



*Siri Størmer*

## **Sammenhengen mellom syssel- setting i AKU og registrene: Måling av kvalitet og prediksjon**

# Innhold

1. Innledning.....	3
2. Kvalitetsbegrepet og konsistens .....	4
2.1. Kvalitetsmålinger .....	4
2.2. Definisjon av ett kvalitetsmål.....	5
3. Prediksjon.....	5
4. Definisjon av sysselsetting i AKU og registrene .....	6
5. Datagrunnlag .....	8
5.1. Nærmere beskrivelse av AKU.....	8
5.2. Nærmere beskrivelse av registerdata.....	8
5.3. Beskrivelse av koblingen mellom de enkelte kvartalsfilene i AKU, AT og SOFA .....	8
5.4. Beskrivelse av koblingen mellom AKU, AT og LTO.....	9
5.5. Kobling mellom AKU, AT, SOFA og LIKN m/MOMS.....	10
6. Analyse av sysselsettingskvaliteten i registrene .....	10
6.1. Kommentarer knyttet til arbeidsforholdene i tabellene.....	10
6.2. Tabeller med aldersfordeling .....	10
6.3. Kommentar til koblingen mellom AKU, AT og LTO.....	11
6.4. Feilkilder i tabellene.....	12
6.5. Kommentarer til Tabell 1 med AKU, AT og LTO .....	12
6.6. Kommentarer til Tabellen med AKU, AT og SOFA .....	14
6.7. Kommentarer til tabeller der det er gruppert på alder.....	17
7. Prediksjonsfunksjon av sysselsetting basert på registeropplysninger.....	17
7.1. Resultat av prediksjonsfunksjon .....	18
Vedlegg A: Litt om filene .....	20
Vedlegg B: Tabeller .....	21
Vedlegg C: Avhengighetsmålet $\gamma$ .....	32
Vedlegg D: Om programmer .....	35
De siste utgitte publikasjonene i serien Notater.....	36

# 1. Innledning

SSB har gjennom de siste 10 årene satset på økt bruk av registre i statistikkproduksjonen. Det er da viktig å undersøke kvaliteten registeropplysningene som eventuelt skal erstatte utvalgsundersøkelser.

Prosjektet er en del av arbeidet med å planlegge Folke- og Boligtellingen i år 2000. Det er viktig å gjennomføre kvalitetsstudier av registrene og utvikle metoder for å måle og øke kvaliteten. I dette arbeidet er AKU benyttet for å sammenlikne med registeropplysninger om sysselsetting. Et av formålene er å konstruere mål for konsistens mellom sysselsetting målt i AKU og sysselsetting målt ved kombinasjon av informasjon fra flere registre. Hensikten med å sammenlikne målinger basert på registre og utvalgsundersøkelser, er å måle kvaliteten på registre i forhold til kjennemerkene i utvalgsundersøkelsen.

Imidlertid blir det problemer med å erstatte utvalgsundersøkelser ettersom spørsmålene som stilles i for eksempel AKU ikke direkte kan hentes ut av registre. For eksempel vil en fra arbeidstakerregisteret ikke finne opplysninger om hvorvidt en person som har flere arbeidsforhold i en periode arbeider heltid eller deltid, eller om personen har en hovedjobb og/eller flere bijobber. I tillegg vil forsinkede inn og utmeldinger skape problemer.

Foreløpig er arbeidstakerregisteret (AT), lønns- og trekkopgaveregisteret (LTO), arbeidssøkerregisteret (SOFA) og likning med momsregisteret (LIKN m/MOMS) sammenliknet med AKU. Vi har kun undersøkt kjennemerket sysselsatt/ikke sysselsatt.

Kvalitetsmålingene bygger på kryss-tabeller som angir sammenheng mellom sysselsetting målt fra registerene og AKU. Det er flere faktorer som påvirker kvaliteten av sysselsetting målt ved registerene blant annet hvor stabilt et arbeidsforhold er siden forsinkelsene ved inn- og utmelding påvirker kvaliteten. Vi har derfor delt personene inn etter kjønn og alder for å se hvordan kvaliteten varierer med hensyn på disse to faktorene.

Kvaliteten måles ved et avhengighetsmål som måler sannsynligheten for at opplysningene stemmer, og dette er beskrevet nærmere i kapittel 2. Kvalitetsstudiene er utført på registeropplysninger fra 1992 og spørreundersøkelsen i AKU fra 1. kvartal samme året.

Det er også prøvd ut en prediksjonsfunksjon der det estimeres totalt antall sysselsatte i populasjonen. Dette er en metode for å telle opp sysselsettingstall fra registeret som er kalibrert med AKU. Modellen er basert på en logistisk regresjonsmodell som benytter hovedstatuskoden i AKU som responsvariabel og aktuelle registeropplysninger som forklaringsvariable. Det er utifra denne regresjonsmodellen estimert parametere som deretter benyttes på hele populasjonen. Det blir estimert en sannsynlighet for hvert individ om personen er sysselsatt eller ikke på bakgrunn av deres registeropplysninger. Sannsynlighetene summeres, og det blir estimert et totaltall for antall sysselsatte i populasjonen. Totaltallet stemmer godt med det estimerte sysselsettingstallet basert på AKU fra 1. kvartal 1992.

Som nevnt er det skrevet litt om kvalitetsmåling generelt og om å sammenligne to kilder og eventuelle problemer rundt dette beskrives i avsnitt 2. I avsnitt 3 beskrives prediksjon og hvorledes en kan tenke seg å utnytte dette til estimering basert på registerstatistikk. I avsnitt 4 defineres binære sysselsettingsvariable basert på registervariable som er benyttet til å kjøre ut tabeller. I avsnitt 5 beskrives data grunnlaget og i avsnitt 6 presenteres tabeller og kommentarer til disse. I avsnitt 7 analyseres prediksjonsfunksjonen.

## 2. Kvalitetsbegrepet og konsistens

### 2.1. Kvalitetsmålinger

Det skal konstrueres mål på kvalitet mellom AKU og registerdata. Hensikten er å kartlegge sammenlikningen mellom de to nevnte kildene for å konstruere registerbaserte sysselsettingsdefinisjoner. Videre er det viktig å få målt kvaliteten i registeropplysningene sammenliknet med en spørreundersøkelse med henblikk på FOB 2000.

Kvalitetsmålene vil være med på å justere for skjevhet og uoverensstemmelser som kan benyttes i nye sysselsettingsmodeller og eventuelt modeller for andre kjennetegn.

I dette notatet måles kvaliteten mellom hovedstatuskoden i AKU og registre på grunnlag av antall overensstemmelser og uoverensstemmelser. Først blir kvalitetskontrollen foretatt, og deretter blir kvalitetsmålet beregnet. Til dette formålet er det definert følgende mål for kvalitet på grunnlag av:

- Krysstabeller mellom AKU og ett register
- Multiple krysstabeller, tabeller mellom AKU og flere registre

Noen kartlegging av forsinkelser er ikke tatt med. Det vises til notat fra Seksjon for arbeidsmarkedstatistikk av Dale-Olsen (1996) om klassifisering av sysselsatt/ ikke sysselsatt, om forsinkelser i AT og Bø/Nesheim (1993) om sammenheng mellom AKU, AT og SOFA. Videre vises det til analyse av Utne/Vassnes (1995) der de belyser sammenheng mellom AT og LTO-registeret.

I AT registreres start og stoppdato for inn- og utmelding fra arbeidsgiver. Det forekommer ofte forsinkelser på inn- og utmelding i AT. Dette gir også utslag ved å sammenligne svar fra AKU om arbeidssituasjonen på et gitt tidspunkt.

I SOFA registreres start og stoppdato ved at personer selv melder seg arbeidsledig eller blir meldt inn dersom man kategoriseres i en ledighetsstatus. Ledighetsstatusen er beskrevet nærmere i avsnitt 4. Det er utarbeidet krysstabeller mellom AKU og AT, eller AKU og SOFA. Her er det konstruert kodede variable slik at det kan telles opp antall personer som tilhører de ulike kategoriene.

Tabellene brukes til å registrere konsistens mellom AKU og det aktuelle register på et gitt tidspunkt, som igjen gir et tallfestet mål på kvaliteten mellom disse kildene.

I multiple krysstabeller tallfestes samsvar mellom AKU og flere registre. I dette notatet er det laget en krysstabell mellom AKU, AT, SOFA og LIKN m/MOMS. En person som har svart at han er sysselsatt i følge AKU, bør ikke stå registrert som helt ledig i SOFA på samme tid. De ulike svarene i AKU og individenes registeropplysninger på et gitt tidspunkt tallfestes. Dette kan gi et mål på kvalitet mellom registerdata og AKU på et gitt tidspunkt. Videre kan tidspunktet utvides til en periode, og nye krysstabeller kan kjøres ut. Foreløpig er ikke det gjennomført.

Kvalitetsmålingen må ta skille mellom forskjeller som skyldes avvik i definisjonen og forskjeller som skyldes feilmåling i de ulike kildene som sammenliknes. Det sammenlignes mellom AKU og registerdata. Definisjonen på å være sysselsatt i AKU avviker fra å være sysselsatt i registeret. AKU-definisjonen på å være sysselsatt tar utgangspunktet i ILO-definisjonen. I AKU defineres intervjuobjektet (IO) som sysselsatt dersom vedkommende har utført minst en times inntektsgivende arbeid siste uke før spørreuken i AKU. Det er selvsagt grader av å være sysselsatt, men i denne kvalitetsundersøkelsen tas det kun utgangspunkt i hovedstatuskoden i AKU. Det spørres da etter om IO er sysselsatt, ledig eller utenfor arbeidsstyrken.

LTO-registeret er skattemyndighetenes register over lønns- og trekkoppgaver. Oppgaveplikten påhviler for selvstendig næringsdrivende eller enkeltmannsbedrifter, innehaveren, og i selskap, forening, institusjon eller innretning, den daglige leder. Det er foretatt kvalitetskontroll mellom AKU, AT og LTO. Videre henvises til Dale-Olsen (1996) om klassifisering av sysselsatt/ikke sysselsatt for analyse av kobling mellom AT og LTO. Når det gjelder dato på lønns- og trekkoppgavene, stemmer denne opplysningen svært sjelden for tidsbegrensede arbeidsforhold.

I AT skal en person registreres som arbeidstaker dersom han har utført en jobb på gjennomsnittlig minst 4 timer i uken og av 6 dagers varighet. Det er et strengere krav å bli innmeldt i AT enn å definere seg som sysselsatt i AKU. I tillegg fanges ikke de selvstendige arbeidstakerne opp i AT. Disse er registrert i LIKN m/MOMS, og en har forsøkt å definere de selvstendig næringsdrivende basert på inntektsvariabelen. I LIKN m/MOMS registreres ikke dato for arbeidsforholdet. Det er derfor umulig å sjekke kvalitet angående tidspunkt mellom AKU og LIKN m/MOMS.

## 2.2. Definisjon av ett kvalitetsmål

For å vurdere hvor god sammenhengen mellom svar i AKU og deres registeropplysninger fra AT og LTO er, er det beregnet et avhengighetsmål. Målet er konstruert ved at en beregner den betingede sannsynligheten for at informasjonen i AT og LTO er sammenfallende minus sannsynligheten for at informasjonen ikke sammenfaller gitt AKU status. Av dette kan ses at målet tar verdier mellom [-1, 1].

