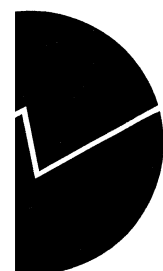


*Magne Bråthen og Johan Fosen*

**Definisjon av sysselsetting  
basert på registerinformasjon**  
Utarbeidelse av klassifikasjonsrutine

Notater



# Innhold

|   |    |
|---|----|
| INNOLD .....  | 2  |
| 1. INNLEDNING.....  | 4  |
| 2. FORMÅLET MED PROSJEKTET .....  | 5  |
| 2.1 MOTIVASJON .....  | 5  |
| 2.2 TILRETTELEGGING AV DATAMATERIALET FRA DE ADMINISTRATIVE REGISTRENE.....                               | 6  |
| 2.3 UTARBEIDELSE AV KLASIFIKASJONSROUTINEN: SAMSVARSANALYSER MELLOM REGISTRENE OG AKU PÅ INDIVIDNIVÅ..... | 6  |
| 2.4 KRAV TIL DEN FERDIGE KLASIFIKASJONSROUTINEN.....  | 7  |
| 2.5 ANALYSE AV PERSONER SOM ER SYSSELSATTE I AKU, MEN ER UTEN ARBEIDSFORHOLD IFØLGE REGISTRENE.....       | 7  |
| 3. REGISTERGRUNNLAG .....   | 8  |
| 3.1 ADMINISTRATIVE REGISTRE BENYTTET I ANALYSEN. ....   | 8  |
| 3.1.1 <i>Arbeidstaker- / Arbeidsgiverregster (AT/AG)</i> . ....   | 8  |
| 3.1.2 <i>Lønns- og trekkoppgave registeret (LTO)</i> .....  | 8  |
| 3.1.3 <i>Likningsregisteret</i> . ....  | 8  |
| 3.1.4 <i>SOFA-søker-registeret</i> . ....   | 8  |
| 3.2 TILRETTELEGGING AV DATA. ....   | 8  |
| 3.2.1 <i>Utarbeidelse av datafiler</i> .....  | 8  |
| 3.2.2 <i>Valg av viktigste arbeidstakerforhold</i> .....  | 9  |
| 3.2.3 <i>Konsistensbehandling mot SOFA-søker-registeret</i> .....   | 9  |
| 3.2.4 <i>Valg av viktigste arbeidsforhold</i> . ....  | 9  |
| 4. KONSTRUKSJON AV DATASETTE FOR ANALYSE .....  | 11 |
| 4.1 KOPLING AV REGISTERFILENE TIL AKU-FIL .....   | 11 |
| 4.2 AVGRENSING AV DATASETTE.....  | 11 |
| 4.3 IDENTIFISERING AV VIKTIGSTE FORHOLD I REFERANSEUKA .....  | 13 |
| 4.4 VIKTIGSTE FORHOLD DERSOM FEILDATERINGER GODTAS .....  | 14 |
| 4.5 DET ENDELIGE DATASETTE .....  | 15 |
| 4.6 GJENTAKELSE AV OPPLEGGET .....  | 15 |
| 5. FORSLAG TIL KLASIFIKASJONSROUTINE .....  | 16 |
| 5.1 “TRINNVIS METODE” .....   | 16 |
| 5.2 MULIGE REGISTERFORHOLD.....   | 16 |
| 5.3 TRINN 1: “PERSONER MED VIKTIGSTE ARBEIDSTAKERFORHOLD REGISTRERT BADE I AT OG LTO”.....                | 18 |
| 5.4 TRINN 2: “PERSONER MED VIKTIGSTE SELVSTENDIGFORHOLD REGISTRERT I EN AV SELVSTENDIGFILENE” .....       | 18 |
| 5.5 TRINN 3: “PERSONER MED VIKTIGSTE ARBEIDSTAKERFORHOLD REGISTRERT ENTEN I AT ELLER LTO”.....            | 19 |
| 5.6 TOTALSAMSVAR VED BENYTTELSE AV “TRINNVIS METODE” .....  | 20 |
| 6. KVALITETSVURDERING AV DEN FORESLÅTTE KLASIFIKASJONSROUTINEN.....                                       | 22 |
| 6.1 KVALITET INNENFOR DELPOPULASJONER .....   | 22 |
| 6.1.1 <i>Regionale delpopulasjoner</i> .....  | 22 |
| 6.1.2 <i>Delpopulasjoner dannet etter alder</i> .....   | 28 |
| 6.1.3 <i>Delpopulasjoner dannet etter arbeidsmarkedsvariable</i> .....                                    | 29 |
| 6.2 KLASIFIKASJONSROUTINENS KVALITET FOR ANDRE ÅR .....   | 30 |
| 6.3 KONKLUSJON .....  | 32 |

|   |           |
|---|-----------|
| <b>7. ALTERNATIVE KLASSEKATEGORISJONSROUTINER .....</b>                                       | <b>33</b> |
| 7.1 LÖNN SOM INDIKATOR PÅ SYSSELSETTING .....   | 33        |
| 7.2 TIDSAVSTAND .....   | 34        |
| 7.2.1 <i>Asymmetrisk vindu</i> .....  | 35        |
| 7.3 ENKEL UTVIDELSE AV DEN FORESLÅTTE KLASSEKATEGORISJONSROUTINEN FRA TRE TIL FEM TRINN ..... | 36        |
| 7.4 METODE BASERT PÅ MAKSIMALT SAMSVAR .....  | 38        |
| 7.4.1 <i>Inndeling i disjunkte grupper</i> .....  | 38        |
| 7.4.2 <i>Idealisert metode</i> .....  | 39        |
| 7.4.3 <i>Modifisert metode</i> .....  | 39        |
| 7.4.4 <i>Resultat</i> .....   | 39        |
| 7.4.5 <i>En vurdering av metoden</i> .....  | 40        |
| <b>8. ANALYSE AV PERSONER UTEN REGISTRERTE ARBEIDSFORHOLD .....</b>                           | <b>41</b> |
| 8.1 HOVEDTREKK .....  | 41        |
| 8.2 EN UTDYPNING AV RESULTATENE I ANALYSEN AV PERSONER UTEN REGISTRERTE ARBEIDSFORHOLD ....   | 43        |
| 8.3 EN KORT OPPSUMMERING .....  | 46        |
| <b>VEDLEGG A .....</b>  | <b>47</b> |
| <b>REFERANSER .....</b>   | <b>48</b> |
| <b>DE SIST UTGITTE PUBLIKASJONENE I SERIEN NOTATER .....</b>                                  | <b>49</b> |

# 1. Innledning

FoB 2000 vil for første gang basere den delen av folketellingen som omhandler personers arbeidsmarkedssituasjon kun på registeropplysninger. I den forbindelse har FoB2000 overført midler til arbeidet med å utvikle en klassifikasjonsrutine for personers sysselsettingsstatus, basert på informasjon fra administrative registre. Prosjektet er utført av Seksjon for arbeidsmarkedsstatistikk, (260), og Seksjon for statistiske metoder og standarder, (720). Dette notatet inneholder en presentasjon av klassifikasjonsrutinen som er utarbeidet, samt en beskrivelse av analysen som ligger til grunn for forslaget. I kapittel 2 gis det en nærmere gjennomgang av bakgrunnen for prosjektet.

Arbeidet som presenteres i dette notatet bygger videre på et forslag til klassifikasjonsrutine utviklet av Harald Dale-Olsen. Dette er presentert i et ikke utgitt notat "Klassifisering av status som sysselsatt eller ikke-sysselsatt", som igjen er delt opp i en presentasjon og en kravspesifikasjon (Dale-Olsen 1998). Vi har i kapittel 3 valgt å gi en kort gjennomgang av de momenter fra dette notatet som ligger til grunn for vårt arbeid.

Før vi presenterer vårt forslag til klassifikasjonsrutine, vil vi i kapittel 4 redegjøre for de bearbeidelser av datamaterialet som er gjort forut for vår analyse. Her gis en beskrivelse av utvalget som er benyttet, og hvordan datasettet som anvendes er konstruert.

ILO (International Labour Organisation) har utarbeidet en definisjon av begrepet sysselsatt som vi ønsker å følge. Denne definisjonen brukes i Arbeidskraftundersøkelsen (AKU). Vi har derfor i utarbeidelsen av selve klassifikasjonsrutinen som presenteres i kapittel 5, benyttet informasjon om personers arbeidsmarkedssituasjon gitt i AKU. Utgangspunktet har vært å legge samsvarsanalyser mellom AKU og registrene til grunn for valg av klassifikasjonsrutine. Det vil si at vi sammenligner den arbeidsmarkedssituasjonen en person har i AKU, med den status som er definert ved hjelp av informasjon fra registrene og den utarbeidete klassifikasjonsrutinen. Vi har i kapittel 6 også delt utvalget opp etter aktuelle bakgrunnsvariable, for å se hvorvidt graden av samsvar er forskjellig innenfor deler av populasjonen.

En rekke alternative klassifikasjonsrutiner har vært utprøvd i løpet av prosjektet. Enkelte av disse gjennomgås i kapittel 7.

Ved å kople sammen informasjon fra AKU og de forskjellige administrative registrene, kan vi analysere om det er visse grupper som er av særlig dårlig kvalitet i registrene. En slik analyse er gjengitt i kapittel 8, hvor vi ser nærmere på personer som oppgir i AKU at de er sysselsatt, men som vi ikke finner spor av i registrene i det aktuelle året.

## 2. Formålet med prosjektet

### 2.1 Motivasjon

Utgangspunktet for prosjektet er at vi ønsker å etablere en rutine for å klassifisere en person som sysselsatt eller ikke ved hjelp av informasjon fra forskjellige administrative registre. Årsaken er at registrene i motsetning til en utvalgsundersøkelse som Arbeidskraftsundersøkelsen (AKU), inneholder informasjon om hele befolkningen. På grunn av utvalgsusikkerheten knyttet til AKU, må vi derfor basere oss på nettopp registerdata når vi skal presentere sysselsettingstall for små deler av populasjonen.

Ideelt sett burde AKU og registerstatistikken gi samme totaltall for sysselsatte siden vi måler samme fenomen. Vi må imidlertid forholde oss til at de administrative registre ikke er innrettet spesielt på statistikk, og at det som følge av dette vil være forskjeller i anslaget på antall sysselsatte. Eksempler på mulige faktorer som virker inn på nøyaktigheten av de anslagene som gjøres ved hjelp av informasjon fra registrene, kan være:

- Verken dataorganisering eller registrerte kjennemerker i registrene er definert utfra de behov til informasjon statistikkutarbeiderne måtte ha.
- 
- Det forekommer manglende ajourføringer. Dette vil si at inn- og utmeldinger i registrene enten oppdateres for sent, eller uteblir. Når det gjelder effekten av forsinkelser vil denne være større desto kortere produksjonstid den aktuelle statistikken har. Undersøkelser har vist at i underkant av 70 prosent av meldingene er registrert etter et kvartal, og kun 40 prosent er registrert etter 4 uker.
- Enkelte begrepsdefinisjoner i det administrative system avviker fra statistisk relevante begreper. Et eksempel på dette momentet er definisjonen av sysselsatte som benyttes i Arbeidstakerregisteret (AT), kontra den som benyttes i AKU. For at en person skal innmeldes i AT må denne ha et arbeidsforhold med minst 4 timers gjennomsnittlig arbeidstid pr. uke, samt at dette arbeidsforholdet er av minst 6 dagers varighet. AKU baseres derimot på ILO-definisjonen som benytter betegnelsen arbeid for alt arbeid i referanseuka av minst en times varighet. Ser vi bort fra andre momenter som også spiller inn vil en naturlig følge av denne definisjonsforskjellen være at flere personer blir klassifisert som sysselsatte i AKU enn i AT.

Motivasjonen for selve prosjektet kan knyttes opp til det siste momentet ovenfor. Man ønsker å innføre en sysselsettingsdefinisjon i registerstatistikken som ligger tettere opp til den ønskede definisjon av sysselsetting, det vil si ILO definisjonen, enn den som benyttes i de forskjellige registrene. En naturlig følge av dette blir at anslagene på det totale antall sysselsatte som gjøres ved hjelp av registerinformasjon, ikke skal skille seg nevneverdig fra anslagene gjort i AKU.

