988): "Yrkesdeltakelse for personer over aldersgrensen", Rapporter 88/5 Statistisk sentralbyrå.

Kjeldstad, Randi (1990): "Yrkesdeltaking, yrkesinntekt og uførepensjonering", INAS-rapport 1990:3.

Kravdal, Øystein, Tom Wennemo og Rolf Aaberge (1989): "Unobserved heterogeneity in models of marriage dissolution", Discussion paper 42, Statistisk sentralbyrå.

Statistisk ukehefte 1/91, "Framskrivning av arbeidsstyrken, 1989-2040", Statistisk sentralbyrå.

Trygdestatistisk årbok 1990, Rikstrygdeverket (1990).