Det antas at det velges to personer tilfeldig som begge har forskjellige register-verdier, dvs at det velges tilfeldig en sysselsatt og en som ikke er registrert som sysselsatt både i AT og LTO gitt personenes svar i AKU. Konsistensmålet formuleres følgende:

$$\gamma = P(\text{Register-verdien} = \text{for begge} \mid \text{Personene har svart likt i AKU}) - P(\text{Register-verdien} \neq \text{for begge} \mid \text{Personene har svart likt i AKU})$$

Formalisert skrives avhengighetsmålet på følgende vis:

La for en vilkårlig person  $(X, Y, Z)$  være henholdsvis (register-verdi i AT, register-verdi i LTO, svar i AKU). For to verdier av  $(X, Y)$ :  $(X_1, Y_1)$ ,  $(X_2, Y_2)$  beregnes  $\gamma$  som differansen mellom to sannsynligheter:

$$\begin{aligned} (2.1) \quad \gamma &= P(X_1 = Y_1 \cap X_2 = Y_2 \mid X_1 \neq X_2, Y_1 \neq Y_2, Z_1 = Z_2) \\ &\quad - P(X_1 \neq Y_1 \cap X_2 \neq Y_2 \mid X_1 \neq X_2, Y_1 \neq Y_2, Z_1 = Z_2) \\ &= P((X_1 - X_2)(Y_1 - Y_2) > 0 \mid X_1 \neq X_2, Y_1 \neq Y_2, Z_1 = Z_2) \\ &\quad - P((X_1 - X_2)(Y_1 - Y_2) < 0 \mid X_1 \neq X_2, Y_1 \neq Y_2, Z_1 = Z_2) \end{aligned}$$

De estimerte  $\gamma$ -ene til hver tabell er presentert og tolket i vedlegg C. Videre er den teoretiske utledningen av  $\gamma$  også dokumentert i vedlegg C.

## 3. Prediksjon

Det er utarbeidet en prediksjonsmodell for å predikere antall sysselsatte personer. Det er tatt utgangspunkt i hovedstatuskoden i AKU der en binær variabel for sysselsatt/ikke sysselsatt er konstruert. Denne er benyttet som responsvariabel der de AKU-intervjuedes registeropplysninger er brukt som forklaringsvariable. Vi har benyttet en logistisk regresjon, og de estimerte parametrene er benyttet i en prediksjonsmodell for å beregne et sysselsettingstall for hele populasjonen.

Ved å ta i bruk registeropplysninger, kan det lages mer detaljert statistikk enn statistikken basert på AKU-utvalget. For eksempel kan det beregnes sysselsettingstall for mindre områder som kommuner. Det er ikke alle som er sysselsatt i følge AKU som registreres i AT. En av egenskapene til en prediksjonsfunksjon er at den kan rette opp en skjevhet i registerdefinisjonen. For å bli registrert i AT må en jobb være av minst 4 timer gjennomsnittlig arbeidstid per uke. Det er også en forutsetning at arbeidsforholdet skal være i minst 6 dager.

Videre er det fra tidligere undersøkelser vist at et betydelig antall meldinger er forsinkede i registerne, spesielt i AT. For å konstruere en relevant prediksjonsmodell, bør det tas hensyn til forsinkelsene. Det er mulig å modellere forsinkelsene i AT. En rimelig antakelse kan være at forsinkelsen er eksponentielt fordelt. Dette gjelder både forsinkede inn- og utmeldinger.

## 4. Definisjon av sysselsetting i AKU og registrene

Arbeidsstyrken defineres på følgende vis:

Arbeidsstyrken omfatter sysselsatte, ledige og personer på sysselsettingstiltak.

Først gir betegnelsen *arbeid* uttrykk for alt arbeide av *minst en times varighet*. Videre betyr en *lønnstaker* en person som utførte arbeid for lønn eller betaling i form av kontanter eller varer i løpet av referanseperioden. Dette gjelder også for selvstendige. Endelig defineres en *sysselsatt* person over en spesiell alder som en i løpet av en kort periode, enten en uke eller en dag, finnes blant kategoriene lønnstaker eller selvstendig.

På bakgrunn av registeropplysningene skal en sammenlikne konsistens mellom, i første omgang, hovedstatus-variabelen i AKU og opplysninger på samme tidspunkt i registerne.

I AT og LTO finnes som tidligere beskrevet start- og stoppdatoer for de yrkesaktives arbeidsforhold. Samme datoen som IO er spurt i AKU, sjekkes det i registerne om det foreligger noen registrerte arbeidsforhold på personen. Det er konstruert 3 variable basert på hver av de tre filene som utgjør hele AT og LTO-registeret. Dersom det forekommer et arbeidsforhold på individet som bare står i AT og ikke i LTO, vil den konstruerte variabelen **sysat** gis verdien 1, hvis ikke gis variabelen verdien 0. Videre vil variabelen **syslto** gis verdien 1 dersom det forekommer et arbeidsforhold på individet som bare står i LTO og ikke i AT. Variabelen er 0 for de øvrige individene. På filen som består av koblete arbeidsforhold mellom AT og LTO, er det konstruert en variabel. Variabelen **syilt** gis verdien 1 dersom det forekommer et arbeidsforhold på individet som står oppført med kobling i LTO og AT. Dessuten må arbeidsforholdet være registrert på samme tidspunkt som intervjudtidspunktet i AKU for at de 3 variablene skal få verdien 1. Dette skjer for hvert arbeidsforhold.

Før beregning av tabeller selekteres det slik at for hvert individ telles maksimum ett arbeidsforhold. Det er altså foretatt følgende definisjoner:

Fra AT, LTO, SOFA og LIKN m/MOMS er følgende binære variable definert:

### ***Sysselsatte basert kun på AT***

- Personen er sysselsatt dersom innmelding i AT har skjedd før intervjudato i AKU og utmelding etter intervjudato i AKU. Variabelen sysselsatt = 1 dersom dette holder og 0 ellers. Variabelen er på hele arbeidstakerregisteret.

### ***Klassifisering av ledighetsdefinisjon basert på SOFA***

- Personen får en kode som klassifiserer ledighetsstatus dersom innmelding i SOFA er skjedd før spørredato i AKU og utmelding etter spørredato i AKU. Variabelen ledstat = 0, 1, 2, 3, 4 eller 5 dersom dette holder.

Variabelen ledstat er konstruert ut i fra tiltakskoden i SOFA. Se nærmere beskrivelse nedenfor.

### **Definisjon av selvstendig næringsdrivende basert på LIKN m/MOMS**

- Personen er selvstendig dersom det er registrert minst en oppgave med større pensjonsgivende inntekt enn 22 000 kr på årsfilen i LIKN m/MOMS. Variabelen selvstendig = 1 dersom dette holder og 0 ellers.

### **Definisjon av sysselsatte basert på AT og LIKN m/MOMS**

- Det er konstruert en binær variabel som inneholder alle arbeidsaktive i følge AT og samtidig selvstendig næringsdrivende på gitt intervju tidspunkt. Variabelen selvstendig- og arbeidstaker = 1 dersom personen er sysselsatt på gitt tidspunkt i AT og/ eller oppgitt seg selvstendig næringsdrivende i LIKN m/MOMS og 0 ellers.

### **Registrering av sysselsatte basert kun på LTO**

- Personen registreres som sysselsatt i følge filen LTOUAT dersom det foreligger et registrert arbeidsforhold på individet under samme tidsrom som IO spørres i AKU. Variabelen **syslto** = 1 dersom dette holder og 0 ellers.

### **Registrering av sysselsatte basert på kobling mellom AT og LTO**

- Personen registreres som sysselsatt i følge filen LTOMAT dersom det foreligger et registrert arbeidsforhold på individet under samme tidsrom som IO spørres i AKU. Variabelen **sylt** = 1 dersom dette holder og 0 ellers.

### **Registrering av sysselsatte basert kun på AT**

- Personen registreres som sysselsatt i følge filen ATULTO dersom det foreligger et registrert arbeidsforhold på individet under samme tidsrom som IO spørres i AKU. Variabelen **syat** = 1 dersom dette holder og 0 ellers.

Fra AKU er det definert en binær variabel der :

- Personen er sysselsatt i følge AKU dersom tallet i hovedstatuskoden tilsier at intervjuobjektet har svart at han er sysselsatt. Variabelen sysselsatt i følge AKU har verdi lik 1 dersom dette holder og 0 dersom intervjuobjektet har svart noe annet.
- Dessuten defineres en variabel som er 1 dersom hovedstatuskoden sier IO er sysselsatt, 2 dersom hovedstatuskoden tilsier arbeidsledig og 3 dersom koden tilsier utenfor arbeidsstyrken.

Videre er variabelen ledstat konstruert på bakgrunn av variabelen tiltakskode i SOFA som er ledighetsstatuser med følgende gruppering:

0	H	tilhører kategori 0
1,2,3,6	Sysselsettingstiltak	tilhører kategori 1
4,5	Kvalifiseringstiltak	tilhører kategori 2
7	Attføringstiltak	tilhører kategori 3
‘.’	Ikke på tiltak eller registrert ledig	tilhører kategori 4
‘ ‘	Ikke oppført i SOFA	tilhører kategori 5

Det er altså konstruert 6 nye grupperinger på bakgrunn av tiltakskoden i SOFA. For beskrivelse av tiltakskoden, se vedlegg A.

På bakgrunn av de binære variablene, beregnes krysstabeller mellom den binære sysselsettingsvariabelen fra AKU mot variablene konstruert fra registrene.

## 5. Datagrunnlag

Til analysen er benyttet data fra System for persondata, og der inngår følgende filer:

Jobber fra AT og / eller LTO

- Selvstendighetsforhold (fra LIKN m/MOMS)
- Ledighets og tiltaksforhold (fra SOFA)
- Utdanning
- Demografidata

I tillegg kommer kvartalsfilene fra AKU 1992.

### 5.1. Nærmere beskrivelse av AKU

Grunnlaget for analysen er 4 kvartalsfiler fra AKU 92. Foreløpig er kun en kvartalsfil om gangen koblet opp mot de aktuelle intervjuobjektens registeropplysninger. AKU-filene er også analysert månedvis, dvs at respondentene blir delt inn etter hvilken måned de er intervjuet og sjekket mot registeropplysninger på samme tidspunkt. Nærmere om filbeskrivelse, se vedlegg A.

### 5.2. Nærmere beskrivelse av registerdata

Registerdata er hentet fra System for persondata og bygger på årsfiler. Registerne som er benyttet er:

- Jobber fra arbeidstakerregisteret (AT) og /eller lønns- og trekkoppgaveregisteret (LTO)
- Selvstendighetsforhold fra likning med momsregisteret (LIKN m/MOMS)
- Ledighets og tiltaksforhold (SOFA)

Det er konstruert egne SAS-datasett som inneholder utvalgte variable der variable fra registerne ovenfor og AKU inngår.

Datafilene fra AT og LTO er organisert slik at det ligger en record for hvert arbeidsforhold. For individer som har flere arbeidsforhold, foreligger det derfor flere recorder. For å undersøke om en person er sysselsatt i følge registerne på samme tidspunkt som IO svarer i AKU, benyttes start- og stoppdatoene i registerne til dette.

Det er som nevnt konstruert egne SAS-datasett som inneholder utvalgte variable fra registerne og nesten hele AKU. Filene er koblet på personnummer, så det forekommer flere recorder per personnummer der det er registrert flere arbeidsforhold.

Av registerne slik de er i dag, kan ikke personens hovedjobb og/ eller bijobber skilles ut. For å definere en person som sysselsatt i følge et register, må han/hun ha minst ett arbeidsforhold i det aktuelle registeret til samme tid som intervjutidspunktet i AKU. Nærmere om hvorledes sysselsettingsdefinisjonene er konstruert, se avsnitt 4.