Arbeidet med prosjektet kan grovt sett deles i tre. Den første delen, som utgjør kapittel 3 og 4 i notatet, tar for seg tilrettelegging av datamaterialet fra de forskjellige administrative registrene. Del 2, kapittel 5, 6 og 7, utgjør hoveddelen av dette notatet og er konsentrert rundt selve analysen som ligger til grunn for utforming og valg av klassifikasjonsrutine. Som vi skal komme nærmere inn på senere bygger utarbeidelsen av rutinen på samsvarsanalyser med AKU, noe som fører oss over i den tredje delen av prosjektet. Denne presenteres i kapittel 8 og gir en analyse av personer som oppgir i AKU at de er sysselsatte, men som det ikke finnes informasjon om i registrene.

## 2.2 Tilrettelegging av datamaterialet fra de administrative registrene

Den nåværende rutine for utarbeiding av offisielle sysselsettingstall ved hjelp av de administrative registrene, er kun basert på informasjon fra Arbeidstakerregisteret. Dette har gitt en underestimert anslaget på antall arbeidstakere. I 1992 var det eksempelvis 132 000 flere arbeidstakere ifølge AKU sammenlignet med Arbeidstakerregisteret. For å oppnå målet om samme totalnivå på antall sysselsatte ved hjelp av registrene som i AKU, må det derfor benyttes informasjon fra flere registre.

En del av arbeidet med utvelgelse av hvilke registre som benyttes og utarbeiding av årsfiler, er gjort før vi begynte vår analyse. Dette er presentert i Dale Olsen (1998). Vi vil imidlertid kort gjøre rede for dette arbeidet i kapittel 3, samt i kapittel 4 presentere de bearbeidinger av datamaterialet som er foretatt i forbindelse med påkopling av informasjon fra AKU.

## 2.3 Utarbeidelse av klassifikasjonsrutinen: Samsvarsanalyser mellom registrene og AKU på individnivå

Et av hovedmålene er som nevnt ovenfor, at anslaget på antall sysselsatte ved hjelp av informasjon fra registrene og klassifikasjonsrutinen skal ligge på samme nivå som ifølge AKU. Dette kan oppnås med ulik grad av samsvar på individnivå (mikrokonsistens). Med samsvar på individnivå menes at de som ved hjelp av klassifikasjonsrutinen blir definert som sysselsatte eller ikke-sysselsatt, har samme status ifølge AKU. I analysen som ligger til grunn for valg av klassifikasjonsrutine, er det lagt vekt på at det oppnås høyest mulig samsvar når det gjelder sysselsettingsstatus mellom registrene og AKU. Dette medfører at AKU vil få en mer aktiv rolle i registerstatistikken.

Det er en rekke innvendinger man kan rette mot det å la tallene fra AKU operere som en slags fasit for anslagene gjort ved hjelp av de administrative registrene. En av disse går på mulige feilkilder som kan påvirke tallene fra AKU, og at vi ved å benytte AKU så aktivt som vi gjør, overfører disse på registrene. Eksempler på mulige feilkilder kan være:

- Måle- og bearbeidingsfeil. Et generelt problem som gjelder statistiske undersøkelser, dette være seg både utvalgsundersøkelser og totaltelling, er at disse kan inneholde visse målings- og bearbeidingsfeil. Målefeil kan eksempelvis oppstå ved at oppgavegiveren på grunn av minnefeil, misforståelse av spørsmålet eller lignende, avgir feil svar eller at intervjueren krysser av i feil rubrikk på spørreskjemaet. Bearbeidingsfeil henspiller på feil koding av for eksempel næring, yrke eller feil som oppstår ved overføring av opplysningene fra spørreskjemaet til maskinlesbart medium.
- Utvalgsskjevheter. I motsetning til feilkilden ovenfor, skyldes dette at resultatene bygger på et utvalg av befolkningen. Utvalgsskjevhet kan blant annet oppstå når frafallet, de som ikke svarer, er systematisk ulikt resten av utvalget. Dette kan gi seg utslag i at personer som i arbeidssammenheng har en spesiell atferd, ikke blir representert i utvalget i samme grad som i befolkningen.
- Utvalgsusikkerhet utgjør ikke en feil i den forstand at noe går galt. Den gir uttrykk for den usikkerheten en får i resultatene fordi disse bygger på opplysninger om bare en del av befolkningen. Det innebærer at ulike utvalg kan gi forskjellige resultater.

Når vi allikevel har valgt å la informasjon fra AKU være retningsgivende i vår utarbeidelse av en klassifikasjonsrutine for registerstatistikk har dette sin bakgrunn i flere momenter.

For det første benytter AKU ILO - definisjonen av arbeid. Som før nevnt gir betegnelse arbeid her uttrykk for alt arbeide av minst en times varighet.

Et annet moment er at både valg av variable og definisjoner i Arbeidstakerregisteret ikke er spesielt beregnet for utarbeidelse av statistikk, samtidig som det også er problemer knyttet til oppdateringer av registeret. Det er derfor knyttet større tillit til AKU-tall når det gjelder totalnivået på antall arbeidstakere.

Til slutt har vi at Nasjonalregnskapet benytter totaltall fra AKU. Ved å utarbeide en klassifikasjonsrutine ved hjelp av AKU, vil vi få mer samsvar mellom de forskjellige sysselsettingstallene som publiseres.

## **2.4 Krav til den ferdige klasifikasjonsrutinen**

Selve klassifikasjonsrutinen består av en teknisk prosedyre, hvor det gjøres en vurdering av hvorvidt den registerinformasjonen vi har om en person kvalifiserer denne til å bli klassifisert som sysselsatt eller ikke. Som det kom frem ovenfor, har vi en del uttalte mål om de resultatene som den utarbeidete rutinen skal gi.

*1) Klassifikasjonsrutinen skal gi samme nivå på antall sysselsatte som AKU.*

Den registebaserte sysselsettingsdefinisjonen skal altså konstrueres slik at anslaget vi får på antall sysselsatte totalt skal være det samme som AKU.

*2) Antall personer definert som selvstendige ved klassifikasjonsrutinen skal også være i overensstemmelse med antallet gitt i AKU.*

En videreføring av prosjektet har vært å se på en mulig oppsplitting av de sysselsatte i arbeidstakere og selvstendige, og de ikke sysselsatte i arbeidsløse og utenfor arbeidsstyrken. Analysen i dette notatet tar hensyn til en oppsplitting av de sysselsatte. Bakgrunnen for at vi så ønsker å sette et mål for antall selvstendige, har å gjøre med kvaliteten på registerinformasjonen når det gjelder denne gruppen. Vi skal komme nærmere inn på konstruksjonen av selvstendigfilene og svakheten ved disse senere i notatet. Familiearbeidere vil forøvrig i vårt forslag, bli definert som arbeidstakere.

*3) Klassifikasjonsrutinen skal gi mest mulig samsvar på individnivå.*

Som nevnt ovenfor ønsker vi at størst mulig andel av de personene som vi har informasjon igjennom AKU, blir gitt samme sysselsettingsstatus ved registerinformasjon og den utarbeidete klassifikasjonsrutinen.

*4) Klassifikasjonsrutinen skal være "enkel" å implementere og oppdatere.*

Hvilke mulige administrative registre som kan benyttes til å angi sysselsettingstall, vil høyst sannsynlig endres over tid. Samtidig vil vi også oppleve at begrepene i registrene endres over tid. Det er derfor viktig å utarbeide en rutine hvor det kan justeres for slike endringer, uten å måtte gjøre fundamentale endringer i selve fremgangsmåten.

## **2.5 Analyse av personer som er sysselsatte i AKU, men er uten arbeidsforhold ifølge registrene**

Det å kople sammen AKU og registrene gir oss muligheter til å undersøke nærmere om det er visse grupper som er av særlig dårlig kvalitet i registrene. Vi vil derfor se nærmere på de personene som i AKU oppgir at de er sysselsatt men som vi ikke finner spor av i registrene i det aktuelle året. Dette gjør vi for å finne ut om gruppen utgjøres av personer med spesielle kjennetegn, eventuelt avdekke mulige svakheter ved AKU. Analysen er lagt til kapittel 8 i notatet.

## 3. Registergrunnlag

### 3.1 Administrative registre benyttet i analysen

Vi vil i dette delkapitlet gi en oversikt over de administrative registre vi har benyttet i analysen, med en kort beskrivelse av hvert enkelt. En mer utførlig presentasjon finnes i Dale-Olsen (1998).

#### 3.1.1 Arbeidstaker- / Arbeidsgiverregister (AT/AG)

Rikstrygdeverkets arbeidstakerregister (AT) definerer en arbeidstaker som enhver som arbeider i annens tjeneste for lønn eller annen godtgjørelse. For unntak se Dale-Olsen (1998). Bare jobber med minst 4 timers gjennomsnittlig arbeidstid pr. uke registreres. Registeret omfatter alle meldepliktige ansettelsesforhold med start- og eventuell stoppdato.

Rikstrygdeverkets arbeidsgiverregister (AG) har bedriftsoppdeling, det vil si bedrift som enhet. Registeret definerer en arbeidsgiver som enhver som benytter seg av arbeidstakere i Norge for utførelse av arbeid mot vederlag.

#### 3.1.2 Lønns- og trekkoppgave registeret (LTO)

Lønns- og trekkoppgaveregisteret (LTO) er skattemyndighetenes register over lønns- og trekkoppgaver. Det innrapporteres en oppgave per jobb, per kalenderår. Registeret omfatter utbetalte beløp for alle jobber med utbetaling på minst 400 kr. Trygdeutbetalinger og andre trekkpliktige ytelser er også med. Dateringene for de forskjellige lønnsforholdene er svært mangelfulle i LTO. De fleste jobbene er registrert med varighet i hele kalenderåret. Jobb i LTO er forøvrig knyttet til foretaksnivå.

#### 3.1.3 Likningsregisteret

Likningsregisteret inneholder opplysninger fra selvangivelsene for den enkelte person. De viktigste kjennemerkene for arbeidsmarkedsstatistikk er de som gjelder personinntekt som arbeidstaker og selvstendignæringsdrivende. Det finnes ikke opplysninger om når på året en person har opptjent sin inntekt.

#### 3.1.4 SOFA-søker-registeret

SOFA-søker-registeret er Arbeidsdirektoratets register over personer som er arbeidsledige eller på tiltak. Utfra månedsfiler blir det utarbeidet en årsfil, og konstruert start- og eventuelt stopptidspunkt for ledighets- og tiltaksperioder. Undersøkelser av SOFA-årsfilene når det gjelder kvaliteten på de konstruerte dateringene, viser for øvrig at de rutinene som er benyttet for å bestemme disse start- og stopptidspunktene konsekvent underestimerer varigheten av SOFA-forholdet. Til tross for dette benytter vi allikevel det tilgjengelige datagrunnlaget.

## 3.2 Tilrettelegging av data

En stor del av tilretteleggingen av datamaterialet er foretatt forut for vår analyse (Dale-Olsen 1998). I korte trekk går dette ut på å konstruere en årsfil ved hjelp av informasjon fra de administrative registrene nevnt ovenfor. Denne skal inneholde opplysninger om det til enhver tid viktigste arbeidsforholdet, og start og stoppdatoer for dette. Siden dette danner grunnlaget for konstruksjonen av datasettet som benyttes i vår analyse, har vi valgt å gi et kort sammendrag av prosedyren som er benyttet. Det tilhørende dataflytdiagrammet er presentert i vedlegg A.

### 3.2.1 Utarbeidelse av datafiler

I System for persondata er det foretatt en kopling mellom Arbeidstakerregisteret og LTO-registeret. Utfra denne koplingen er det konstruert tre nye filer med personer som henholdsvis har lønnsforhold uten kopling til arbeidstakerforhold (LTOUAT92), lønnsforhold med kopling til arbeidstaker- og arbeidsgiverforhold (ATMLTO92), og arbeidstakerforhold uten kopling til lønnsforhold (ATULTO92). For nærmere opplysninger om denne type kopling viser vi til Utne & Vassnes (1995).



Ut fra opplysninger om pensjonsgivende inntekt for selvstendige fra Likningsregisteret, er det konstruert to filer for selvstendig næringsaktivitet (SL\_JSF92 og SL\_NER92). Disse filene omfatter personer med pensjonsgivende inntekt i henholdsvis primærnæringer og annen næringsaktivitet utenom primærnæringerne. På dette stadiet kan et gitt fødselsnummer finnes i alle fem filene, samtidig som det i flere av filene også vil kunne forekomme dubletter. Dette fører til det første trinnet i prosedyren for klassifikasjon av viktigste arbeidsforhold, nemlig valg av viktigste arbeidstakerforhold. Filene for selvstendige blir foreløpig holdt utenfor.