### 5.3. Beskrivelse av koplingen mellom de enkelte kvartalsfilene i AKU, AT og SOFA

Det er koblet på personnummer mellom aktuelle opplysninger i AKU og i AT. For å få tak i personene som står i AT, er det koblet mellom arbeidstakerforhold uten og med kopling til LTO-forhold. Data fra AT og LTO er organisert i tre filer etter om jobben:

- finnes i begge registre
- finnes kun i AT
- finnes kun i LTO



Arbeidstakerforholdene foreligger pr. i dag på 2 filer som inneholder henholdsvis 127 048 og 2 284 327 observasjoner der hver jobb er en record på filene. Videre foreligger de ukoblede LTO-forholdene på en stor fil der alle oppgaver som inngår i LTO er registrert. Filen er på 4.024 millioner recorder med en record per forhold.

Det er dannet et datasett som inneholder kun personene som står i AKU i det aktuelle kvartalet og samtidig er registrert i AT og et eget datasett for personer som ikke er spurt i AKU, men som er registrert i AT. Det sistnevnte datasettet er konstruert til senere bruk.

Opplysningene som foreløpig er blitt benyttet i AKU er fødselsnummer, måned personene er spurt og hovedstatus. Hovedstatusen angir hvorvidt personen er sysselsatt, arbeidssøkende eller utenfor arbeidsstyrken. Det kan fås ut nærmere spesifisering som foreløpig ikke er blitt benyttet.

Videre er det kun lest inn registervariable som er aktuelle for kvalitetsstudiet og som direkte omhandler sysselsetting. Fra arbeidstakerforhold med og uten kopling til LTO er foreløpig «registrert start- og stoppdato» blitt studert for å måle konsistens mellom spørsmål om hovedstatus i AKU sammenfaller med opplysninger på samme tidspunkt i AT. Start- og stoppdatoene registrerer faktisk inn- og utmeldinger av arbeidstakerregisteret.

Datasettene som er beskrevet over er koblet sammen slik at en finner opplysninger om hvorvidt en person som er spurt i AKU er registrert i AT og /eller SOFA i løpet av 92.

SOFA-filen inneholder 1 112 180 observasjoner, og det er 1 record per ledighetsregistrering. Variablene som foreløpig er blitt benyttet fra SOFA-filen er «start- og stoppdato» som angir perioden på ledighetsregistreringen for hvert enkelt individ. Videre er «tiltakskode» benyttet som beskriver tiltakskoden satt ut fra arbeidssøkerstatus.

Forøvrig er de andre variablene fra AKU og AT de samme som beskrevet tidligere. På datasettet hvor det er koblet på personnummer over personene som ikke er spurt i AKU, er det 2 128 082 personer totalt etter at det er selektert slik at det kun er en record per person. Seleksjonen er foretatt ved at den første jobben per person registreres dersom han står oppført i AT eller om personen er ledig i følge SOFA. Grunnen til seleksjonen er at det skal estimeres antall sysselsatte og ikke antall sysselsettinger eller antall ledige.

#### **5.4. Beskrivelse av koblingen mellom AKU, AT og LTO**

Det er som nevnt koblet på personnummer mellom AKU og register-opplysningene. Man får tak i alle registrerte arbeidsforhold ved å koble sammen de tre nevnte filene mellom AT og LTO.

Det er først dannet et datasett som inneholder personene som er spurt i AKU 1. kvartal 92 og deres eventuelle opplysninger i AT og/ eller LTO som er av interesse.

De foreløpige opplysningene som er blitt benyttet fra AKU er fødselsnummer, spørreuke og hovedstatus. Hovedstatusen angir hvorvidt personen er sysselsatt, arbeidssøkende eller utenfor arbeidsstyrken. Det kan fås ut nærmere spesifisering om statusen som til nå ikke er blitt benyttet.

Når det gjelder register-opplysninger, er foreløpig start- og stoppdatoene blitt benyttet til å klassifisere sysselsettingsstatus. Av dette fås det ut om IO er sysselsatt, dvs registrert i AT og/ eller LTO på innterjtidspunktet med et eller flere arbeidsforhold, eller om det ikke finnes registreringer på IO.

## 5.5. Kobling mellom AKU, AT, SOFA og LIKN m/MOMS

Det koblede datasettet med variable fra AKU, AT og SOFA er koplet sammen med variable fra LIKNING m/MOMS. I LIKNING m/MOMS er det en record per oppgave slik at en person med flere oppgaver står oppført flere ganger på filen. Filen inneholder 3 820 291 recorder.

Det nye koblede datasettet inneholder de personene som er spurt i AKU i det bestemte kvartalet og registeropplysningene som fås fra registrene AT, SOFA og LIKN m/MOMS. Av LIKN m/MOMS blir selvstendighetsforhold fanget opp, noe som ikke registreres i AT. For å fange opp de selvstendige, er foreløpig inntekts-variablene benyttet. Det må utprøves hva nedre grense for inntekt skal være for at en person skal defineres som selvstendig næringsdrivende og derav som sysselsatt. Variablene som benyttes er «pensjonsgivende inntek for jord/skog/fiske og for annen selvstendig næring». LIKN m/MOMS registeret er en årsfil, og der forekommer ikke start- og stoppdatoer for arbeidsforholdene. En kvalitetskontroll av tidspunkt gir lite i denne sammenheng.

Videre er det konstruert et nytt datasett over personene som står i ett eller flere av de nevnte registrene men kun de som ikke er spurt i AKU. Personer som er eldre enn 74 år i 1992 tatt ut for å få samsvar med AKU tall.

På datasettet koblet med AKU er det 21 907 recorder, dvs. 21 907 personer etter selektering. På datasettet som inkluderer personene som ikke spørres i AKU, er det 4 021 764 observasjoner før det er selektert med kun 1 record per personnummer. Etter selektering er det 3 117 397 recorder på filen. Dette datasettet er benyttet til prediksjonsfunksjonen i kapittel 7.

## 6. Analyse av sysselsettingskvaliteten i registrene

Et utvalg tabeller er tatt med som angir hva IO har svart i AKU mot deres registeropplysninger fra AT, SOFA og LIKN m/MOMS. Tabellene er basert på de 3 ukene IO er spurt i 1. kvartal og gått inn i registrene på samme tidspunkt.

Nedenfor følger kommentarer til tabellene som gjelder koblingene mellom AKU, AT og LTO og AKU, AT, LTO og SOFA.

### 6.1. Kommentarer knyttet til arbeidsforholdene i tabellene

For å få riktig antall personer som er registrert i ett eller flere registre, er det selektert slik at det registreres kun ett arbeids- eller ledighetsforhold per individ. Tallene er lettere å tolke ettersom en ønsker å studere konsistens mot IO's svar i AKU.

### 6.2. Tabeller med aldersfordeling

Det er kjørt ut tabeller på aldersfordeling for å se på konsistens mellom AKU og registeropplysninger innen ulike aldersgrupper. Aldersgruppene er som følger.

- personer mellom 16 - 20 år tilhører aldersgruppe 1
- personer mellom 20 - 25 år tilhører aldersgruppe 2
- menn mellom 25 - 55 år tilhører aldersgruppe 3
- kvinner mellom 25 - 55 år tilhører aldersgruppe 4
- menn mellom 55 - 67 år tilhører aldersgruppe 5
- kvinner mellom 55 - 67 år tilhører aldersgruppe 6
- personer mellom 67 - 75 år tilhører aldersgruppe 7

Det foreligger tabeller basert på de 3 ulike koblingene av registre der disse er aggregert opp på alder. Det er kjørt ut en tabell for hver aldersgruppe da dette ser mest oversiktlig ut.

Tabellene er lagt i vedlegg B.

### 6.3. Kommentar til koblingen mellom AKU, AT og LTO

Tabell 2 nedenfor viser antall i de ulike gruppene gått ut i fra hovedstatuskoden i AKU. Det vil si at i den første tabellen ser vi fordelingen av registeropplysningene for de som ikke har svart i AKU, den neste for de som har svart at de er sysselsatt i AKU osv. Videre er det altså konstruert en binær variabel på bakgrunn av de 3 nevnte filene som altså teller opp arbeidforholdene basert på start- og stoppdatoene i AT- og LTO-registeret.

Tabell med AKU, AT og LTO variablene

- sysat er 1 dersom IO er registrert som sysselsatt kun i AT og ikke i LTO på intervjudtidspunktet 0 ellers
- Variabelen sysat er konstruert på bakgrunn av filen **ATULTO**, dvs individene som er registrert i AT men ikke i LTO.
- sylt er 1 dersom IO er registrert som sysselsatt både i AT og LTO på intervjudtidspunktet 0 ellers
- Variabelen sylt er konstruert på bakgrunn av filen **LTOMAT**, dvs individene som det foreligger kobling på både i AT og LTO.
- syslto er 1 dersom IO er registrert som sysselsatt kun i LTO og ikke i AT på intervjudtidspunktet 0 ellers
- Variabelen syslto er konstruert på bakgrunn av filen **LTOUAT**, dvs individene som er registrert i LTO men ikke finnes i AT.

**Tabell 1. Sammenheng mellom IO i AKU og deres registeropplysninger i AT og LTO**

	AKU				
	Sysselsatte	Ledige	Utenfor arb.sty	Netto	Frafall
<b>REGISTERE</b>					
1. AT og LTO	11 081	109	96	11 286	777
2. Kun AT	135	15	348	498	27
3. Kun LTO	1 023	205	1 213	2 441	197
4. Ingen	1 943	539	5 200	7 682	1 015
5. I alt	14 182	868	6 857	21 907	2 016
1. AT og LTO	78%	13%	1%	52%	39%
2. Kun AT	1%	2%	5%	2%	1%
3. Kun LTO	7%	24%	18%	11%	10%
4. Ingen	14%	62%	76%	35%	50%
5. I alt	100%	100%	100%	100%	100%

Betegnelsen ATULTO gjelder ukomplete arbeidstakerforhold hvor man ikke har oppnådd kopling mot et tilhørende lønnsforhold. Betegnelsen ATMLTO gjelder komplette arbeidstakerforhold hvor man har oppnådd kopling mot et tilhørende lønnsforhold. Betegnelsen LTOUAT gjelder ukomplete lønnsforhold hvor man ikke har oppnådd kopling mot et tilhørende arbeidstakerforhold. Det vil ikke være så interessant å trekke slutninger om de personene som kun har ukomplete lønnsforhold. AKU-informasjonen er knyttet til bestemte spørreuker, og variabelen hovedstatus er konstruert på bakgrunn av de svar intervjuobjektene gir på spørsmål om sin aktivitet(er) spørreuken. Datering av LTO-forhold på bak-

grunn av kun den informasjonen som finnes på LTO-oppgaven er ofte unøyaktig og i mange tilfeller så vil dette resultere i svært feilaktig datering. For komplette lønnsforhold anvender man derfor dateringen i AT-registeret. For de ukomplete lønnsforholdene har man beklageligvis kun dateringen i LTO-registeret.

#### **6.4. Feilkilder i tabellene**

Det finnes flere mulige forklaringer på inkonsistens mellom AKU, LTO- og AT-registeret. Inkonsistensen kan oppstå både hos AKU og hos registrene. Noe summarisk kan man si at disse forklaringene (de mest åpenbare) er:

- Manglende inn- og utmeldinger i AT-registeret.
- Forsinkede inn- og utmeldinger i AT-registeret.
- Manglende innrapportering av lønnsforhold (LTO-oppgavene).
- Unøyaktig eller manglende datering av lønnsforhold .
- Feil svar i AKU.
- Feilkodinger i AKU.
- Ulikheter i regelverk og definisjoner.

#### **6.5. Kommentarer til Tabell 1 med AKU, AT og LTO**

Det er 2 016 personer som ikke har svart i AKU. Av disse ser vi at omlag 50% (1 015) hverken finnes i ATULTO, ATMLTO eller LTOUAT, og at de derfor sannsynligvis har et meget marginalt forhold til arbeidsmarkedet. Av de restrende 1 001 personene, har 77,6% forhold i ATMLTO.

Vi ser altså at de personene som ikke har svart i AKU, har enten et koplet lønnstakerforhold (50%) eller så har de ikke noe forhold i det hele tatt (35%).

Det er 14 182 personer som har svart at de er sysselsatt i AKU.

Av disse 14 182 personene som har svart de er sysselsatte i AKU, ser vi at ca. 13,7% (1 943) hverken finnes i ATULTO, ATMLTO eller LTOUAT. 20,9% (2 966) finnes hverken i ATULTO eller ATMLTO. AKU baserer seg på ILO-definisjonen når det gjelder å definere en persons hovedstatus som sysselsatt. Kriteriet for å bli registrert i AT-registeret er strengere, så alle personene som har hovedstatus som sysselsatt behøver ikke å være registrert i AT-registeret. Mange av disse har heller ikke noen lønnsforhold (1 943 har heller ikke lønnsforhold). Vi må også regne med at noen av de som svarer at de er sysselsatt i AKU, er sysselsatt som selvstendig eller som familiearbeider. Disse personene vil vi stort sett ikke finne i noen av registrene. Unøyaktigheter og feil i AKU kan også forekomme. 1 023 av de personene som ikke er registrert i AT-registeret har et lønnsforhold. Dette kan skyldes ulikheter i regelverket, men vi vet også at AT-registeret inneholder mangler og unøyaktigheter. Av 14 182 personene som har svart de er sysselsatte i AKU, finner vi hele 77,3% i ATMLTO og 79,1% (11 216) i både ATMLTO og ATULTO.

Dette betyr at de som svarer at de er sysselsatte i AKU stort sett er registrert i AT-registeret, og de er registrert med hva en kan kalle «troverdige» forhold. Av de 11 216 personene som er registrert i AT-registeret, er det nemlig hele 98,4% som også har et lønnsforhold.

Det er 868 personer som svarer de er ledige i AKU. I utgangspunktet burde disse ikke være innmeldt i AT-registeret, men manglende utmelding kan muligens være et problem.

Av de 868 personene som har svart at de er ledige i AKU, er det 62,1% (539) som hverken er registrert som arbeidstaker i AT-registeret eller har et lønnsforhold i LTO-registeret (hverken i ATMLTO, ATULTO eller LTOUAT). Det er imidlertid hele 85,7% (744) som ikke er registrert i AT-

registeret (ATMLTO og ATULTO), men det er 23,6% (205) som kun er registrert med et lønnsforhold i LTOUAT. 12,6% er registrert i ATMLTO, mens kun 1,7% er registrert i ATULTO. Det er ingen personer som er registrert i både ATULTO og ATMLTO samtidig som de svarer at de er ledig i AKU.

Under 15% av de ledige er altså registrert i AT-registeret. Noen av disse kan ha en «småjobb» på si mens de går ledig, men siden de oppfatter seg som ledig vil deres hovedstatus fremstå som ledig. De fleste som svarer at de er ledige i AKU og er registrert i AT-registeret, finnes i ATMLTO, dvs. de har relativt «troverdige» forhold. Tar en i betraktning de forsinkelser og mangler som vi vet eksisterer i AT-registeret og det faktum at det faktisk kan forekomme «feil» i AKU, så virker dette som et rimelig avvik.

Det er 6 857 personer som sier de er utenfor arbeidsstyrken i AKU. Disse personene håper vi i størst mulig grad ikke finnes i AT- eller LTO-registeret. Vi antar at det finnes noen flere i LTO-registeret, siden disse LTO-forholdene kan skyldes etterbetalinger og utbetalinger ikke direkte knyttet til arbeid, for eksempel feriepenger.

Av disse 6 857 personene som har svart de er utenfor arbeidsstyrken i AKU, er det 75,8% som hverken er registrert som arbeidstaker i AT-registeret eller har et lønnsforhold i LTO-registeret. 1,4% (96) av de som sier de er utenfor arbeidsstyrken i AKU, har et koplet arbeidstakerforhold, mens 5,1% (348) har et ukoplet arbeidstakerforhold. 17,7% (1 213) har kun et ukoplet lønnsforhold. Disse tallene bekrefter våre antagelser, og avviket er ikke større enn hva vi må akseptere (forvente).

**Tabell 2. Sammenheng mellom IO i AKU og deres registeropplysninger i SOFA med prosentvis andel registrert i SOFA og de som ikke er registrert i SOFA**

	AKU				
	Sysselsatte	Ledige	Utenfor arb.sty	Netto	Frafall
<b>SOFA</b>					
Helt ledige	507	226	293	1 026	160
Syssels.tiltak	70	36	30	136	17
Kval.tiltak	189	122	223	534	67
Attføring	93	7	23	123	12
Annet	982	302	321	1 605	184
Ikke i SOFA	12 341	175	5 967	18 483	1 576
<b>I alt</b>	<b>14 182</b>	<b>868</b>	<b>6 857</b>	<b>21 907</b>	<b>2 016</b>
Helt ledige	4%	26%	4%	5%	8%
Syssels.tiltak	0%	4%	0%	1%	1%
Kval.tiltak	1%	14%	3%	2%	3%
Attføring	1%	1%	0%	1%	1%
Annet	7%	35%	5%	7%	9%
Ikke i SOFA	87%	20%	87%	84%	78%
<b>I alt</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>

Kommentarene til tabell 2 slås sammen med kommentarene til tabell 3 da de samme tallene foreligger i tillegg til sammenkoblingen med AT.

**Tabell 3. Sammenheng mellom IO i AKU og deres registeropplysninger i SOFA med prosentvis andel registrert i SOFA**

	AKU				
	Sysselsatte	Ledige	Utenfor arb.sty	Netto	Frafall
<b>SOFA</b>					
Helt ledige	507	226	293	1 026	160
Syssels.tiltak	70	36	30	136	17
Kval.tiltak	189	122	223	534	67
Attføring	93	7	23	123	12
Annet	982	302	321	1 605	184
Sum i SOFA	1 841	693	890	3 424	440
Helt ledige	28%	33%	33%	30%	31%
Syssels.tiltak	4%	5%	3%	4%	4%
Kval.tiltak	10%	18%	25%	16%	15%
Attføring	5%	1%	3%	4%	3%
Annet	53%	44%	36%	47%	42%
Sum i SOFA	100%	100%	100%	100%	100%

**Tabell 4. Sammenheng mellom IO i AKU og derers registeropplysninger i AT, SOFA**

	AKU								Frafall
	Sysselsatte		Ledige		Utenfor arb.sty		Netto		
	i AT	ikke i AT	i AT	ikke i AT	i AT	ikke i AT	i AT	ikke i AT	
<b>SOFA</b>									
Helt ledige	358	149	47	179	37	256	442	584	160
Syssels.tiltak	45	25	8	28	4	26	57	79	17
Kval.tiltak	95	94	7	115	12	221	114	430	67
Attføring	71	22	0	7	2	21	73	50	12
Annet	731	251	58	244	41	280	830	775	184
Ikke i SOFA	10 353	1 988	21	154	378	5 589	10 752	7 731	1 576
I alt	11 653	2 529	141	727	474	6 383	12 268	9 639	2 016
Helt ledige	3%	8%	31%	25%	7%	4%	4%	6%	8%
Syssels.tiltak	0%	2%	6%	4%	1%	1%	1%	1%	1%
Kval.tiltak	1%	6%	5%	16%	3%	3%	1%	5%	3%
Attføring	0%	1%	0%	1%	0%	0%	1%	0%	1%
Annet	6%	15%	40%	33%	8%	4%	7%	10%	9%
Ikke i SOFA	89%	68%	18%	21%	81%	88%	84%	78%	78%
I alt	100%	100%	100%	100%	100%	100%	100%	100%	100%

## 6.6. Kommentarer til Tabellen med AKU, AT og SOFA

Det er 2 016 personer som ikke har svart i AKU. (Kommentarene for frafallet henvises til vedlegg tabell A.5). Av disse ser vi at 43,8% (883) hverken finnes i AT- eller SOFA-søker-registeret. 34,4% (693) er registrert i AT-registeret, men finnes ikke i SOFA-søker-registeret. Dette betyr at over 78% av de som ikke har svart i AKU, finnes ikke i SOFA-søker-registeret. 7,9% (160) av de som ikke har svart i AKU finnes i SOFA-søker-registeret som helt ledige. Av disse 160 personene, finnes 26,9% registrert i AT-registeret som arbeidstakere.

Det er 14 182 (11 653 + 2 529) personer som har svart de er sysselsatt i AKU. Se også vedleggstabell A.6. Vi ønsker å finne de fleste av disse i AT-registeret men ikke SOFA-søker-registeret. Hvis de befinner seg i SOFA-søker-registeret, skal de ihvertfall ikke være registrert som helt ledig. Personer på sysselsettingstiltak skal regnes som sysselsatte i AKU.

Av de 14 182 personene som har svart de er sysselsatt i AKU, finner vi 73% (10 353) registrert i AT-registeret og ikke i SOFA-søker-registeret. 82,2% (11 653) er registrert i AT-registeret totalt. Det er 87% (12 341 = 10 353 + 1 988) som ikke er registrert i SOFA-søker-registeret, mens 13% (1 841) er registrert i SOFA-søker-registeret.

Av de 1 841 personene som er registrert i SOFA-søker-registeret og som har svart at de er sysselsatt i AKU, er det 27,5% (507) som er registrert helt ledig. Dette utgjør imidlertid kun 3,6% av alle de som har svart at de er sysselsatt i AKU. Det er 3,8% (70) av de 1 841 som er registrert på sysselsettingstiltak, mens det er 10,3% (189) som er registrert på kvalifiseringstiltak. Dette innebærer at det kun er 0,5% og 1,3% av alle de som svarer at de er sysselsatte, som er på hhv. sysselsettingstiltak og kvalifiseringstiltak.

Vi finner 14% (1 988) som hverken er registrert i AT-registeret eller i SOFA-søker-registeret.

Det er 9,2% (1 300) som både er registrert i AT-registeret og i SOFA-søker-registeret, men det er kun 2,5% (358) som er registrert i A-/A-registeret og som helt ledig i SOFA-søker-registeret. Disse 358 personene utgjør imidlertid 27,5% av de 1 300 personene som finnes i begge registrene. Den største gruppen her finner vi blant de personene som er registrert i SOFA-søker-registeret, men som ikke er på tiltak eller er helt ledig. Dette utgjør 56,2% av de 1 300. 3,5% av de 1 300 er registrert på sysselsettingstiltak. Dette betyr at det kun er 0,3% som både er registrert i AT-registeret og på sysselsettingstiltak.

Det er kun 3,8% (541) av de som har svart de er sysselsatt i AKU som ikke er registrert i AT-registeret mens de finnes i SOFA-søker-registeret. Av disse 541 personene finner vi 27,5% (149) registrert som helt ledig. Den største gruppen blant disse 541 personene er de som er registrert i SOFA-søker-registeret men som hverken er helt ledig eller på tiltak. Disse utgjør 46,4% (251).

Det er 868 (141 + 727) personer som svarer at de er ledige i AKU. Se vedlegg tabell A.7. Denne gruppen forventer vi skal være registrert i SOFA-søker-registeret enten som helt ledig eller som på kvalifiseringstiltak. De skal ihvertfall ikke være registrert på sysselsettingstiltak. De fleste skal heller ikke være registrert i AT-registeret.

Av de 868 personene som har svart de er ledig i AKU, finner vi 17,8% (154) hverken i AT- eller SOFA-søker-registeret. Vi finner kun 26% (226) registrert som helt ledig i SOFA-søker-registeret. Av de 226 personene som er registrert ledig, er det imidlertid kun 20,1% (47) som er registrert i AT-registeret samtidig.