### **3.2.2 Valg av viktigste arbeidstakerforhold**

Klassifikasjon av viktigste arbeidstakerforhold er delt i tre utfra hva slags registerforhold de forskjellige personene har:

1. Kun komplette lønns- og arbeidstakerforhold, det vil si at et og samme arbeidsforhold er registrert både i arbeidstakerregisteret og lønns- og trekkoppgaveregisteret.
2. Klassifikasjon av personer med både ukomplete og komplette lønns- og arbeidstakerforhold.
3. Ingen komplette lønns- og arbeidstakerforhold.

Hvis en person kun har ett koplet lønns- og arbeidstakerforhold i løpet av hele året, er dette personens viktigste arbeidstakerforhold. For de øvrige personene benyttes opplysninger om forventet arbeidstid, kontantlønn, samt start- og stopptidspunkt for de forskjellige arbeidsforhold til å klassifisere viktigste arbeidstakerforhold. Vi vil etter klassifikasjonen ha opplysninger om hva som på et hvert tidspunkt i løpet av året er viktigste arbeidstakerforhold, registrert med start og stoppdato. For en nærmere beskrivelse av hvilke kriterier som er benyttet, henviser vi til kapittel 5 i Dale-Olsen (1998).

### **3.2.3 Konsistensbehandling mot SOFA-søker-registeret**

Etter valg av viktigste arbeidstakerforhold kopler man sammen alle forhold i SOFA-søkerfilen som gjelder ledighets- tiltaks- og permitteringsforhold med alle arbeidstaker-/lønnsforhold, det vil si både komplette og ukomplete arbeidstaker- / lønnsforhold. Den nye filen vil gi en fullstendig oversikt over alle arbeids- / lønns- og ledighetsforhold en person har hatt i løpet av 1992. Videre konsistensbehandles først dateringen for viktigste arbeidstakerforhold og dateringene av ledighetsforholdene. Resultatet av denne korreksjonen vil så bestemme de korrigerede datoene for eventuelle biforhold. De forskjellige korrigeringskriteriene er beskrevet i Dale-Olsen (1998).

Opplysningene fra SOFA-søker-registeret blir også benyttet til å påføre ledighetsdatoer på de to selvstendigfilene. Dette gjøres ved å ta utgangspunkt i at et forhold som selvstendig forutsettes å vare hele året. Årsaken til denne forutsetningen er at Likningsregisteret, som vi benytter når vi konstruerer de to selvstendige filene, ikke inneholder dateringer av arbeidsforholdene. Hvis vedkommende har et ledighetsforhold i SOFA-søker-registeret i løpet av året, blir perioden selvstendige forholdet varer, endret til å gjelde hele året bortsett fra de periodene som personen er helt ledig.

### **3.2.4 Valg av viktigste arbeidsforhold**

Ved å sammenligne samlet daglønn som arbeidstaker og samlet daglønn i et eventuelt selvstendig forhold, bestemmes til slutt viktigste arbeidsforhold utfra hvor denne er høyest. Samlet daglønn som arbeidstaker beregnes først ved å dele lønnssummen knyttet til arbeidsforholdet på antall dager dette har vart, gitt ved start og stoppdatoer. Videre summeres daglønn i alle tellende arbeidstaker- og lønnsforhold sammen. Samlet daglønn som selvstendig er gitt som forholdet mellom summen av samlet pensjonsgivende inntekt i primærnæringer og samlet pensjonsgivende inntekt fra annen næringsvirksomhet, og antall dager i året.

Etter prosedyren for valg av viktigste arbeidsforhold er foretatt står vi igjen med fem filer, en for komplette arbeidstaker-/lønnsforhold, to for henholdsvis ukomplete arbeidstaker- / lønnsforhold, og to for selvstendigforhold i henholdsvis primærnæringer og andre næringer. Samlet inneholder disse filene

opplysninger om viktigste arbeidstakerforhold, viktigste selvstendigforhold og viktigste arbeidsforhold, samt konsistensbehandlede dateringer for hver enkelt av disse. Dette danner grunnlaget for vår videre behandling av datamaterialet.

## 4. Konstruksjon av datasett for analyse

Med de fem registerfilene beskrevet i forrige kapittel som utgangspunkt, har vi utarbeidet et mer hensiktsmessig datasett for våre analyser. For det første er det koplet til AKU-informasjon, og dernest har vi for hver person valgt ut et viktigste forhold for referanseuka (AKU-uka) samtidig som vi samler informasjon fra de andre jobbene personen har

### 4.1 Kopling av registerfilene til AKU-fil

I AKU i 1992 ble ca. 24 000 personer i alderen 16-74 forsøkt intervjuet hvert kvartal, fordelt på en bestemt uke hver måned. På grunn av rotasjonsordningen i 1992 er antallet intervjuobjekter som ikke har vært intervjuet tidligere, hhv. ca 24 000, 12 000, 12 000 og 6000 for de fire kvartalene, noe som gir ca 54 000 personer i bruttoutvalget (de man forsøker å intervjuer) for 1992. Vi vil kun se på første gangsintervjuene i 1992 for de 54 000, og vi vil kople på arbeidsmarkedsinformasjon for disse personene.

I likhet med AKU, har også likningsfilen person som enhet, dvs at hver rad eller record i datasettet består av ulike personer. Vi legger all likningsinformasjon om de 54 000 i AKUs bruttoutvalg inn på filen.

Vi har fem arbeidsmarkedsfiler – beskrevet i forrige kapittel – bestående av hhv. koplete forhold (A/A + LTO), ukoplete A/A-forhold, ukoplete LTO-forhold, selvstendige fra jordbruk/skogbruk/fiske, og til sist selvstendige fra andre næringer. De fem filene har jobbforhold som enhet, dvs at hvert jobbforhold er representert med en rad i datafilen. Alle disse jobbene legges i en samlefil (fremdeles er jobbforhold enhet). Nå vil en person være representert med så mange rader/records som det antall jobbforhold vi finner om personen fra de fem arbeidsmarkedsfilene. På samlefilen legges all AKU-Likningsinformasjon inn på hvert jobbforhold. For en bestemt person med f.eks syv jobbforhold, vil altså hver rad (jobbforhold) inneholde AKU-Likningsinformasjon som er identisk for alle syv jobbforhold.

For å få dataene på en form som er hensiktsmessig for vår analyse, må vi ha person som enhet i datafilen, dvs at vi må samle all informasjon om en person på en rad i datafilen. I 4.3 og 4.4 vil vi se på hvordan vi skal løse denne utfordringen, men først vil vi avgrense det antall personer vi skal analysere.

### 4.2 Avgrensning av datasettet

Vi ønsker å benytte AKU som en slags fasit når vi skal undersøke ulike klassifikasjonsrutiner for sysselsettingsstatus, og vi er da avhengig av å ha AKU-informasjon om sysselsettingsstatus. Vi kan gruppere alle personene i AKUs bruttoutvalg (de som er forsøkt intervjuet) i fire grupper etter de to variablene som angir hvorvidt en person har registerinformasjon, her i betydningen en av de fem arbeidsmarkedsfilene ovenfor, og hvorvidt personen har svart i AKU. Dette er illustrert i tabell 4.1.

Tabell 4.1 Antall personer etter hvorvidt personen har svart i AKU og hvorvidt personen er registrert med registeropplysninger fra en av de fem arbeidsmarkedsfilene.

| Antall<br>Prosent                  | Ikke svart i<br>AKU | Har svart i<br>AKU | Total            |
|------------------------------------|---------------------|--------------------|------------------|
| Ikke register-<br>informasjon..... | 581<br>1,08         | 2 549<br>4,73      | 3 130<br>5,81    |
| Register-<br>informasjon.....      | 3 956<br>7,34       | 46 804<br>86,85    | 50 760<br>94,19  |
| Total.....                         | 4 537<br>8,42       | 49 353<br>91,58    | 53 890<br>100,00 |

I analysen er det de personer hvor vi har informasjon både fra AKU og arbeidsmarkedsfiler vi betrakter, og vi kaller disse for gruppe B ('B' for "begge kilder"). Nedenfor vil vi se på hvilke problemer det kan medføre at vi ser bort fra de andre personene i bruttoutvalget. Vi lar som tidligere grad av samsvar (eller samsvarsgrad) betegne andelen av personene som klassifiseres likt som i AKU. Videre lar vi relativt nivåavvik betegne den relative forskjellen mellom antall sysselsatte iflg. klassifikasjonsrutinen og antallet iflg AKU. For enkelhetsskyld vil vi nedenfor implisitt betrakte AKU-status og sann status som det samme, selv om dette ikke er tilfellet.

Når klassifikasjonsregelen til slutt skal benyttes i praksis, er det på hele den delen av populasjonen som er 16-74 år. Ettersom bruttoutvalget i AKU antas å være representativt i forhold til populasjonen, ønsker vi at alle personer i AKUs bruttoutvalg skal være riktig klassifisert gjennom klassifikasjonsrutinen, dvs grad av samsvar på 100 prosent. Vi ser i tabell 4.1 at 13,15 prosent av alle personene i bruttoutvalget ikke tilhører vår analysegruppe (B). Disse 13,15 prosent utgjør et frafall i analysen ved at vi ikke undersøker om disse er riktig klassifisert. Dersom kvaliteten på klassifikasjonsrutine i forhold til sannhet er vesentlig dårligere i de tre frafallsgruppene enn i vår analysegruppe (B), vil våre konklusjoner i kapittel 5 kunne bli feil.

AKU-fracfallet, dvs de som mangler all AKU-informasjon, er de personene i tabell 4.1 som er i første kolonne. Disse to gruppene har vi ikke mulighet til å analysere. Vi har registerbasert sysselsettingsstatus (status iflg. klassifikasjonsrutinen), men ettersom vi mangler AKU-basert status, kan vi ikke sjekke kvaliteten. Dersom sannheten er at kvaliteten i AKU-fracfallsgruppene er systematisk lavere enn i gruppe B, betyr det at vi overvurderer kvaliteten av klassifikasjonsrutinen. Dersom relativt nivåavvik mellom registerbasert status og sann status er systematisk større i AKU-fracfallsgruppene enn i gruppe B, og i samme retning, overvurderer vi evnen til å treffe riktig nivå. Dersom det derimot er motsatt retning kan vi ha undervurdert evnen til å treffe riktig nivå. Nivå- og samsvarsbetraktningene gjelder både for hele populasjonen og for mindre deler av populasjonen som f.eks Nordland fylke.

AKU-fracfallet kunne vært benyttet til å undersøke kvaliteten av AKU, men det faller utenfor emnet i dette notatet.

Frafallet som skyldes at kun arbeidsmarkedsregisterinformasjon mangler (4,73 prosent av bruttoutvalget), kan vi undersøke ettersom vi kan betrakte manglende registeropplysninger som at man klassifiseres som ikke-sysselsatt. Vi kaller denne gruppen for gruppe A ('A' for 'AKU'). Det viser seg å være 307 AKU-sysselsatte, og dermed ser vi fra tabell 4.1 at andelen feilklassifiserte og relativt nivåavvik er 12 prosent (307/2549).<sup>1</sup> Dersom dette tallet avviker mye fra tilsvarende andeler i gruppe B, og i tillegg gruppe A ikke er for liten i forhold til gruppe B, betyr dette at vi feilvurderer

<sup>1</sup> Ettersom alle er klassifisert som ikke-sysselsatte, blir andel vi klassifiserer feil (samsvarsfeil) lik andel som er AKU-sysselsatte. Med andre ord er både samsvarsfeil og relativt nivåavvik fra AKU-nivået lik AKU-basert sysselsettingsnivå.

kvaliteten med vår analyse basert på gruppe B i kapittel 5 og 6. I kapittel 8 ser vi nærmere på gruppe A, men utfra et bredere perspektiv der vi ønsker å få et oversiktsbilde over denne gruppen.

### 4.3 Identifisering av viktigste forhold i referanseuka

For å lette analysen, er det hensiktsmessig å få person som enhet istedet for jobbforhold, dvs få en rad/record per person. Årsaken til dette er at vi i opptellinger av sysselsatte har person som enhet, og datasettet bør derfor ha variable der person er enhet. Dersom vi skal ha person som enhet, taper vi en del informasjon. I motsatt fall måtte vi legge informasjonen fra hvert jobbforhold til en person etter hverandre, og dermed vil mange variable måtte omdøpes (f.eks lønn1, lønn2 osv for jobbforhold 1 og 2). I tillegg må av datatekniske årsaker alle personer ha like mange variable som personen med flest jobbforhold, så datasettet ville blitt veldig stort og uryddig. Vi må altså velge ut et jobbforhold som vi betrakter som det viktigste i referanseuka. Vi vil kalle datasettet med en rad per person for analysedatasettet

Vi ønsker å ikke miste all informasjon om de andre jobbforholdene, og derfor aggregeres informasjon om alle jobbforholdene: vi beregner sum av kontantlønnene (som vi vil kalle samlet lønn) og antall forhold av de ulike typene.<sup>2</sup> Det vi mister er all annen informasjon spesielt knyttet til andre forhold enn det viktigste forholdet, som f.eks næring, arbeidstid m.m.

Det er ulike måter å prioritere mellom de jobbforholdene som en person har. Vi har valgt å prioritere datering og type forhold, og vi må velge hvilket aspekt som skal telle mest av disse to. Riktig datering vil si at et forhold er aktivt i referanseuka, som er uka for AKU-intervjuet. Dersom vi benytter dateringsbegrepene “tidsrommet hvor forholdet er viktigste arbeidstaker-/lønnsforhold” og “tidsrommet hvor forholdet er viktigste selvstendigforhold”, så får vi en todelt prosedyre som danner datasettet vi benytter i kapittel 5:

Del 1: valg av viktigste arbeidstaker-/lønnsforhold og valg av viktigste selvstendigforhold. Disse to forholdene finner vi ved å først legge vekt på riktig datering, og så for arbeidstaker-/lønnsforhold prioritere koplet forhold<sup>3</sup> fremfor ukoplet forhold.

Del 2: prioritering mellom viktigste selvstendigforhold og viktigste arbeidstaker-/lønnstakerforhold. Vi vektlegger først datering, og dersom den er riktig for begge to, er prioriteringen slik: 1) koplet forhold, 2) selvstendigforhold og 3) annet forhold.

For personen som er illustrert i figur 4.1, ser vi at det i del 1, er det koplete forholdet som er det viktigste arbeidstaker-/lønnsforholdet, og det ene selvstendigforholdet blir viktigste selvstendigforhold. Begge har riktig datering, så i del 2 prioriteres da det koplete forholdet.

For forslaget til klassifikasjonsrutine som vi har i kapittel 5, er det kun personer med et riktig datert forhold som er potensielt registersysselsatte<sup>4</sup>. Imidlertid kan personer der alle forholdene er feil datert også være potensielt registersysselsatte, og vi vil nedenfor se nærmere på hva dette innebærer når vi skal lage analysedatasettet, og vil vil trekke inn disse personene som potensielt registersysselsatte i 7.2 og 7.3.

Istedet for å se på viktigste selvstendigforhold og viktigste arbeidstaker-/lønnsforhold, kunne vi ha betraktet Dale-Olsen (1998) sin definering av datoer for viktigste forhold (se kap. 3). Isåfall ville del 1 ovenfor falt bort. Denne metoden er prøvd ut, men gir ikke bedre resultat.

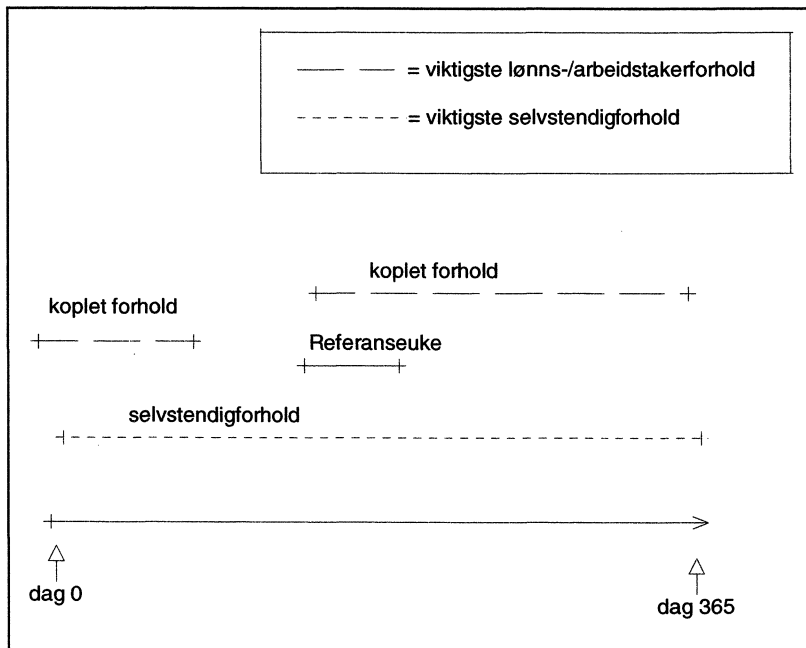
---

<sup>2</sup> I datasettet som benyttes i klassifikasjonsrutinen i kap. 5, beregnes ikke antall forhold.

<sup>3</sup> forhold som er identifisert både i A/A og LTO.

<sup>4</sup> klassifisert som sysselsatte iht. klassifikasjonsrutine

Figur 4.1. Et eksempel på hvordan registerforhold som er viktigste i en eller annen forstand i løpet av året, er fordelt sammen med Referanseuka for en bestemt person.



#### 4.4 Viktigste forhold dersom feildateringer godtas

Som nevnt ovenfor, kan det finnes personer der ingen forhold er riktig datert. Utfra perspektivet i 4.3, vil disse personene bli klassifisert som ikke-registersysselsatte uansett. Vi vet imidlertid at det er unøyaktigheter og feil mht. registreringer av start- og stoppdatoer, og for personer som ikke har noe forhold som er riktig datert, men derimot bommer med et par uker, kan det godt tenkes at de er sysselsatte, men at dateringen bare er unøyaktig. Vi presenterer nedenfor en modifikasjon av del 1 og del 2 i 4.3, som gjør at vi kan betrakte personer der alle forhold bommer på dateringen. Denne modifikasjonen vil vi ikke få bruk for i den foreslåtte klassifikasjonsrutinen i kapittel 5, men vi vil bruke den i 7.2.

Istedet for å velge forholdet med riktig datering i del 1 og del 2, generaliserer vi til at vi velger forholdet med mest riktig datering. Dette betyr f.eks at dersom ingen arbeidstaker-/lønnsforhold er riktig datert, så velger vi det forholdet der dateringen bommer minst på referansetidspunktet. Dersom vi betrakter personen på figur 4.1, kan vi tenke oss at vi har to gardiner som henger vertikalt og dekker hele tidsrommet unntatt referansetidspunktet (AKU-uka), og at gardinene trekkes til side horisontalt, like mye på hver side av referanseuka, inntil et forhold som er aktivt som viktigste arbeidstaker-/lønnsforhold dukker opp. Avstanden mellom referanseuka og nærmeste gardin er da avviket fra referansetidspunktet, og vi vil kalle dette avviket for tidsavstand.

I 7.2 vil vi se nærmere på hvilken tidsavstand som gir best resultat. Ellers vil vi ikke benytte oss av dette begrepet.<sup>5</sup>

<sup>5</sup> Datasettet som benyttes i vår analysen har imidlertid med tidsavstand, slik at klassifikasjonsrutineforslaget vi presenterer i kap. 5, enkelt kan inneholde mildere krav til tidsavstand. Ved beregning av tidsavstand er det ikke tatt hensyn til om det er skuddår.

#### 4.5 Det endelige datasettet

Det datasettet som kommer ut av algoritmene ovenfor har en rad/record per person. På denne raden er all AKU- og likningsinformasjon tilstede og all opprinnelig informasjon om det viktigste forholdet på referansetidspunktet.<sup>6</sup> I tillegg har vi samlet lønn for personen.

#### 4.6 Gjentakelse av opplegget

Konstruksjonen av datasettet basert på registerfilene det bearbejdet nivået  $k_3$  (se 3.2.1) består av fem SAS-programmer som må kjøres etter hverandre. En beskrivelse av hvordan disse programmene skal kjøres ligger på området `q:\dok\tslo\sf\prosjekt\aaarsyss\`.

Når man går igang med å finne registerbasert sysselsettingsstatus for senere år, kan det være aktuelt å basere selvstendigfilene på Selvangivelsesregisteret, som ikke fantes i 1992. Dette betyr imidlertid trolig ingenting for denne delen av databehandlingen, men derimot har det visse konsekvenser for den bearbejdingen av registrene som er beskrevet i kapittel 3.

---

<sup>6</sup> I tillegg har vi antall dager dateringen bommer på referansetidspunktet

## 5. Forslag til klassifikasjonsrutine

### 5.1 “Trinnvis metode”

Vi skal i dette kapittelet presentere vårt forslag til klassifikasjonsrutine (registerbasert sysselsettingsdefinisjon). Selve fremgangsmåten har vi valgt å kalle “trinnvis metode”. Vår angrepsvinkel tar utgangspunkt i at hele utvalget ansees som ikke-registersysselsatt. For hvert trinn innfører vi så et registerkriterium, som for eksempel at ett og samme arbeidsforhold må være registrert både i Arbeidstakerregisteret (AT) og Lønns- og trekkoppgaverregisteret (LTO) på et gitt tidspunkt. De personene som oppfyller dette kriteriet blir omklassifisert, og får status som registersysselsatt. Disse trekkes så ut av utvalget før vi går videre til neste trinn hvor de gjenværende blir vurdert på nytt. Dette betyr at hvis en person først er definert som sysselsatt, kan ikke denne statusen mistes på et senere tidspunkt i klassifikasjonsrutinen.

For hvert trinn gjør vi en samsvarsundersøkelse med AKU. Det vil si at vi undersøker i hvilken grad personene vi har definert som registersysselsatt, har samme status ifølge AKU.

Etter endt prosedyre skal vi ved hjelp av informasjon fra registrene ha klassifisert samme antall sysselsatte som i AKU. Dette som følge av målsettingen om at klassifikasjonsregelen skal gi samme sysselsettingsnivå som AKU. Utvalget benyttet i analysen består som før nevnt av 46 804 personer, hvorav 31 940 personer oppgir i AKU at de er sysselsatt. Det vil med andre ord si at vi ved hjelp av den utarbeidete metoden skal definere 31 940 personer av utvalget som registersysselsatte.

Det er under fremstillingen av forslaget tatt hensyn til en oppsplitting av de sysselsatte i henholdsvis arbeidstakere og selvstendige, som også er et av delmålene for rutinen.

### 5.2 Mulige registerforhold

For bedre å klargjøre hva som ligger bak oppbyggingen av klassifikasjonsrutinen og de kriterier som ligger til grunn for definering av en persons sysselsettingsstatus, vil vi først gi en oversikt over mulige registerforhold en person kan ha. Utfra de registrene som vi har benyttet og de koplinger som er gjort (se kapittel 3 for oversikt), kan en persons viktigste arbeidstakerforhold eller selvstendigforhold være registrert på en av følgende måter:

- 1) Et og samme arbeidsforhold er både registrert i AT og LTO på referansetidspunktet (ATMLTO).
- 2) Arbeidsforholdet er kun registrert i AT på referansetidspunktet (ATULTO).
- 3) Arbeidsforholdet er kun registrert i LTO på referansetidspunktet (LTOUAT).
- 4) Arbeidsforholdet er registrert i selvstendigfilen for jordbruk, skogbruk og fiske på referansetidspunktet (SLJSF).
- 5) Arbeidsforholdet er registrert i selvstendigfilen for andre næringer på referansetidspunktet (SLNER).

Referansetidspunktet referer til den uken hver enkelt person ble spurt i AKU.

I tabell 5.1 og 5.2 fordeler vi personene i utvalget utfra hva slags registerforhold de har på referansetidspunktet, og hva de oppgir som sysselsettingsstatus i AKU. Dette gir en oversikt over hvor stor grad av samsvar det er mellom den informasjonen registrene gir og hva personene oppgir i AKU.



Tabell 5.1 Oversiktstabell over samsvar i sysselsettingsstatus mellom arbeidsmarkedsfilene og AKU, for personer med registrert arbeidstaker- og/eller lønnsforhold.

| Mulige registerforhold som arbeidstaker:      | Sysselsettingsstatus ifølge AKU |               |              |                 |
|---|---------------------------------|---------------|--------------|-----------------|
|   | I alt                           | Arbeidstakere | Selvstendige | Ikke-sysselsatt |
| <b>Koplet forhold (ATMLTO):</b>               |                                 |               |              |                 |
| Antall personer uten ATMLTO                   | 19 684                          | 3 523         | 2 348        | 13 813          |
| Antall personer med ATMLTO                    | 27 120                          | 25 788        | 281          | 1 051           |
| <b>AT forhold uten LTO kopling (ATULTO) :</b> |                                 |               |              |                 |
| Antall personer uten ATULTO                   | 46 559                          | 29 205        | 2 605        | 14 749          |
| Antall personer med ATULTO                    | 245                             | 106           | 24           | 115             |
| <b>LTO forhold uten AT kopling (LTOMAG):</b>  |                                 |               |              |                 |
| Antall personer uten LTOMAG                   | 40 697                          | 27 222        | 1 892        | 11 583          |
| Antall personer med LTOMAG                    | 6 107                           | 2 089         | 737          | 3 281           |

Vi tar utgangspunkt i personene med et koplet forhold, dvs at viktigste arbeidstakerforhold er registrert både i AT og LTO på referansetidspunktet. Det fremgår av tabell 5.1 at denne gruppen utgjøres av 27 120 personer. Av disse har 25 788 personer samme status i AKU, 281 oppgir at de er selvstendige, og 1 051 er ikke-sysselsatt ifølge AKU.