Den største gruppen av de 868 personene som har svart at de er ledige i AKU, er de som er registrert i SOFA-søker-registeret men som ikke er ledig eller på tiltak. Disse utgjør 34,8% (302), men kun

19,2% (58) av disse 302 personene er registrert i AT-registeret. Dette utgjør 6,7% av alle de som har svart at de er ledig i AKU. Det er 14,6% (122) som er registrert på kvalifiseringstiltak. Disse er feilaktig klassifisert som ledige i AKU. Av disse 122 personene finner vi kun 5,7% (7) i AT-registeret. Det er kun 0,8% (7) som finnes på kvalifiseringstiltak. Ingen av disse er registrert i AT-registeret.

Det er 6 857 (474 + 6 383) personer som svarer at de er utenfor arbeidsstyrken i AKU. Se også tabell A.8. Denne gruppen forventer vi hverken i noen særlig grad skal være registrert i AT- eller i SOFA-søker-registeret. Det er grunn til å anta at vi relativt sett finner flere registrert i SOFA-søker-registeret enn i AT-registeret.

81,5% (5 589) av de som svarer i AKU at de er utenfor arbeidsstyrken, finnes hverken i AT- eller SOFA-søker-registeret. 1,4% (96) er både innmeldt i AT- og SOFA-søker-registeret. Det er 5,5% (378) som kun er innmeldt i AT-registeret, mens det 11,6% (794) er som kun er innmeldt i SOFA-søker-registeret.

Av de 6 857 personer som svarer at de er utenfor arbeidsstyrken i AKU, finner vi 13% (890) personer registrert i SOFA-søker-registeret. Av disse 890 personene finner vi flest (321) som hverken er registrert som helt ledig eller på tiltak. Dette gjelder uansett om vedkommende er registrert i AT-registeret. Den nest største gruppen finner vi blant de som er helt ledige (293 = 37 + 256). For begge disse gruppene ser vi at andelen som samtidig er registrert i AT-registeret utgjør ca. 14,5%.

Det virker som de fleste personene som oppgir i AKU at de er utenfor arbeidsstyrken, hverken finnes i AT-registeret eller SOFA-søker-registeret. Hvis de først skal være innmeldt i et register, er sjansen dobbelt så stor for at personen er innmeldt i SOFA-søker-registeret. Dette er som forventet. Det observerte avviket er innenfor hva vi må regne med.

**Tabell 5. Sammenheng mellom IO i AKU og derers registeropplysninger i AT, SOFA og LIKN m/MOMS**

	AKU								Frafall
	Sysseksatte		Ledige		Utenfor arb.sty		Netto		
	i reg.	ikke i reg	i reg	ikke i reg	i reg	ikke i reg	i reg	ikke i reg	
<b>SOFA</b>									
Helt ledige	377	128	51	175	40	253	468	556	160
Sysseks.tiltak	46	25	10	26	4	26	60	77	17
Kval.tiltak	100	89	8	114	15	208	123	411	67
Attføring	71	22	0	7	2	21	73	50	12
Annet	752	321	67	235	45	276	864	832	184
Ikke i SOFA	11 283	1 058	30	145	448	5 519	11 761	6 722	1 576
I alt	12 629	1 553	166	702	554	6 303	13 349	8 558	2 016
Helt ledige	3%	8%	31%	25%	7%	4%	4%	6%	8%
Sysseks.tiltak	0%	2%	6%	4%	1%	1%	1%	1%	1%
Kval.tiltak	1%	6%	5%	16%	3%	3%	1%	5%	3%
Attføring	0%	1%	0%	1%	0%	0%	1%	0%	1%
Annet	6%	15%	40%	33%	8%	4%	7%	10%	9%
Ikke i SOFA	89%	68%	18%	21%	81%	88%	84%	78%	78%
I alt	100%	100%	100%	100%	100%	100%	100%	100%	100%



Med sysselsatt i register menes de som er registrert som sysselsatt i AT og/eller LIKN m/MOMS. Tallene i tabell 4 avviker lite fra tabell 5. Forskjellen består i at det er tatt med selvstendighetsforhold som her er definert på bakgrunn av LIKN m/MOMS-registeret. Videre er selvstendighetsforholdene konstruert på en utprøvet variabel basert på en lav inntektsgrense. Registeret er som nevnt en årsfil der det ikke forekommer datoer på arbeidsforholdene. Konsistensen som måles blir ikke egnet for korttidsstatistikk ettersom likningsregisteret blir oppdatert ut i påfølgende år.

Forskjellen mellom tabell 4 og 5 er at vi får flere sysselsatte med selvstendighetsforhold i forhold til ikke sysselsatte som virker rimelig. Av de som har svart at de er sysselsatte i AKU, registreres 82,1% som sysselsatte og 17,8% som ikke sysselsatte i følge AT. Der det er tatt med selvstendighetsforhold finner vi blant de sysselsatte i AKU er 89,0% sysselsatte i følge registre og 11,0% er ikke sysselsatte.

### 6.7. Kommentarer til tabeller der det er gruppert på alder

Når det gjelder kommentarer til aldersfordelingen, blir disse lite omtalt. En finner som forventet at i den yngste aldersgruppen er usikkerheten på antall sysselsatte stor. Av de 424 blant 1 691 som har svart at de er sysselsatte i AKU, finnes 205 i LTOMAT, det vil si under 50%. Av tabell 22, som viser koblingen mellom AKU, AT, LTO og SOFA for den yngste aldersgruppen ser vi at av de 424 blant 1 691 som har svart at de er sysselsatt i AKU, finnes 232 i AT. I tabell 30 som viser samme aldersgruppe der det i tillegg er trukket inn selvstendighetsforhold, finnes 236 i AT eller LIKN m/MOMS av de samme 424 som har svart i AKU at de er sysselsatte.

## 7. Prediksjonsfunksjon av sysselsetting basert på registeropplysninger

De binære variablene omtalt over, er benyttet til å konstruere en prediksjonsmodell for å beregne registerbaserte sysselsettingstall som sammenlignes med AKU tall.

Følgende er utført:

Det er kjørt en regresjonsmodell med binomisk link der responsvariabelen for person nr.  $i$  er  $Y_i = 1$  dersom personen har svart sysselsatt i AKU og  $Y_i = 0$  dersom personen ikke er sysselsatt i følge AKU. Tilpasningene er gjennomført for personene som deltar i 1. kvartal og deres registeropplysninger. Forklaringsvariablene for person nr  $i$  er som følger:

$$x_{i,1} = \begin{cases} 1 & \text{dersom individet er definert som sysselsatt i følge AT og/eller LIKN} \\ & \text{m/MOMS på intervjudtidspunktet} \\ 0 & \text{ellers} \end{cases}$$

$$x_{i,2} = \begin{cases} 0 & \text{dersom individet ikke står som ledig på intervjudtidspunktet} \\ 1 & \text{dersom individet står som ledig på intervjudtidspunktet} \end{cases}$$

$$x_{i,3} = 1, 2, 3, 4, 5, 6, 7 \text{ ettersom individet tilhører en av de ulike aldersgruppene}$$

For følgende modeller er prøvd ut:

$$(1) \text{Logit}(P(Y_i = 1|x)) = \alpha + \beta_1 x_{i,1} + \beta_2 x_{i,2}$$

$$(2) \text{Logit}(P(Y_i = 1|x)) = \alpha + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3}$$

der  $\text{Logit}(p) = \ln [p/(1-p)]$ .

Ved bruk av modell (1) er det kjørt en regresjonsmodell for hver aldergruppe for seg, og ved modell (2) er aldersgruppene benyttet som en forklaringsvariabel.

Regresjonsanalysen er utført i SAS ved *proc probit* og kontrollert ved *proc catmod*, og det er blitt tilpasset regresjonsparametere ved at modellen er kjørt på personene som er spurt i AKU.

De estimerte regresjonsparameterene utnyttes nå til å predikere det totale sysselsettingstall for befolkningen. Den binomiske linken er følgende:

$$P(Y_i=1|x) = \frac{\exp(\mathbf{b}x_i)}{1 + \exp(\mathbf{b}x_i)}$$

der  $\mathbf{b}$  er vektoren med de ukjente parameterene som skal estimeres og  $x_i$  er vektoren av registeropplysninger for individ nr  $i$ .

Det registreres at i *proc probit* estimeres  $P(Y_i=1|x)$ . For hver person i registerne er det definert en  $x_{ij}$  for  $j=1, 2, 3$ . Forklaringsvariablene til personene som står oppført i registerne blir nå puttet inn i link-funksjonen. Det blir så estimert et tall mellom 0 og 1 for hver person, og disse blir summert sammen. Endelig blir totalt antall sysselsatte estimert som er et beregnet sysselsettingstall på bakgrunn av registeropplysningene inkludert de som spørres i AKU.

## 7.1. Resultat av prediksjonsfunksjon

Resultat fra modell (1) og modell (2) er tatt med og sammenlignet med de estimerte tallene fra AKU. Ettersom AT og LIKN m/MOMS er definert på jobber, vil det bli registrert flere arbeidsforhold og likningsopp-gaver for hvert enkelt individ. Det er foretatt selektering slik at det kun finnes et arbeidsforhold og/eller trekkopp-gave per personnummer. Dette gjøres fordi det skal estimeres antall sysselsatte. Her er altså de selvstendige definert ut i fra en pensjonsgivende inntekt på 20 000,- årlig som trolig er for lavt.

Først presenteres sysselsettingstallene basert på AKU i tabell 6, og i tabell 7 foreligger sysselsettingstallene fra prediksjonsmodellen.

**Tabell 6. Tabell over estimerte sysselsettingstall basert på AKU for 1. kvartal 1992 i tusen**

Aldersgruppe	Sysselsettingstall basert på AKU
16-19 GR. 1	61
20-24 GR. 2	191
M 25-54 GR. 3	780
K 25-54 GR. 4	661
M 55-66 GR. 5	142
K 55-66 GR. 6	113
67-74 GR. 7	32
Total syssels.	1 980

Her ser vi tallene basert på AKU fra 1. kvartal 1992. Tallene er beregnet ved å ta gjennomsnittet av de 3 første månedene.

Nedenfor følger tall fra beregningene:

**Tabell 7. Tabell over predikerte sysselsettingstall basert på AKU og registre fra 1992**

Alders- gruppe	Utv. i AKU brutto	Utv. i AKU netto	Syssels. i AKU	Totalpop. i reg.	Prediksjon 1.	Prediksjon 2.
GR. 1	1 691	1 620	424	237 489	78 869	82 739
GR. 2	2 564	2 346	1 363	347 093	217 222	228 032
GR. 3	7 098	6 366	5 580	969 107	769 778	769 622
GR. 4	6 601	6 107	4 681	890 443	662 577	661 485
GR. 5	1 748	1 589	1 053	227 485	113 278	137 279
GR. 6	1 831	1 696	870	232 317	110 901	110 785
GR. 7	2 390	2 174	211	324 873	29 653	29 542
Sum	23 923	21 907	13 971	3 228 807	1 982 278	2 019 484

I tabellen ser vi bruttoutvalget i AKU for 1. kvartal 1992. I neste kolonne ser vi antallet som har svart i de ulike aldersgruppene. Deretter ser vi antallet som har svart at de er sysselsatte, og videre er det sjekket hvor mange personer det foreligger i totalpopulasjonen i de ulike aldersgruppene. I prediksjon 1 er det kjørt en prediksjonsmodell der det er delt opp i aldersgrupper på forhånd og kjørt en regresjonsmodell for hver aldersgruppe, og i prediksjon 2 er de estimerte tallene der det er kjørt en prediksjonsmodell med alder som forklaringsvariabel.