Dette betyr at hvis vi som første kriterium i klassifikasjonsregelen har at personer med et koplet forhold på referansetidspunktet blir definert som registersysselsatte og arbeidstakere, vil vi ved hjelp av registerinformasjon gi 95,1 prosent (25 788/27120) samme status som i AKU. Velger vi imidlertid et av de andre registerforholdene som kriterium for å bli klassifisert som arbeidstaker får vi følgende resultater:

Ser vi på de 245 personene med arbeidstakerforholdet kun registrert i Arbeidstakerregisteret (ATULTO) har 43,3 prosent av disse oppgitt i AKU at de er arbeidstakere.

Tilsvarende tall for de 6 107 personene som kun er registrert i lønns- og trekkoppgaverregisteret (LTOMAG) på referansetidspunktet, er 34,2 prosent (2 089 / 6 107).

Tabellen ovenfor bekrefter som forventet at et arbeidstakerforhold som både er registrert i AT og LTO er mer pålitelige enn hvis det kun er å finne i et av registrene. Samtidig får vi et bilde av hvor god informasjon ukoplete registerforhold gir om en persons sysselsettingsstatus. Tabell 5.2 gir en liknende oversikt over viktigste selvstendigforhold.

Tabell 5.2 Oversiktstabell over samsvar i sysselsettingsstatus mellom arbeidsmarkedsfilene og AKU, for personer med registrert selvstendigforhold.

| <u>Sysselsettingsstatus ifølge AKU</u>          |        |               |              |                 |
|---|--------|---------------|--------------|-----------------|
| Mulige registerforhold som selvstendig:         | I alt  | Arbeidstakere | Selvstendige | Ikke-sysselsatt |
| <b>Selvstendigforhold primærnæring (SLJSF):</b> |        |               |              |                 |
| Antall personer uten SLJSF                      | 45 337 | 28 791        | 1 875        | 14 671          |
| Antall personer med SLJSF                       | 1 467  | 520           | 754          | 193             |
| <b>AT forhold uten LTO kopling (SLNER) :</b>    |        |               |              |                 |
| Antall personer uten SLNER                      | 44 278 | 28 495        | 1 145        | 14 638          |
| Antall personer med SLNER                       | 2 526  | 816           | 1 484        | 226             |

Tabell 5.2 viser at 51,4 prosent (754/1 467) av personene med viktigste selvstendig forhold registrert i selvstendigfilen for jordbruk, skogbruk og fiske blir klassifisert riktig i henhold til AKU hvis vi benytter dette som klassifikasjonskriterium på trinn 1. Tilsvarende tall for personer med arbeidsforhold i selvstendigfilen for andre næringer er 58,7 prosent (1 484/2 526).

Som konklusjon kan vi si at det å velge registerforholdet ATMLTO som kriterium i trinn 1 vil gi størst andel riktig klassifiserte personer ifølge AKU, sett i forhold til de konkurrerende alternativene. Dette ligger til grunn for definisjonen av trinn 1.

### **5.3 Trinn 1: “Personer med viktigste arbeidstakerforhold registrert både i AT og LTO”**

*Det første kriteriet for at en person skal klassifiseres som sysselsatt er at viktigste arbeidstakerforhold er registrert både i Arbeidstakerregisteret og Lønns- og trekkoppgaverregisteret på referansetidspunktet.*

Ifølge tabell 5.1 har vi altså at 27 120 personer blir definert som registersysselsatte på dette trinnet. Disse personene vil også få status som arbeidstakere. Av disse igjen er som før nevnt 95,1 prosent også klassifisert som arbeidstakere i AKU. Selv om vi anser dette nivået på samsvar som meget bra, har vi sett på mulige løsninger for å begrense antall feilklassifiserte. Denne analysen har vi lagt til kapittel 7, hvor vi gir eksempler på andre mulige innfallsvinkler for klassifikasjonsrutinen som er fremkommet under arbeidet.

Før vi går videre med trinn 2, trekker vi ut de som nå er klassifisert som sysselsatte/arbeidstakere. Disse kan ikke endre status på et senere tidspunkt. De resterende er med til trinn 2.

### **5.4 Trinn 2: “Personer med viktigste selvstendigforhold registrert i en av selvstendigfilene”**

*Kriterium nummer to for å bli klassifisert som registersysselsatt er at personen har et arbeidsforhold registrert i en av de to selvstendigfilene på referansetidspunktet.*

*Tillegskriterium: Nivået på antall selvstendige skal være det samme som i AKU.*

I tillegg til at personens arbeidsforhold må ha en spesiell registertilhørighet, er det i dette trinnet definert et tilleggskriterium. Årsaken til dette ligger i hvordan vi har kommet frem til filene for

selvstendige i primærnæringene og selvstendige i andre næringer. Som beskrevet i kapittel 3, er disse konstruert ved hjelp av opplysninger om pensjonsgivende inntekt fra likningsregisteret. Herfra har vi ingen informasjon om start- og stoppdatoer for de forskjellige selvstendigforholdene. Når det gjelder dateringen er derfor varigheten for alle arbeidsforhold satt fra 1.1 til 31.12. Selv om vi deretter konsistensbehandler opplysningen med SOFA - søker - registeret, det vil si at varigheten av selvstendigforholdet blir justert hvis personen i en periode er registrert som arbeidsledig eller på tiltak, vil resultatet bli en overestimering av varigheten av selvstendigforholdet. Dette vil igjen føre til at det anslaget på antall selvstendige vi får ved kun å sette som kriterium at de skal være registrert i en av selvstendig-filene på et gitt tidspunkt, vil være større enn det anslaget gjort i AKU.

Problemet blir derfor å finne ut hvilke personer med arbeidsforhold i selvstendigregistrene på referansetidspunktet som ikke skal være med. Vi har valgt å utelate de med høyest lønn fra lønns- og arbeidstakerforhold. Resonnementet som ligger bak er at enkelte som driver selvstendig virksomhet, har et arbeidstakerforhold ved siden av. Det kan ofte være vanskelig å definere hvilke av disse som anses som personens hovedvirksomhet. Personer med høy lønn fra andre arbeidstakerforhold, viser seg i større grad å være arbeidstakere ifølge AKU. Disse vil ha stor mulighet til å bli klassifisert som arbeidstakere i siste trinn, der størrelsen på kontantlønnen, i tillegg til registerforholdet, er kriterium for å bli definert som sysselsatt og arbeidstaker.

Som følge av trinn 2 er antall personer klassifisert som registersysselsatte kommet opp i 29 749 personer. Av disse er 2 629 definert som selvstendig og 27 120 som arbeidstakere. 72 prosent av de selvstendige har samme yrkesstatus i AKU, mens 95,1 prosent av arbeidstakerne har som før nevnt, også status som arbeidstaker i AKU.

Et av målene var at totaltallet på antall sysselsatte skulle opp på samme nivå som i AKU. For å oppnå dette, mangler vi 2 191 arbeidstakere.

Før vi definerer et kriterium for å bli klassifisert som registersysselsatt i trinn 3, trekker vi ut de som til nå har fått status som selvstendige og legger disse sammen med personene som ble omklassifisert i trinn 1. De personene som til nå ikke er definert som registersysselsatte, blir med til trinn 3.

### **5.5 Trinn 3: “Personer med viktigste arbeidstakerforhold registrert enten i AT eller LTO”**

*Kriterium nummer 3 for å bli klassifisert som registersysselsatt er at viktigste arbeidstakerforhold er registrert enten i AT eller i LTO på referansetidspunktet.*

*Tilleggs-kriterium: Totalnivået på antall sysselsatte skal være det samme som totalnivået i AKU.*

Ved hjelp av trinn 3 ønsker vi å omklassifisere så stor andel av personene som til nå ikke er registersysselsatte, at vi kommer opp på samme nivå på antall sysselsatte som i AKU. Mulige registerforhold for de resterende på referansetidspunktet er enten at de kun har arbeidsforholdet registrert i AT, eller kun i LTO. Av de gjenværende har 5 599 personer et av disse registerforholdene. Dette vil med andre ord si at hvis vi klassifiserer alle disse som arbeidstakere vil vi få et høyere totalnivå på antall sysselsatte enn AKU. Som under trinn 2, må vi også her utarbeide en form for tilleggskriterium. Dette tilleggskriteriet må utformes på en slik måte at det sikrer en utvelgelse som gjør at de 2 191 personene som faktisk er arbeidstakere, i størst mulig grad blir klassifisert som arbeidstakere.

Vi har valgt følgende fremgangsmåte:

Vi sorterer personer som har sitt viktigste arbeidstakerforhold registrert enten kun i AT, eller kun i LTO, etter avtagende samlet lønn fra alle arbeidstakerforhold (sum kontantlønn). Vi starter med å omklassifisere de personene som har høyest sum kontantlønn, og “fyller opp” til man kommer opp på

samme nivå for antall sysselsatte som i AKU. Personene som da blir definert som registersysselsatte / arbeidstakere i trinn 3 er de med høyest samlet lønn fra alle arbeidstakerforhold.

Av de 2 191 personene som i trinn 3 blir klassifisert som arbeidstakere, er 788 ikke sysselsatt ifølge AKU, 146 er selvstendige, og 1 257 har i AKU status som arbeidstakere. Dette betyr at 57 prosent blir riktig klassifisert ifølge AKU på trinn 3.

14 864 personer har etter trinn 3 ikke fått status som verken arbeidstaker eller selvstendig. Disse blir nå klassifisert som ikke-sysselsatte. 85,2 prosent av disse er også ikke-sysselsatt ifølge AKU.

## 5.6 Totalsamsvar ved benyttelse av “Trinnvis metode”

Vi er nå ferdig med klassifikasjonsrutinen, og kan analysere samsvaret mellom den status de forskjellige personene har fått som følge av informasjonen fra registrene, og den status de har i AKU. Tabell 5.3 viser samsvar på såkalt 2x2 nivå. Det vil si at vi kun ser på status som sysselsatt / ikke-sysselsatt. Det skraverte området i hver tabell viser antallet personer som har samme status både i registrene og AKU.

Tabell 5.3 Samsvar i sysselsettingsstatus, ikke-sysselsatt, mellom registrene og AKU etter endt klassifikasjonsrutine.

| Sysselsettingsstatus ifølge registrene | Sysselsettingsstatus ifølge AKU |               |        |
|--|---------------------------------|---------------|--------|
|  | Ikke sysselsatt                 | Sysselsatt    | Totalt |
| Ikke sysselsatt                        | <b>12 668</b>                   | 2 196         | 14 864 |
| Sysselsatt                             | 2 196                           | <b>29 744</b> | 31 940 |
| Totalt                                 | 14 864                          | 31 940        | 46 804 |

Tar vi kun for oss de som ved hjelp av klassifikasjonsrutinen er klassifisert som sysselsatte er 29 744 personer av totalt 31 940 riktig klassifisert ifølge AKU, det vil si 93,12 prosent. Blant de som har fått status ikke-sysselsatt er det også personer som har en annen status ifølge AKU. Tar vi dette med i beregningen, finner vi at samsvaret mellom registrene og AKU etter endt klassifikasjonsrutine totalt er 90,62 prosent  $((12\,668 + 29\,744)/46\,804)$ .

Tabell 5.4 Samsvar i sysselsettingsstatus, ikke-sysselsatt, arbeidstaker, selvstendig, mellom registrene og AKU etter endt klassifikasjonsrutine.

| Sysselsettingsstatus ifølge registrene: | Sysselsettingsstatus ifølge AKU |              |               | Totalt |
|---|---------------------------------|--------------|---------------|--------|
|   | Ikke sysselsatt                 | Selvstendig  | Arbeidstaker  |        |
| Ikke sysselsatt                         | <b>12 668</b>                   | 310          | 1 886         | 14 864 |
| Selvstendig                             | 357                             | <b>1 892</b> | 380           | 2 629  |
| Arbeidstaker                            | 1 839                           | 427          | <b>27 045</b> | 29 311 |
| Totalt                                  | 14 864                          | 2 629        | 29 311        | 46 804 |

Tabell 5.4 tar hensyn til en oppsplitting av sysselsatte i henholdsvis selvstendige og arbeidstakere. Av arbeidstakerne har 92,27 prosent  $(27\,045 / 29\,311)$  samme status i AKU. Videre er 72 prosent  $(1\,892 / 2\,629)$  av personene vi har klassifisert som selvstendige, også selvstendige ifølge AKU. Tar vi med de ikke-sysselsatte ser vi at samsvaret blir 88,89 prosent  $((12\,668 + 1\,892 + 27\,045)/46\,804)$ .

Registerbaserte sysselsettingstall gir i motsetning til en utvalgsundersøkelse som AKU, muligheten til å splitte opp i mindre delpopulasjoner. Derfor vil vi i neste kapittel å se på samsvar innenfor utvalgte demografiske grupper. Dette vil gi et mer inngående bilde av kvaliteten på den klassifikasjonsrutinen vi har valgt.

## 6. Kvalitetsvurdering av den foreslåtte klassifikasjonsrutinen

Vi har i kapittel 5 presentert vårt forslag til klassifikasjonsrutine for sysselsettingsstatus. Rutinen gir samme totalnivå på antall sysselsatte som i AKU, og ca. 90 prosent av personene i utvalget vi har analysert får samme status basert på klassifikasjonsrutinen som de får i AKU (dvs at graden av samsvar er 90 prosent). Selv om dette utvalget representerer hele populasjonen, er dette likevel på langt nær nok for å kunne fastslå at rutinen er god. Kvaliteten er avhengig av bl.a to faktorer: hvor god den er for mindre grupper av befolkningen, f.eks for en kommune, og av hvor god metoden vil være når den skal benyttes på et annet år enn 1992.

Med todelt sysselsettingsstatus mener vi verdiene sysselsatt/ikke-sysselsatt, mens tredelt innebærer verdiene arbeidstaker/selvstendig/ikke-sysselsatt. Vi vil benytte samsvar om andelen personer som er likt klassifisert iflg. klassifiseringsrutinen og i AKU. Begrepet relativt nivåavvik betegner relativt avvik mellom antall sysselsatte (evt lønnstakere eller selvstendige dersom det er det vi ser på) basert på klassifikasjonsregelen og basert på AKU.

### 6.1 Kvalitet innenfor delpopulasjoner

Klassifikasjonsrutinen skal benyttes til å lage et folketellingsregister i FoB2000, og man skal da utfra dette registeret presentere sysselsettingstall for kommuner, og for grupper innenfor kommuner. Vi ønsker at alle slike grupper, heretter kalt delpopulasjoner, skal ha riktig antall registersysselsatte.

Utfra størrelsen på det utvalget vi har, kan vi ikke undersøke direkte hvorvidt vi får riktig nivå for små delpopulasjoner som f.eks kvinner i Etnedal Kommune. Utvalget fra denne delpopulasjonen er for lite, nærmere bestemt ca ni personer. Måten å sikre best mulig kvalitet for små delpopulasjoner er at flest mulig personer skal være riktig klassifisert på grovere nivå, i dette tilfellet at flest mulig av innbyggerne i Etnedal eller Oppland skal være riktig klassifisert, og likeledes flest mulig av kvinnene i hele landet. Jo flere som er riktig klassifisert på grovere nivå som kommunenivå, jo sikrere kan vi være på at vi treffer riktig sysselsettingsnivå på mer detaljert nivå som kvinner i Etnedal. For utvalget har vi at dersom alle i Etnedal kommune skulle få riktig sysselsettingsstatus, så får automatisk alle *kvinner* i Etnedal også riktig registerbasert sysselsettingsstatus.

Riktig sysselsettingsnivå i en kommune holder *ikke* for å få riktig nivå blant f.eks kvinner i kommunen: klassifikasjonen kunne f.eks tenkes å favorisere menn som sysselsatte, og dersom vi lar for mange menn bli klassifisert som registersysselsatte og et tilsvarende antall kvinner bli ikke-sysselsatte, blir registersysselsettingsnivået blant kvinnene i kommunen feil, men nivået i i kommunen er fremdeles riktig.

Nedenfor vil vi danne ulike delutvalg ved å gruppere utvalget etter ulike variable. Vi vil gruppere etter demografiske og arbeidslivsrelaterede variable.

#### 6.1.1 Regionale delpopulasjoner

For den regionale dimensjon ser vi i tabell 6.1 at samsvaret for todelt sysselsettingsstatus er over 89 prosent innenfor alle fylker unntatt Sogn og Fjordane som har 86 prosent. For tredelt sysselsettingsstatus synker samsvaret naturligvis noe<sup>7</sup>, og for Sogn og Fjordane er det nå nede i underkant av 82 prosent. Relativt avvik fra AKU-nivået på todelt status er mindre enn 3 prosent med ett unntak (Aust-Agder, -6 prosent), mens det for tredelt sysselsettingsstatus er noe større avvik: for lønnstakere er det nå Sogn og Fjordane som er dårligst (-7,7 prosent), mens det for selvstendige er store avvik på over 20 prosent for både Telemark, Sogn- og Fjordane, Møre og Romsdal og Troms.

---

<sup>7</sup> samsvar på tredelt nivå krever mer enn på todelt nivå. En som både er registersysselsatt og sysselsatt iflg. AKU kan være arbeidstaker iflg. klassifikasjonsrutinen og selvstendig iflg. AKU.

Disse tallene er så store at det er grunn til å sette spørsmålstegn ved om man i det hele tatt kan sette lit til selvstendigtallene for lavt regionalt nivå.

*Tabell 6.1 Overensstemmelse mellom sysselsettingsstatus definert ved den foreslåtte klassifikasjonsrutinen og status i AKU, fordelt etter fylke . Todelt sysselsettingsstatus.*

| Fylke                | Totalt antall personer | Antall sysselsatte iflg. AKU | Samsvar:   |  |
|----------------------|------------------------|------------------------------|--|--|
|                      |                        |                              | Andel klassifisert identisk med AKU-status.Prosent | Relativt avvik mellom nivå-tall. Prosent |
| Uoppgitt.....        | 16                     | 10                           | 50,00  | -20,00                                   |
| Østfold.....         | 2646                   | 1730                         | 91,72  | -0,52                                    |
| Akershus.....        | 4810                   | 3519                         | 91,08  | 0,20                                     |
| Oslo.....            | 4795                   | 3308                         | 91,43  | 0,94                                     |
| Hedmark.....         | 1714                   | 1082                         | 90,55  | 3,14                                     |
| Oppland.....         | 2497                   | 1665                         | 90,51  | -1,74                                    |
| Buskerud.....        | 2517                   | 1748                         | 91,46  | -0,17                                    |
| Vestfold.....        | 2200                   | 1485                         | 90,86  | -0,34                                    |
| Telemark.....        | 1838                   | 1193                         | 90,10  | 2,35                                     |
| Aust-Agder.....      | 416                    | 272                          | 91,59  | -6,25                                    |
| Vest-Agder.....      | 1634                   | 1079                         | 88,92  | -2,13                                    |
| Rogaland.....        | 4149                   | 3068                         | 91,83  | -1,73                                    |
| Hordaland.....       | 4600                   | 3188                         | 90,52  | -1,32                                    |
| Sogn og Fjordane.... | 990                    | 721                          | 86,06  | 3,05                                     |
| Møre og Romsdal....  | 2663                   | 1766                         | 89,15  | 1,42                                     |
| Sør-Trøndelag.....   | 3054                   | 2060                         | 91,68  | -0,39                                    |
| Nord-Trøndelag.....  | 1149                   | 772                          | 89,38  | 0,78                                     |
| Nordland.....        | 2670                   | 1668                         | 90,34  | 1,44                                     |
| Troms.....           | 1613                   | 1056                         | 88,66  | -0,10                                    |
| Finnmark.....        | 833                    | 550                          | 88,84  | 2,73                                     |

Tabell 6.2 Overensstemmelse mellom sysselsettingsstatus definert ved den foreslåtte klassifikasjonsrutinen og status i AKU, fordelt etter fylke . Tredelt sysselsettingsstatus.

| Fylke               | Totalt antall personer | Antall lønns-<br>takere iflg.<br>AKU | Antall selv-<br>stendige iflg.<br>AKU | Samsvar:   |   |
|---------------------|------------------------|--------------------------------------|---------------------------------------|--|---|
|                     |                        |                                      |                                       | andel klassifisert<br>identisk med AKU-<br>status. Prosent | Relativt avvik mellom<br>nivå tall. Prosent |
|                     |                        |                                      |                                       | Lønnstakere  | Selvstendige                                |
| Uoppgitt            | 16                     | 10                                   |                                       | 50,00  | -30,00                                      |
| Østfold             | 2646                   | 1578                                 | 152                                   | 90,06  | -0,57                                       |
| Akershus            | 4810                   | 3289                                 | 230                                   | 89,81  | 0,30  |
| Oslo                | 4795                   | 3146                                 | 162                                   | 90,72  | 1,56  |
| Hedmark             | 1714                   | 961                                  | 121                                   | 88,39  | 4,06  |
| Oppland             | 2497                   | 1500                                 | 165                                   | 88,71  | -0,80                                       |
| Buskerud            | 2517                   | 1583                                 | 165                                   | 89,99  | 0,44  |
| Vestfold            | 2200                   | 1365                                 | 120                                   | 89,50  | 0,22  |
| Telemark            | 1838                   | 1106                                 | 87                                    | 88,90  | 0,27  |
| Aust-Agder          | 416                    | 253                                  | 19                                    | 90,14  | -6,32                                       |
| Vest-Agder          | 1634                   | 954                                  | 125                                   | 86,60  | -1,68                                       |
| Rogaland            | 4149                   | 2867                                 | 201                                   | 90,17  | -1,60                                       |
| Hordaland           | 4600                   | 2966                                 | 222                                   | 88,85  | -1,25                                       |
| Sogn og<br>Fjordane | 990                    | 600                                  | 121                                   | 81,92  | 7,67  |
| Møre og<br>Romsdal  | 2663                   | 1632                                 | 134                                   | 86,48  | -0,12                                       |
| Sør-<br>Trøndelag   | 3054                   | 1849                                 | 211                                   | 89,88  | -0,70                                       |
| Nord-<br>Trøndelag  | 1149                   | 691                                  | 81                                    | 87,12  | 1,01  |
| Nordland            | 2670                   | 1470                                 | 198                                   | 87,87  | 0,61  |
| Troms               | 1613                   | 975                                  | 81                                    | 86,61  | -2,87                                       |
| Finnmark            | 833                    | 516                                  | 34                                    | 87,04  | 1,74  |



Dersom vi ser på todelt sysselsettingsstatus, tyder funnene ovenfor på at kvaliteten til klassifikasjonsrutinen delvis er upåvirket av hvilken region vi betrakter. Denne antakelsen kan settes nærmere på prøve ved å studere forholdene på kommunenivå. Pga det begrensede intervjugrunnlaget innenfor de fleste kommuner, kan vi med visse unntak ikke trekke konklusjoner om kvalitet for enkeltkommuner, men dersom vi har en klar tendens som gjelder alle eller mange kommuner, gir dette grunnlag for å si noe om generelle trekk hos kommunene som gruppe.

Figur 6.1 nedenfor viser at for kommuner der AKU-utvalgsstørrelsen er større enn 10 (178 kommuner) er det for todelt og tredelt sysselsettingsstatus samsvar på over hhv. 82 prosent og 78 prosent for kommuner med utvalgsstørrelse større enn 50 – utvalgsstørrelse på 50 betyr at det er ca. 3750 innbyggere i kommunen ettersom populasjonen er ca. 75 ganger større enn utvalget. Samsvaret varierer litt mindre for store kommuner (utvalgsstørrelse f.eks over 400).

I figur 6.2 ser vi at relativt avvik av registerbasert sysselsettingsnivå i kommunene i forhold til AKU-nivå (relativt nivåavvik) er mindre enn 10 prosent for alle kommuner der antall sysselsatte i utvalget er større enn 30 (flere enn 2100 sysselsatte i kommunen). For mindre kommuner er ikke relativt avvik større enn 30 prosent. Dersom vi splitter sysselsatte opp i lønnstakere og selvstendige, følger lønnstakere omtrent samme mønster som sysselsatte, med den forskjell at det er 10-20 kommuner i varierende størrelse opp til ca. 15000 lønnstakere i kommunen som har relativt avvik opp til 30 prosent.