Vi ser at totalt antall sysselsatte blir en forskjell på 37 206 mellom prediksjon 1 og prediksjon 2. De predikerte totaltallene ligger nær det totale sysselsettingstallet basert på AKU fra tabell 6, spesielt like er tallene fra prediksjon 1 sammenlignet med AKU's tall. Sammenligner vi de predikerte sysselsettingstallene med tallene basert på AKU innenfor hver aldersgruppe, ser vi at prediksjon 1 og 2 ligger henholdsvis 17 og 21 tusen høyere for gruppe 1. For gruppe 2 ligger prediksjonstallene henholdsvis 18 og 29 tusen høyere, mens tallene er svært like i aldersgruppe 3 og 4. I gruppe 5 ligger prediksjonstallene under med henholdsvis 32 og 5 tusen. For aldersgruppe 6 og 7 er prediksjonsmodellene og AKU-tallene tilnærmet like. Når det gjelder frafall i AKU, estimerer ikke prediksjonsmodellene for dette. Det gjør heller ikke AKU. Prediksjonsmodellen har kun tatt utgangspunkt i antall sysselsatte som har svart i AKU og modellerer uten å ta hensyn til frafallet. Det er vanskelig å konkludere noe ved prediksjonsmodellene, men modellene korrigerer mer for registervariable enn hva sysselsettingstallene basert på AKU gjør.

## Litt om filene

I analysen har vi brukt følgende filer:

AKU-fil på stormaskin:

P6245.S0211.B050n8A8.G92K1.V00 for 1. kvartal 92 og tilsvarende K2, K3, K4 for de øvrige kvartalene fra 92.

Registerfiler som er hentet fra masselageret på UNIX:

Arbeidstakerforhold med kobling til LTO: fil3.ssd01

LTO-forhold uten kobling til AT: fil1.ssd01

arbeidstakerforhold uten kobling til LTO: fil2.ssd01

SOFA:sofa\_u92.ssd01

Likning med moms-registeret: likmoms.ssd01

På filen sofa\_u92.ssd01 finnes blant annet variabelen tiltakskode som er kodet som følger:

0. Helt ledig

1. Arbeid for trygd

2. Andre offentlige sysselsettingstiltak

3. Lønnskudd

4. YFP/Praksisplasser

5. AMO

6. Andre1542-tiltak

7. 1543-tiltak, attføring

## Tabeller mellom AKU og registerne

Vedlegg B

**Tabell B.1. Tabell over de som ikke har svart i AKU. Registrert kun i LTO fra filen LTOUAT**

Registrert kun i AT fra filen ATULTO	Registrert i AT og LTO fra LTOMAT	0	1	Total
0	0	1 015	197	1 212
	1	641	122	763
1	0	23	4	27
	1	12	2	14
Total		1 619	325	2 016

**Tabell B.2. Tabell over de som har svart at de er sysselsatt i AKU Registrert kun i LTO fra filen LTOUAT**

Registrert kun i AT fra filen ATULTO	Registrert i AT og LTO fra LTOMAT	0	1	total
0	0	1 943	1 023	2 966
	1	9 031	1 928	10 959
1	0	88	47	135
	1	97	25	122
Total		11 159	3 020	14 182

**Tabell B.3. Tabell over de som har svart at de er ledige i AKU. Registrert kun i LTO fra filen LTOUAT**

Registrert kun i AT fra filen ATULTO	Registrert i AT og LTO fra LTOMAT	0	1	Total
0	0	539	205	744
	1	88	21	109
1	0	10	5	15
	1	0	0	0
Total		627	231	868

**Tabell B.4. Tabell over de som har svart at de er utenfor arbeidsstyrken i AKU. Registrert kun i LTO fra filen LTOUAT**

Registrert kun i AT fra filen ATULTO	Registrert i AT og LTO fra LTOMAT	0	1	Total
0	0	5 200	1 213	6 413
	1	74	17	91
1	0	276	72	348
	1	4	1	5
Total		5 554	1 303	6 857

**Tabell B.5. Tabell over de som ikke har svart i AKU. Ledighetsstatus fra SOFA**

Registrert fra AT	0	1	2	3	4	5	Total
0	117	9	62	6	112	883	1 189
1	43	8	5	6	72	693	827
Total	160	17	67	12	184	576	2 016

**Tabell B.6. Tabell over de som har svart at de er sysselsatt i AKU. Ledighetsstatus fra SOFA**

Registrert fra AT	0	1	2	3	4	5	Total
0	149	25	94	22	251	1 988	2 529
1	358	45	95	71	731	10 353	11 653
Total	507	70	189	93	982	12 341	14 182

**Tabell B.7. Tabell over de som har svart at de er ledige i AKU. Ledighetsstatus fra SOFA**

Registrert fra AT	0	1	2	3	4	5	Total
0	179	28	115	7	244	154	727
1	47	8	7	0	58	21	141
Total	226	36	122	7	302	157	868

**Tabell B.8. Tabell over de som har svart at de er utenfor arbeidsstyrken i AKU. Ledighetsstatus fra SOFA**

Registrert fra AT	0	1	2	3	4	5	Total
0	256	26	221	21	280	5 589	6 383
1	37	4	12	2	41	378	474
Total	293	32	223	23	321	5 967	6 857

**Tabell med AKU, AT, SOFA og LIKN m/MOMS.****Tabell B.9. Tabell over de som ikke har svart i AKU. Ledighetsstatus fra SOFA**

Registrert fra AT og LIKN	0	1	2	3	4	5	Total
0	114	9	59	6	108	805	1 101
1	46	8	8	6	76	771	915
Total	160	17	67	12	184	1 576	2 016

**Tabell B.10. Tabell over de som har svart at de er sysselsatt i AKU. Ledighetsstatus fra SOFA**

Registrert fra AT og LIKN	0	1	2	3	4	5	Total
0	128	25	89	22	321	1 058	1 553
1	377	46	100	71	752	11 283	12 629
Total	505	71	189	93	983	12 341	14 182

**Tabell B.11. Tabell over de som har svart at de er ledige i AKU. Ledighetsstatus fra SOFA**

Registrert fra AT og LIKN	0	1	2	3	4	5	Total
0	175	26	114	7	235	145	702
1	5	10	8	0	67	30	166
Total	226	36	122	7	302	175	868

**Tabell B.12. Tabell over de som har svart at de er utenfor arbeidsstyrken i AKU. Ledighetsstatus fra SOFA**

Registrert fra AT	0	1	2	3	4	5	Total
0	253	26	208	21	276	5 519	6 303
1	40	4	15	2	45	448	554
Total	293	32	223	23	321	5 967	6 857

**Tabell B.13. Tabell for totalen mellom AKU, AT og LTO. Ikke registrert i AT og LTO**

Registrert i AT og LTO						
Alle status-grupper i AKU	Reg. i AT	Arb.f-hold ikke regist. i LTO	Arb.-hold reg. i LTO	Arbeidsforhold ikke registrert i LTO	Arbeidsforhold registrert i LTO	Total
0	0	1 015	197	641	2 171	1 975
	1	23	4	12	122	41
1	0	1 943	1 023	9 031	1 928	13 925
	1	88	47	97	25	257
2	0	539	205	88	21	853
	1	10	5			15
3	0	5 200	1 213	276	72	6 761
	1	74	17	4	1	96
Total		8 892	2 711	10 149	2 171	23 923

**Tabell B.14. Tabell for aldersgruppe 1 mellom AKU, AT og LTO. Ikke reg. i AT og LTO.**

Registrert i AT og						
Alle status-grupper i AKU	Reg. i AT	Arb. ikke reg. i LTO	Arb. reg. i LTO	Arb. ikke reg. i LTO	Arb. reg i LTO	Total
0	0	45	15	10	1	71
	1					
1	0	99	115	165	40	419
	1	1	1	1	2	5
2	0	60	22	2	1	85
	1	1				1
3	0	759	263	56	15	1 103
	1	5	2			7
Total		980	418	234	59	169

**Tabell B.15. Tabell for aldersgruppe 2 med AKU, AT og LTO. Ikke registrert i AT og LTO.  
Registrert i AT og LTO**

Alle status- grupper i AKU	Reg. i AT	Arb. ikke reg. i LTO	Arb. reg. i LTO	Arb. ikke reg. i LTO	Arb. reg i LTO	Total
0	0	115	29	56	16	216
	1			1	1	2
1	0	306	180	672	170	1 328
	1	7	8	14	6	35
2	0	133	47	24	5	209
	1	2	2			4
3	0	428	234	62	28	752
	1	10	5	3		18
Total		1 001	505	832	226	2 564

**Tabell B.16. Tabell for aldersgruppe 3 mellom AKU, AT og LTO. Ikke reg. i AT og LTO.  
Registrert i AT og LTO**

Alle status- grupper i AKU	Reg. i AT	Arb. ikke reg. i LTO	Arb. reg. i LTO	Arb. ikke reg. i LTO	Arb. reg i LTO	Total
0	0	308	65	276	52	701
	1	11	2	8	1	22
1	0	703	325	3 597	847	5 472
	1	47	21	30	10	108
2	0	182	76	34	9	301
	1	6	2			8
3	0	348	89	30	5	472
	1	11	2	1		14
Total		1 616	582	3 976	924	7 098

**Tabell B.17. Tabell for aldersgruppe 4 mellom AKU, AT og LTO. Ikke reg. i AT og LTO.  
Registrert i AT og LTO**

Alle status- grupper i AKU	Reg. i AT	Arb. ikke reg. i LTO	Arb. reg. i LTO	Arb. ikke reg. i LTO	Arb. reg i LTO	Total
0	0	217	41	185	41	484
	1	6	1	3		10
1	0	510	258	3 178	659	4 605
	1	18	12	42	4	76
2	0	119	46	19	6	190
	1		1			1
3	0	931	208	53	18	1 210
	1	18	6		1	25
Total		1 819	573	3 480	729	6 601



**Tabell B.18. Tabell for aldersgruppe 5 mellom AKU, AT og LTO. Ikke reg. i AT og LTO.  
Reg. i AT og LTO**

Alle status- grupper i AKU	Reg. i AT	Arb. ikke reg. i LTO	Arb. reg. i LTO	Arb. ikke reg. i LTO	Arb. reg i LTO	Total
0	0	73	14	61	10	158
	1		1			1
1	0	168	66	670	128	1 032
	1	11	2	6	2	21
2	0	33	9	5		47
	1	1				1
3	0	361	100	15	3	479
	1	8	1			9
Total		655	193	757	143	1 748

**Tabell B.19. Tabell for aldersgruppe 6 mellom AKU, AT og LTO. Ikke reg. i AT og LTO.  
Reg. i AT og LTO**

Alle status- grupper i AKU	Reg. i AT	Arb. ikke reg. i LTO	Arb. reg. i LTO	Arb. ikke reg. i LTO	Arb. reg i LTO	Total
0	0	77	9	44	1	131
	1	4				4
1	0	85	38	673	65	861
	1	2	3	4		9
2	0	9	3	4		16
	1					
3	0	689	86	25	1	801
	1	9				9
Total		875	139	750	67	1 831

**Tabell B.20. Tabell for aldersgruppe 7 mellom AKU, AT og LTO. Ikke reg. i AT og LTO.  
Reg. i AT og LTO**

Alle status- grupper i AKU	Reg. i AT	Arb. ikke reg. i LTO	Arb. reg. i LTO	Arb. ikke reg. i LTO	Arb. reg i LTO	Total
0	0	180	24	9	1	214
	1	2				2
1	0	72	41	76	19	208
	1	2			1	3
2	0	3	2			5
	1					
3	0	1 674	233	35	2	1 944
	1	3	1			14
Total		1 946	301	120	23	2 390

**Tabell for AKU, AT og SOFA gruppert på alder.**

**Tabell B.21. Tabellen er for totalen og uavhengig av aldersfordeling. Ledighetstatus i SOFA**

Statusgr. i AKU	Registrert i AT	0	1	2	3	4	5	Total
0	0	117	9	62	6	112	883	1 189
	1	43	8	5	6	72	693	827
1	0	149	25	94	22	251	1 988	2 529
	1	358	45	95	71	731	10 353	11 653
2	0	179	28	115	7	244	154	727
	1	47	8	7		58	21	141
3	0	256	26	211	21	280	5 589	6 383
	1	37	4	12	2	41	378	474
Total		1 186	153	601	135	1 789	20 059	23 923

**Tabell B.22. Tabell for aldersgruppe 1 mellom AKU, AT, LTO og SOFA. Ledighetstatus i SOFA**

Statusgr. i AKU	Registrert i AT	0	1	2	3	4	5	Total
0	0	3		3		5	49	60
	1	3		1			7	11
1	0	8	2	22	2	11	147	192
	1	8		12		7	205	232
2	0	4		16	2	4	56	82
	1	1					3	4
3	0	34	1	78	1	40	872	1 026
	1	3		2		3	76	84
Total		64	3	134	5	70	1 415	1 691

**Tabell B.23. Tabell for aldersgruppe 2 mellom AKU, AT, LTO og SOFA. Ledighetstatus i SOFA**

Statusgr. i AKU	Registrert i AT	0	1	2	3	4	5	Total
0	0	36	2	19	2	18	60	137
	1	9	2	3	1	13	53	81
1	0	64	12	45	2	76	201	400
	1	90	14	30	6	137	686	963
2	0	41	6	47		59	21	174
	1	7	2	4		21	5	39
3	0	69	11	63	2	64	445	654
	1	14	2	3		9	88	116
Total		330	51	214	13	397	1 559	2 564

**Tabell B.24. Tabell for aldersgruppe 3 mellom AKU, AT, LTO og SOFA. Ledighetstatus i**

		SOFA						
Statusgr. i AKU	Registrert i AT	0	1	2	3	4	5	Total
0	0	52	6	27	2	56	223	366
	1	20	5	1	2	30	229	357
1	0	50	5	13	10	86	723	887
	1	145	22	23	36	220	4 247	4 693
2	0	73	14	24	4	112	27	254
	1	25	4	1		20	5	55
3	0	45	8	33	11	70	262	429
	1	9	1	3		8	36	57
Total		419	65	125	65	602	5 822	7 098

**Tabell B.25. Tabell for aldersgruppe 4 mellom AKU, AT, LTO og SOFA. Ledighetstatus i**

		SOFA						
Statusgr. i AKU	Registrert i AT	0	1	2	3	4	5	Total
0	0	16	1	12	1	27	193	250
	1	6	1		1	24	212	244
1	0	23	4	14	5	68	523	637
	1	78	6	29	16	296	3 619	4 044
2	0	42	3	26	1	51	36	159
	1	9	1	2		12	8	32
3	0	85	4	36	5	95	908	1 133
	1	11	1	4		17	67	102
Total		270	21	123	31	590	5 566	23 923

**Tabell B.26. Tabell for aldersgruppe 5 mellom AKU, AT, LTO og SOFA. Ledighetstatus i**

		SOFA						
Statusgr. i AKU	Registrert i AT	0	1	2	3	4	5	Total
0	0	8		1	1	6	70	86
	1	3			1	5	64	73
1	0	3	2		1	4	197	207
	1	26	1		11	31	777	846
2	0	15	4	1		13	8	41
	1	4	1			2		7
3	0	13	1		1	5	441	461
	1					2	25	27
Total		72	9	2	15	68	1 582	1 748

**Tabell B.27. Tabell for aldersgruppe 6 mellom AKU, AT, LTO og SOFA. Ledighetstatus i SOFA**

Statusgr. i AKU	Registrert i AT	0	1	2	3	4	5	Total
0	0	2					83	85
	1	2			1		47	50
1	0				2	5	95	102
	1	11	2	1	2	39	713	768
2	0	3	1	1		5	2	12
	1	1				3		4
3	0	6	1	1		6	761	775
	1					2	33	35
Total		25	4	3	5	60	1 734	1 831

**Tabell B.28. Tabell for aldersgruppe 7 mellom AKU, AT, LTO og SOFA. Ledighetstatus i SOFA**

Statusgr. i AKU	Registrert i AT	0	1	2	3	4	5	Total
0	0						205	205
	1						11	11
1	0	1				1	102	103
	1					1	106	107
2	0	1					4	5
	1							
3	0	4			1		1 900	1 905
	1						53	53
Total		6			1	2	2 381	2 390

**Tabell for AKU, AT, LTO, LIKN m/MOMS og SOFA gruppert på alder.**

**Tabell B.29. Tabellen er for totalen og uavhengig av aldersfordeling. Ledighetstatus i SOFA**

Statusgr. i AKU	Registrert i AT og LIKN	0	1	2	3	4	5	Total
0	0	114	9	59	6	108	805	1 101
	1	46	8	8	6	76	771	915
1	0	128	25	89	22	231	1 058	1 553
	1	377	46	100	71	752	11 283	12 629
2	0	175	26	114	7	235	145	702
	1	51	10	8		67	30	166
3	0	253	26	208	21	276	5 519	6 303
	1	40	4	15	2	45	448	554
Total		1 184	154	601	135	1 790	20 059	23 923

**Tabell B.30. Tabell for aldersgruppe 1 mellom AKU, AT, LTO, LIKN m/MOMS og SOFA.  
Ledighetstatus i SOFA**

Statusgr. i AKU	Registrert i AT og LIKN	0	1	2	3	4	5	Total
0	0	3		3		5	49	60
	1	3		1			7	11
1	0	7	2	22	2	11	144	188
	1	9		12		7	208	236
2	0	4		16	2	4	56	82
	1	1					3	4
3	0	34	1	76	1	39	870	1 021
	1	3		4		4	78	89
Total		64	3	134	5	70	1 415	1 691

**Tabell B.31. Tabell for aldersgruppe 2 mellom AKU, AT, LTO, LIKN m/MOMS og SOFA.  
Ledighetstatus i SOFA**

Statusgr. i AKU	Registrert i AT og LIKN	0	1	2	3	4	5	Total
0	0	34	2	18	2	18	57	131
	1	11	2	4	1	13	56	87
1	0	64	12	45	2	75	187	385
	1	90	14	30	6	138	700	978
2	0	41	6	47		57	21	172
	1	7	2	4		23	5	41
3	0	68	11	63	2	64	443	651
	1	15	2	3		9	90	119
Total		330	51	214	13	397	1 559	2 564

**Tabell B.32. Tabell for aldersgruppe 3 mellom AKU, AT, LTO, LIKN m/MOMS og SOFA.  
Ledighetstatus i SOFA**

Statusgr. i AKU	Registrert i AT og- LIKN	0	1	2	3	4	5	Total
0	0	51	6	25	2	53	173	310
	1	21	5	3	2	33	349	413
1	0	34	5	10	10	72	187	318
	1	160	22	26	36	235	4 783	5 262
2	0	72	13	23	4	106	24	242
	1	26	5	2		26	8	67
3	0	44	8	33	11	68	249	413
	1	10	1	3		10	49	73
Total		418	65	125	65	603	5 822	7 098

**Tabell B.33. Tabell for aldersgruppe 4 mellom AKU, AT, LTO, LIKN m/MOMS og SOFA.  
Ledighetstatus i SOFA**

Statusgr. i AKU	Registrert i AT og LIKN	0	1	2	3	4	5	Total
0	0	16	1	12	1	26	182	238
	1	6	1		1	25	223	256
1	0	23	4	12	5	65	338	447
	1	77	7	31	16	299	3 804	4 234
2	0	41	3	26	1	50	34	155
	1	10	1	2		13	10	36
3	0	84	4	35	5	94	894	1 116
	1	12	1	5	2	18	81	119
Total		269	22	123	31	590	5 566	6 601

**Tabell B.34. Tabell for aldersgruppe 5 mellom AKU, AT, LTO, LIKN m/MOMS og SOFA.  
Ledighetstatus i SOFA**

Statusgr. i AKU	Regist- rert i AT og LIKN	0	1	2	3	4	5	Total
0	0	8		1	1	6	62	78
	1	3			1	5	72	81
1	0		2		1	2	63	68
	1	29	1		11	33	911	985
2	0	13	3	1		13	4	34
	1	6	2			2	4	14
3	0	13	1		1	5	429	449
	1					2	37	39
Total		72	9	2	15	68	1 582	1 748

**Tabell B.35. Tabell for aldersgruppe 6 mellom AKU, AT, LTO, LIKN m/MOMS og SOFA.  
Ledighetstatus i SOFA**

Statusgr. i AKU	Registrert i AT og LIKN	0	1	2	3	4	5	Total
0	0	2					80	82
	1	2			1		50	53
1	0				2	5	58	65
	1	11	2	1	2	39	750	805
2	0	3	1	1		5	2	12
	1	1				3		4
3	0	6	1	1		6	754	768
	1					2	40	42
Total		25	4	3	5	60	1 734	1 831

**Tabell B.36. Tabell for aldersgruppe 7 mellom AKU, AT, LTO, LIKN m/MOMS og SOFA.  
Ledighetstatus i SOFA**

Statusgr. i AKU	Registrert i AT og LIKN	0	1	2	3	4	5	Total
0	0						202	202
	1						14	14
1	0	1				1	81	82
	1					1	127	129
2	0	1					4	5
	1							
3	0	4			1		1 880	1 885
	1						73	73
<b>Total</b>		<b>6</b>			<b>1</b>	<b>2</b>	<b>2 381</b>	<b>2 390</b>

## Avhengighetsmålet $\gamma$

Det er estimert en  $\gamma$  til hver tabell ovenfor og alle  $\gamma$ 'ene presenteres og tolkes:

For Tabell B.1-A.4, dvs. for AKU, AT og LTO estimeres følgende verdi på  $\gamma$ :

Gitt hovedstatus i AKU	verdi av $\gamma$
frafall	-0.081
sysselsatt	0.168
ledige	0.230
utenfor arbeidsstyrkeen	0.370

Videre er  $\gamma$  beregnet for tabeller som inneholder opplysninger fra AKU, AT og SOFA og AKU, AT, SOFA og LIKN m/MOMS. Tallene følger nedenfor:

Av Tabell B.5-A.8, for AKU, AT og SOFA, fås følgende estimater for  $\gamma$ :

Gitt hovedstatus i AKU	verdi av $\gamma$
frafall	0.912
sysselsatt	0.912
ledige	0.689
utenfor arbeidsstyrkeen	0.758

Av Tabell B.9-A12, dvs. fra AKU, AT, SOFA og LIKN m/MOMS fås følgende estimater for  $\gamma$ :

Gitt hovedstatus i AKU	verdi av $\gamma$
frafall	0.912
sysselsatt	0.912
ledige	0.689
utenfor arbeidsstyrkeen	0.758

### Tolkning av $\gamma$

Dersom  $\gamma > 0$ , er sannsynligheten for at registerverdien fra begge registrene stemmer gitt AKU opplysninger, større enn sannsynligheten for at registerverdien fra begge registrene ikke stemmer gitt AKU opplysninger. Det er altså en positiv sannsynlighet for at det er konsistens mellom registre og AKU.

Når det gjelder  $\gamma$ -tallene er det konstruert en binær variabel på bakgrunn av AT og LIKN m/MOMS slik at det fås en binær registervariabel på bakgrunn av SOFA og en fra AT og LIKN m/MOMS. Det kjøres så ut 2x2 tabeller basert på hver hovedstatuskode fra AKU, og tolkningen blir som over.

Hvor stor positiv sannsynlighet det må være for at man skal kunne si at konsistensen er god nok, bør bestemmes på forhånd. Målet kan brukes som et kriterium for hvorvidt registeropplysninger mot svar i spørreundersøkelser skal kunne sies å være i samsvar eller ikke.