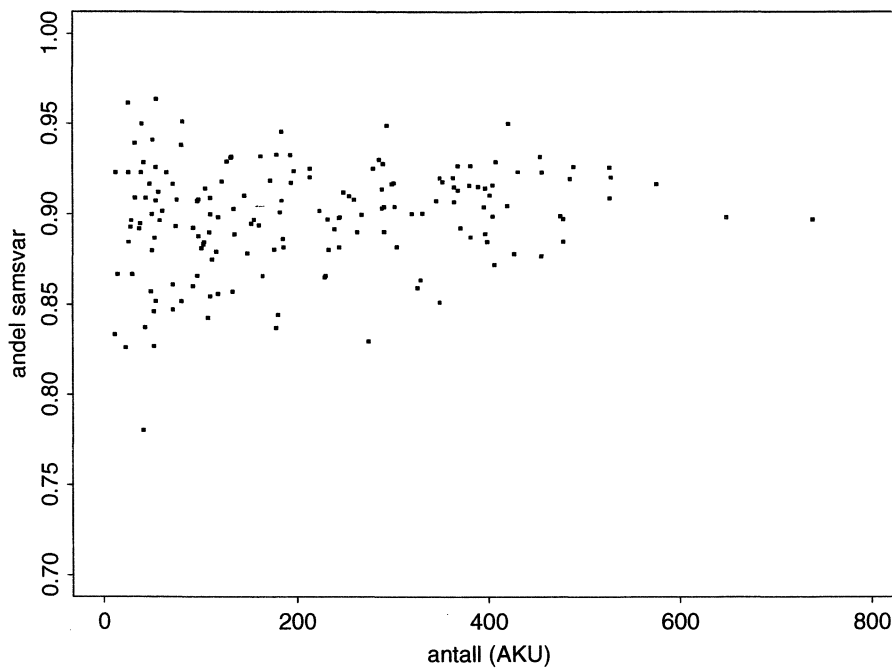
For selvstendige er relativt avvik i nivå tall atskillig større enn for lønnstakere, selv om denne tendensen reduseres når vi tar høyde for at antall selvstendige er lavt i de fleste kommuner slik at usikkerheten er større av den grunn. Figur 6.2 c) viser at for det mindretallet av kommunene hvor utvalget består av flere enn 20 AKU-selvstendige, kan vi ikke si mer enn at relativt avvik er mindre enn 30 prosent, som er et høyere tall enn det vi bør ha. For mindre kommuner må vi opp i 50 prosent relativt avvik før vi dekker de fleste kommunene.<sup>8</sup>

Undersøkelsen på kommunenivå tyder på at klassifikasjonsrutinen for todelt statusinndeling til en viss grad er robust overfor region når det gjelder kvalitet. For tredelt sysselsettingsstatus er det grunn til å sette større spørsmålsteget ved dette fordi at bl.a relativ feil av selvstendignivå på kommunenivå grovt anslått er omtrent tre ganger større enn for lønnstakere, og selv om vi tar høyde for at selvstendignivået er lavere, er relativ feil ca. 50 prosent større (30 prosent istedet for 20 prosent).

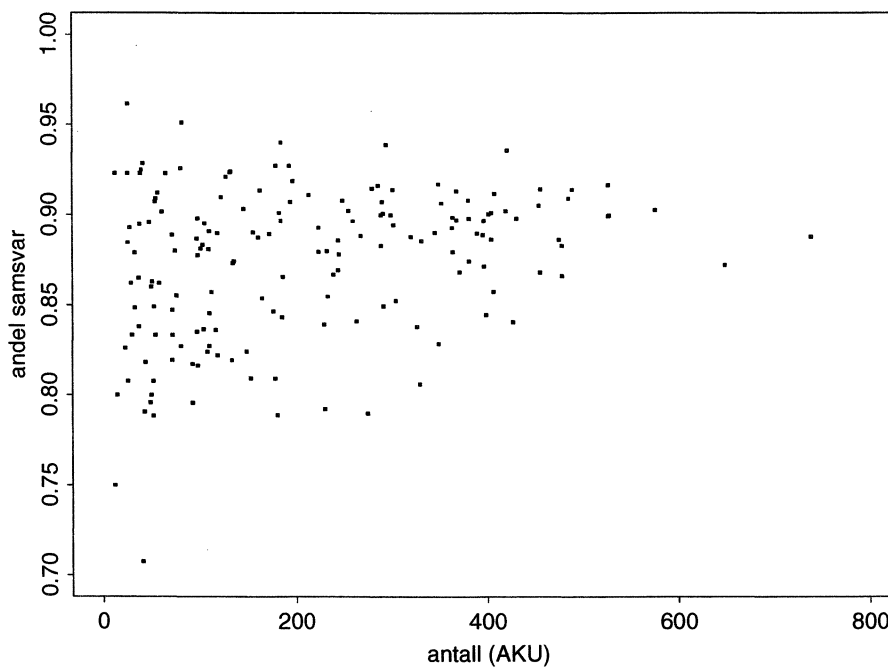
---

<sup>8</sup> de kommunene som har større avvik er så få (ca 10 stk) at vi ikke kan si om dette skyldes klassifikasjonsrutinen eller om det skyldes utvalgsskjevhet i AKU.

Figur 6.1 Spredningsplott av kommuner når a) todelt og b) tredelt sysselsettingsstatus. Vertikal akse er andel individer som har samsvarende sysselsettingsstatus iht. AKU og foreslått klassifikasjonsrutine i ulike kommuner, og horisontal akse er utvalgsstørrelsen til kommunen. Kommuner med utvalgsstørrelse på over 10 og under 800.

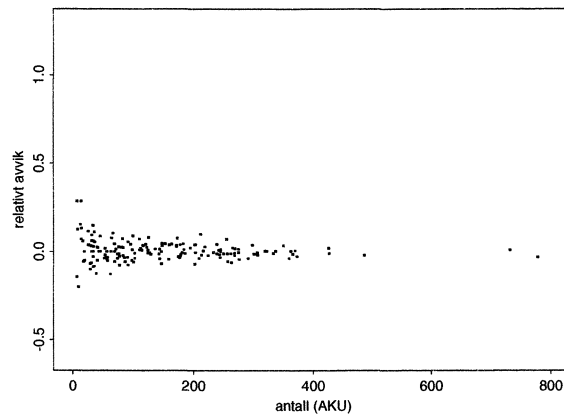


(a)

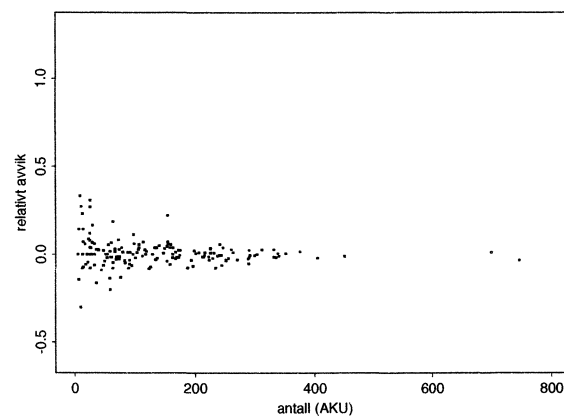


(b)

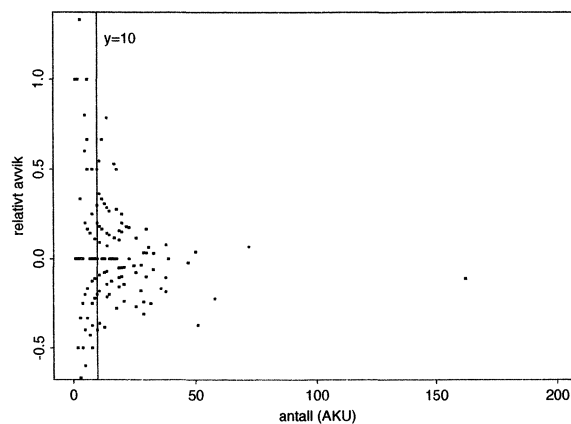
Figur 6.2 Spredningsplott av kommuner. Vertikal akse er relativt avvik av a) sysselsatte, b) lønntakere og c) selvstendige iflg. foreslått klassifikasjonsrutine i forhold til AKU-tall, og horisontal akse er antall AKU-sysselsatte i kommunen. Kommuner med utvalgsstørrelse på over 10. De to-tre største kommunene etter verdi på x-aksen er ikke tatt med.



(a)



(b)



(c)

### 6.1.2 Delpopulasjoner dannet etter alder

Tabell 6.3 og 6.4 viser samsvar og relativt nivåavvik for ulike aldersgrupper. For samsvar er tendensen den samme både for todelt og tredelt sysselsettingsstatus. I aldersgruppen 16-19 år er samsvaret kun 75 prosent, og i gruppen 20-29 er tallet ca. 83 prosent. Nivået for sysselsettingsstatus er som ventet ikke så bra for de unge (-7 prosent), men mer overraskende er det at personer over 67 år har 21 prosent flere sysselsatte iht. klassifikasjonsrutine enn iht. AKU. For tredelt sysselsettingsstatus avviker nivået for selvstendige ganske mye fra AKU-nivå: 16-19 åringene er her for få til at vi kan konkludere, men for 20-29 åringene og de over 67 år er avviket hhv. -12 prosent og +37 prosent<sup>9</sup>.

Tabell 6.3 Overensstemmelse mellom sysselsettingsstatus definert ved den foreslåtte klassifikasjonsrutinen og status i AKU, fordelt etter alder. Todelt sysselsettingsstatus.

| Alder         | Totalt antall personer | Antall sysselsatte iflg. AKU | Samsvar:   |  |
|---------------|------------------------|------------------------------|--|--|
|               |                        |                              | andel klassifisert identisk med AKU-status . Prosent | Relativt avvik mellom nivå-tall. Prosent |
| 16-19 år..... | 2911                   | 1109                         | 75,82  | -7,39                                    |
| 20-29 år..... | 9889                   | 7041                         | 83,46  | 0,43                                     |
| 30-49 år..... | 18905                  | 16422                        | 94,07  | -1,01                                    |
| 50-66 år..... | 10207                  | 6936                         | 94,23  | 1,77                                     |
| 67 år +.....  | 4892                   | 432                          | 93,01  | 21,76                                    |

Tabell 6.4 Overensstemmelse mellom sysselsettingsstatus definert ved den foreslåtte klassifikasjonsrutinen og status i AKU, fordelt etter alder. Tredelt sysselsettingsstatus.

| Alder     | Totalt antall personer | Antall lønns-takere iflg. AKU | Antall selv-stendige iflg. AKU | Samsvar:   |  |              |
|-----------|------------------------|-------------------------------|--------------------------------|--|--|--------------|
|           |                        |                               |                                | andel klassifisert identisk med AKU-status . Prosent | Relativt avvik mellom nivå-tall. Prosent |              |
|           |                        |                               |                                |  | Lønnstakere                              | Selvstendige |
| 16-19 år  | 2911                   | 1102                          | 7                              | 75,34  | -8,44                                    | 157,14       |
| 20-29 år  | 9889                   | 6754                          | 287                            | 82,31  | 0,96                                     | -12,20       |
| 30-49 år  | 18905                  | 14966                         | 1456                           | 91,74  | -0,56                                    | -5,63        |
| 50-66 år  | 10207                  | 6173                          | 763                            | 92,12  | 0,97                                     | 8,26         |
| 66 år --> | 4892                   | 316                           | 116                            | 92,52  | 16,14                                    | 37,07        |

#### Nærmere undersøkelse av de yngste

Vi så ovenfor at det i en så stor gruppe som alle 16-19 åringene, både er relativt lavt samsvar og at nivået ikke treffer helt.<sup>10</sup> Det kunne tenkes at man kunne laget en modifisert klassifikasjonsrutine for denne aldersgruppen slik at vi fikk bedre resultat. Vi lar 'unge' betegne de som er 16-19 år og 'voksne' betegne de som er over 19 år. Vi starter med å se etter hva som skiller de unge fra de voksne, og en naturlig hypotese er da at samsvaret blant de voksne blir mye større fordi at en høyere andel av de voksne er heltidsansatte med stabile komplette forhold over flere år.

<sup>9</sup> for mange registerbasert selvstendige over 67 år kan muligens skyldes (men dette er kun gjetninger) at nylig pensjonerte gårdbrukere henger igjen i registrene, kanskje ved at barn driver gården i den pensjonertes navn et par år, eller at gården forpaktet bort slik at gården gir inntekt uten at eier selv arbeider.