### Beskrivelse av hvorledes $\gamma$ er beregnet

Det skal først defineres celledens sannsynligheter:  $\pi_{11}$ ,  $\pi_{12}$ ,  $\pi_{21}$ ,  $\pi_{22}$



**Tabell C.1.**

		Register 1		
Register 2		sysselsatt	ledig	sum
sysselsatt		$n_{11}$	$n_{12}$	$n_{1.}$
ledig		$n_{21}$	$n_{22}$	$n_{2.}$
sum		$n_{.1}$	$n_{.2}$	$n$

Her er  $n$  er det totale antallet registrerte, og  $n_{ij}$  er antallet i hver kategori. Cellesannsynlighetene fremkommer ved  $\pi_{11} = n_{11} / n$  og tilsvarende for de øvrige cellesannsynlighetene.

Av definisjon (5.1) kan  $\gamma$  videre uttrykkes følgende:

$$\begin{aligned}
 \gamma &= P((X_1 - X_2)(Y_1 - Y_2) > 0 \mid X_1 \neq X_2, Y_1 \neq Y_2, Z_1 = Z_2) \\
 &\quad - P((X_1 - X_2)(Y_1 - Y_2) < 0 \mid X_1 \neq X_2, Y_1 \neq Y_2, Z_1 = Z_2) \\
 &= P(X_1 = Y_1 \cap X_2 = Y_2 \cap X_1 \neq X_2 \cap Y_1 \neq Y_2) / P(X_1 \neq X_2 \cap Y_1 \neq Y_2) \\
 &\quad - P(X_1 \neq Y_1 \cap X_2 \neq Y_2 \cap X_1 \neq X_2 \cap Y_1 \neq Y_2) / P(X_1 \neq X_2 \cap Y_1 \neq Y_2) \\
 &= [P(X_1 = Y_1 = 1 \cap X_2 = Y_2 = 0) + P(X_1 = Y_1 = 0 \cap X_2 = Y_2 = 1)] / P(X_1 \neq X_2 \cap Y_1 \neq Y_2) \\
 &\quad [ -P(X_1 = Y_2 = 1 \cap X_2 = Y_2 = 0) + P(X_2 = Y_1 = 1 \cap X_2 = Y_2 = 0) ] / P(X_1 \neq X_2 \cap Y_1 \neq Y_2) \\
 &= (\pi_{11} \pi_{22} + \pi_{22} \pi_{11}) / P(X_1 \neq X_2 \cap Y_1 \neq Y_2) \\
 &\quad - (\pi_{12} \pi_{21} + \pi_{21} \pi_{12}) / P(X_1 \neq X_2 \cap Y_1 \neq Y_2) \\
 &= 2(\pi_{11} \pi_{22} + \pi_{12} \pi_{21}) / P(X_1 \neq X_2 \cap Y_1 \neq Y_2)
 \end{aligned}$$

der

$$\begin{aligned}
 &P(X_1 \neq X_2 \cap Y_1 \neq Y_2) \\
 &= P(X_1 = 1, Y_1 = 1) P(X_1 = 0, Y_1 = 0) \\
 &\quad + P(X_1 = 1, Y_1 = 0) P(X_1 = 0, Y_1 = 1) \\
 &\quad + P(X_1 = 0, Y_1 = 1) P(X_1 = 1, Y_1 = 0) \\
 &\quad + P(X_1 = 0, Y_1 = 0) P(X_1 = 1, Y_1 = 1) \\
 &= \pi_{11} \pi_{22} + \pi_{12} \pi_{21} + \pi_{21} \pi_{12} + \pi_{22} \pi_{11} \\
 &= 2(\pi_{11} \pi_{22} + \pi_{12} \pi_{21})
 \end{aligned}$$

Dette gir at

$$\gamma = (\pi_{11} \pi_{22} - \pi_{12} \pi_{21}) / (\pi_{11} \pi_{22} + \pi_{12} \pi_{21}).$$

Av dette kan  $\gamma$  estimeres ved hjelp av det registrerte antallet i hver rute i de ulike 2 x 2- tabellene.

## Generell utvidelse av $\gamma$

Dersom register 1 og register 2 (som kan defineres som tidligere med faktor A og faktor B) har flere mulige utfall hhv.  $(A_1, \dots, A_p)$  og  $(B_1, \dots, B_q)$ , kan det vises at  $\gamma$  beregnes på følgende vis:

Vi kan skrive

$$\gamma = (\pi_s - \pi_d) / (1 - \pi) = (\pi_s - \pi_d) / (\pi_s + \pi_d)$$

der

$$\pi_t = P(X_1 = X_2 \cup Y_1 = Y_2)$$

$$\pi_s = P((X_1 - X_2)(Y_1 - Y_2) > 0)$$

$$\pi_d = P((X_1 - X_2)(Y_1 - Y_2) < 0),$$

og

$$\pi_t + \pi_s + \pi_d = 1.$$

Registervariablene kan nå betraktes som vektorer med flere enn 2 utfall.

Videre er  $\pi_s$  er sannsynligheten for at de variable som stemmer, og  $\pi_d$  er sannsynligheten for de som ikke stemmer.

Det kan vises at :

$$\begin{aligned}\pi_t &= \sum_{i=1}^v p_i^2 + \sum_{j=1}^w p_j^2 - \sum_{i=1}^v \sum_{j=1}^w p_{ij}^2 \\ \pi_s &= 2 \sum_{i=1}^{v-1} \sum_{j=1}^{w-1} p_{ij} \{ \sum_{i'>i} \sum_{j'>j} p_{i'j'} \} \\ \pi_d &= 2 \sum_{i=1}^{v-1} \sum_{j=2}^w p_{ij} \{ \sum_{i'>i} \sum_{j'<j} p_{i'j'} \} \end{aligned}$$

der  $p_{ij}$  er celsesannsynlighetene  $\pi_{ij}$  og v og w er antall faktorer for hhv. registervariabel 1 og 2. Dette målet er regnet ut for tabellene også der det er gruppert på alder.

## Om programmer

### Dokumentasjon av filer til tabellkjøring mellom AKU 1. kvartal 1992 og registre fra 92

Kobl4\_2.saskobler registre og AKU og lager nødvendige sysselsettingsvariable basert på registerne AT, LTO, SOFA og LIKN m/MOMS.

ibsel.sassorterer og selekterer slik at vi får en record per person på filen.

aldgr4.sas lager aldersgrupper på datasett som hentes inn i programmet.

#### Datasett:

tabmlm.ssd01 datasett med AKU, AT, LTO, SOFA og LIKN m/MOMS.

tablto1b.ssd01 datasett med AKU, AT og LTO.

aldgr2.ssd01 datasett med aldersgrupper for AKU, AT og LTO.

aldgr3.ssd01 datasett med aldersgrupper for AKU, AT, LTO og SOFA.

aldgr4.ssd01 datasett med aldersgrupper for AKU, AT, LTO, SOFA og LIKN m/MOMS.

#### Filoversikt for prediksjon

Programmene kjøres i rekkefølgen nedenfor:

tu2502a.sas kobler filer og lager datasett for hele populasjonen.

ibsel.sas som over

aldgr4.sas lager aldersgruppe, som over.

fr0702a.sas predikerer sysselsettingstall og skriver ut disse.

## De sist utgitte publikasjonene i serien Notater

- 97/6 E. Gulløy, S. Blom og A.A. Ritland: Levekår blant innvandrere 1996: Dokumentasjonsrapport med tabeller. 205s.
- 97/7 S. Blom og A.A. Ritland: Levekår blant innvandrere 1996: Del 2: Tabeller for nordmenn. 1997. 222s.
- 97/8 T.C. Mykkelbost: Resultater fra brukerundersøkelse i forbindelse med NOS 306: Utslipp til luft i norske kommuner 1993. 21s.
- 97/9 H.M. Teigum: Omnibusundersøkelsene 1996: Dokumentasjonsrapport. 136s.
- 97/10 P.O. Lande og T. Heimdal: GERIX START: System- og brukardokumentasjon. 49s.
- 97/11 A. Barstad: Frihetens århundre? Levekår i Norge i et 100-årsperspektiv. 37s.
- 97/12 G. Sparby: Inntekts- og formuesundersøkelsen 1992: Dokumentasjon. 101s.
- 97/13 V. Pedersen: Inntekts- og formuesundersøkelsen 1993: Dokumentasjon. 94s.
- 97/14 V. Pedersen: Inntekts- og formuesundersøkelsen 1994: Dokumentasjon. 93s.
- 97/15 Metodevalg og kostnader ved etablering og drift av et boligregister. 29s.
- 97/16 K. Vassenden: Innvandrerstatisikkprosjektet: Styringsgruppas evaluering. 34s.
- 97/17 E. Heilund, A.M.K. Holmøy, A.C. Steen og B.L. Western: Beregning av vektorer til inntekts- og formuesundersøkelsene 1994. 30s.
- 97/18 S. Størmer: Sammenhengen mellom sysselsetting i AKU og registrene: Måling av kvalitet og prediksjon. 36s.
- 97/19 H.M. Teigum: Verdiundersøkelsen 1996: Dokumentasjonsrapport. 84s.
- 97/20 T. Ouren og T. Vik: Prosjektrapport: Voksenopplæringsprosjektet 1995-1996. 24s.
- 97/21 S. Hansen og T. Skoglund: Beregning av sysselsetting og lønn i nasjonalregnskapet. 36s.
- 97/22 H. Lövkvist: Standardiserte rater - en metodebeskrivelse med eksempler fra dødsårsaksstatistikken. 45s.
- 97/23 T. Vik og T. Ouren: Voksenopplæringsprosjektet: Dokumentasjon av databasen og rutiner. 34s.
- 97/24 A.S. Andersen: Gerix-data: Gir de grunnlag for å vurdere inntektssystemet for kommunene? 58s.
- 97/25 Ø. Amundrud og I. Tuveng: Utredning av utvalgsplan for sentral sykefraværstatistikk. 36s.
- 97/26 J. Lyngstad: Innvandreres demografi og levekår. 38s.
- 97/27 L. Rogstad, P. Schøning, M.V. Dysterud og S. Homstvedt: Arealstatistikk i Norge: Resultater fra en brukerundersøkelse. 30s.
- 97/28 H.N. Næsheim og I. Tuveng: Muligheter for å få yrkesdata i registerbaserte statistikker. 37s.
- 97/29 J. Rodriguez: Sesongjustering i praksis - en innføring: April 1997. 71s.
- 97/30 K-G. Lindquist: Database for energiintensive næringer: Tall fra industristatistikken. 17s.
- 97/32 A. Mathiassen: Valg og bruk av internasjonale statistikkilder. 14s.
- 97/33 L.T. Foss og A.C. Steen: Inntekts- og kostnadsundersøkelse for privatpraktiserende leger 1995: Dokumentasjon. 56s.
- 97/34 P.E. Gjedtjernet: Inntekts- og formuesundersøkelsen for selskaper skattlagt med hjemmel i petroleumsskatteloven for årene 1991, 1992 og 1993. Dokumentasjon. 41s.
- 97/35 A. Langørgen: Faktorer bak variasjoner i kommunal ressursbruk til pleie og omsorg. 19s.
- 97/36 S.E. Førre: Registerdataene i lys av industristatistikken. 21s.
- 97/37 K. Gimming: Virkninger på prisutviklingen på naturgass i Vest-Europa ved innføring av felles karbonavgift. 40s.

## Notater

**B** *Returadresse:*  
Statistisk sentralbyrå  
Postboks 8131 Dep.  
N-0033 Oslo

Statistisk sentralbyrå

*Oslo:*  
Postboks 8131 Dep.  
0033 Oslo

Telefon: 22 86 45 00  
Telefaks: 22 86 49 73

*Kongsvinger:*  
Postboks 1260  
2201 Kongsvinger

Telefon: 62 88 50 00  
Telefaks: 62 88 50 30

ISSN 0806-3745



**Statistisk sentralbyrå**  
Statistics Norway