<sup>10</sup> På tredelt sysselsettingsstatus klassifiseres 8,44 prosent færre som lønnstakere gjennom klassifikasjonsrutinen enn antallet er iflg. AKU.

Tabell 6.5 Overensstemmelse mellom sysselsettingsstatus definert ved den foreslåtte klassifikasjonsrutinen og status i AKU, fordelt etter alder og registerforhold. Todelt sysselsettingsstatus.

| Alder              | Registerforhold                    | Andel personer. Prosent | Samsvar: andel klassifisert identisk med AKU-status Prosent |
|--------------------|------------------------------------|-------------------------|---|
| Under 20 år.....   | Koplet forh. i referanseuka.....   | 26,3                    | 75,7  |
|                    | Ikke kopl.forh. i referanseuka...  | 73,7                    | 75,2  |
|                    |                                    | 100                     |   |
| F.o.m. 20 år ..... | Koplet forh. i referanseuka.....   | 60,0                    | 95,7  |
|                    | Ikke kopl.forh. i referanseuka.... | 40,0                    | 81,0  |
|                    |                                    | 100                     |   |

Tabell 6.5 viser at det som må antas å være den sikreste indikasjon på sysselsetting, nemlig koplet forhold i referanseuka, ikke er et pålitelig kjennetegn for de unge: blant de som har koplet forhold i referanseuka blant de unge, er det kun 75,7 prosent som virkelig er sysselsatte, mot tilsvarende 95,7 prosent blant de voksne. Vi har at 6,3 prosent<sup>11</sup> av de unge blir feilklassifisert pga feilaktig koplet forhold i referanseuka, dvs at 6,3 prosentpoeng av samsvarsreduksjonen fra maksverdien 100 prosent forklares. Resten av samsvarsreduksjonen skyldes at de fleste unge (73,7 prosent) tilhører de uten koplet forhold i referanseuka, mens det motsatte er tilfelle for de voksne (40 prosent). Det er ingen grunn til å tro at vi kan modifisere klassifikasjonsrutinen for de unge, både fordi at den antatt sikreste informasjonen man kan ha, nemlig å ha koplet forhold i referanseuka, ikke er pålitelig informasjon for de unge, og fordi at blant de som ikke har koplet forhold i referanseuka, er samsvaret blant de unge ikke så mye lavere enn det er for de voksne (75,2 prosent mot 81 prosent).

Av de unge med koplet forhold i referanseuka, er det altså hele 100 prosent - 75,7 prosent = 24,3 prosent som ikke er AKU-sysselsatte, mens den tilsvarende andel for de voksne er 100 prosent - 95,7 prosent = 4,3 prosent. Dette kan umiddelbart gi oss det inntrykket at dateringene våre av koplete forhold er mye mer feilaktige for unge personer, men det store avviket mellom unge og de andre vil antakelig bli mindre dersom man ser bort fra alle de voksne med stabile heltidsjobber som går over flere år og der det for de fleste år ikke er noen dateringer som skal stemme.

Vi har ovenfor sett at manglende samsvar blant de unge (16-19 år) skyldes at man i denne gruppen ikke har mange personer med korrekte koplete forhold i referanseuka – man har få personer med koplet forhold i referanseuka, og mange av dem er feil (iflg. AKU).

### 6.1.3 Delpopulasjoner dannet etter arbeidsmarkedsvariable

Vi beveger oss nå over på arbeidsmarkedsvariablene avtalt arbeidstid og næring. Begge disse variablene er hentet fra en bearbejdet fil basert på arbeidstaker/arbeidsgiverregisteret (A/A), dvs at det ikke er grunnlag for å se på selvstendige; både registerbasert selvstendige og personer med LTO-forhold som sitt viktigste havner i kategorien “uoppgift”. Vi ser ikke overraskende at vi både har veldig bra samsvar og nivå (hhv. over 93 prosent og mindre enn 1 prosent avvik) for personer med arbeidstid over 20 timer per uke (tabell 6.6). For de andre er likevel samsvaret og nivåavviket ikke verre enn hhv. 84 prosent og 10 prosent.

<sup>11</sup>  $(100-75,7)\% \times 26,3\% = 6,3\%$

Alle personer med oppgitt næring fra A/A har meget bra samsvar og lite avvik fra AKU-nivå (tabell 6.7) unntatt jordbruk, skogbruk og fiske der samsvaret både er så lavt som 76 prosent og vi har fått med 23 prosent for mange. Bakgrunnsmaterialet viser at det utelukkende er nivåavviket som forårsaker mindre samsvar, i den forstand at alle AKU-sysselsatte også blir klassifisert som sysselsatte, og i tillegg blir enda en del personer klassifisert som sysselsatte.

Tabell 6.6 Overensstemmelse mellom sysselsettingsstatus definert ved den foreslåtte klassifikasjonsrutinen og status i AKU, fordelt etter avtalt arbeidstid. Todelt sysselsettingsstatus

| Avtalt arbeidstid. Timer/uke | Totalt antall personer | Antall sysselsatte iflg. AKU | Samsvar:   |  |
|------------------------------|------------------------|------------------------------|--|--|
|                              |                        |                              | andel klassifisert identisk med AKU-status . Prosent | Relativt avvik mellom nivå-tall. Prosent |
|                              | 17167                  | 5019                         | 85,34  | -4,76                                    |
| 4-19.....                    | 4815                   | 3781                         | 84,01  | 11,11                                    |
| 20-29.....                   | 3299                   | 2988                         | 93,82  | 0,94                                     |
| 30 + .....                   | 21523                  | 20152                        | 95,81  | -1,04                                    |

Tabell 6.7 Overensstemmelse mellom sysselsettingsstatus definert ved den foreslåtte klassifikasjonsrutinen og status i AKU, fordelt etter næring (fra Arbeidstaker-/arbeidsgiverregisteret). Todelt sysselsettingsstatus.

| Næring fra Arbeidstaker-/arbeidsgiverregisteret (ISIC standard 1983) | Totalt antall personer | Antall sysselsatte iflg. AKU | Samsvar:   |  |
|--|------------------------|------------------------------|--|--|
|  |                        |                              | andel klassifisert identisk med AKU-status . Prosent | Relativt avvik mellom nivå-tall. Prosent |
| Uoppgitt.....  | 17346                  | 5167                         | 85,33  | -4,57                                    |
| Jordbruk, skogbruk, fiske og fangst.....                             | 479                    | 336                          | 76,83  | 22,92                                    |
| Ojeutvinning og bergverksdrift.....                                  | 429                    | 405                          | 96,04  | -0,74                                    |
| industri.....  | 5102                   | 4690                         | 95,28  | 0,06                                     |
| Kraft- og vannforsyning.....   | 362                    | 334                          | 97,79  | -1,20                                    |
| Bygge- og anleggsvirksomhet.....                                     | 1703                   | 1526                         | 93,31  | -1,44                                    |
| varehandel, hotell- og restaurantvirksomhet                          | 5350                   | 4721                         | 92,02  | 2,52                                     |
| Transport, lagring, post og telekommunik...                          | 2261                   | 2103                         | 93,23  | -1,38                                    |
| Finanstjenester, .eiendom.,forn.tjen.yting.....                      | 2352                   | 2190                         | 94,18  | 1,05                                     |
| Offentlig, sosial og privat tjenesteyting.....                       | 11420                  | 10468                        | 94,40  | 0,69                                     |

## 6.2 Klassifikasjonsrutinens kvalitet for andre år

Klassifikasjonsrutinen vi har foreslått er forsøkt laget slik at den virker fornuftig i lys av tidligere erfaring rundt registrene. Dette gir en viss trygghet for at metoden vil være god også for andre år enn 1992. Likevel bærer også metoden preg av at den er spesialkonstruert for 1992-dataene: i og med at vi har fasiten for utvalget for 1992, kan vi lage en klassifikasjonsrutine som tilpasser dataene så bra at vi tilpasser tilfeldige sammenhenger for dette året. Når vi skal benytte rutinen nesten umodifisert for senere år, vil de tilfeldige sammenhengene kunne helle i en annen retning slik at rutinen vår blir dårligere for senere år. En måte å sjekke at klassifikasjonsrutinen er god for andre år (dvs robust overfor endring i tid), er å undersøke rutinen for f.eks 1993. Ettersom det medfører en del arbeid å

frembringe 1993-dataene, velger vi istedet en annen tilnærming: dersom rutinen virker bra innenfor hver måned i 1992, så indikerer dette en lav følsomhet overfor hvilket tidspunkt vi ser på.

Vi ser i tabell 6.8 at for todelt sysselsettingsstatus, blir både samsvar og nivå bra for de ulike månedende. For tredelt sysselsettingsstatus (tabell 6.9) er samsvar bra, men vi får igjen problemer med selvstendignivået. Problemene er imidlertid adskillig mindre enn for fylke: de største utslagene for selvstendignivået (avvik fra AKU) er hhv. -4 prosent og +8 prosent.

*Tabell 6.8 Overensstemmelse mellom sysselsettingsstatus definert ved den foreslåtte klassifikasjonsrutinen og status i AKU, fordelt etter intervjumåned. Todelt sysselsettingsstatus.*

| Måned          | Totalt antall personer | Antall syssel-satte iflg. AKU | Samsvar:   |  |
|----------------|------------------------|-------------------------------|--|--|
|                |                        |                               | andel klassifisert identisk med AKU-status . Prosent | Relativt avvik mellom nivå-tall. Prosent |
| Januar.....    | 6838                   | 4569                          | 89,51  | 0,98                                     |
| Februar.....   | 6921                   | 4720                          | 90,83  | 0,19                                     |
| Mars.....      | 7027                   | 4766                          | 91,02  | -0,57                                    |
| April.....     | 3553                   | 2389                          | 90,18  | -0,13                                    |
| Mai.....       | 3383                   | 2315                          | 90,16  | 0,65                                     |
| Juni.....      | 3423                   | 2343                          | 90,91  | 1,15                                     |
| Juli.....      | 3466                   | 2485                          | 89,99  | -2,13                                    |
| August.....    | 3423                   | 2391                          | 91,09  | -0,04                                    |
| September..... | 3460                   | 2371                          | 91,42  | -0,04                                    |
| Oktober.....   | 1779                   | 1222                          | 91,18  | -0,57                                    |
| November.....  | 1744                   | 1182                          | 90,77  | -0,76                                    |
| Desember.....  | 1787                   | 1187                          | 91,66  | 0,42                                     |

*Tabell 6.9 Overensstemmelse mellom sysselsettingsstatus definert ved den foreslåtte klassifikasjonsrutinen og status i AKU, fordelt etter intervjumåned. Tredelt sysselsettingsstatus.*

| Måned        | Totalt antall personer | Antall lønns-takere iflg. AKU | Antall selv-stendige iflg. AKU | Samsvar:   |  |              |
|--------------|------------------------|-------------------------------|--------------------------------|--|--|--------------|
|              |                        |                               |                                | andel klassifisert identisk med AKU-status . Prosent | Relativt avvik mellom Lønnstakere nivå-tall. Prosent | Selvstendige |
| Januar.....  | 6838                   | 4203                          | 366                            | 88,11  | 1,19   | -1,37        |
| Februar..... | 6921                   | 4305                          | 415                            | 89,28  | 0,32   | -1,21        |
| Mars.....    | 7027                   | 4370                          | 396                            | 89,17  | -0,60  | -0,25        |
| April.....   | 3553                   | 2185                          | 204                            | 88,32  | 0,28   | -4,41        |
| Mai.....     | 3383                   | 2137                          | 178                            | 88,56  | 0,70   | 0,00         |
| Juni.....    | 3423                   | 2140                          | 203                            | 89,19  | 0,47   | 8,37         |
| Juli.....    | 3466                   | 2281                          | 204                            | 87,85  | -2,15  | -1,96        |
| August.....  | 3423                   | 2219                          | 172                            | 89,22  | 0,00   | -0,58        |
| September..  | 3460                   | 2175                          | 196                            | 89,88  | -0,14  | 1,02         |
| Oktober..... | 1779                   | 1126                          | 96                             | 89,04  | -0,80  | 2,08         |
| November..   | 1744                   | 1089                          | 93                             | 88,93  | -1,10  | 3,23         |
| Desember..   | 1787                   | 1081                          | 106                            | 89,76  | 0,37   | 0,94         |

### **6.3 Konklusjon**

Vi har sett at klassifikasjonsrutinen er god for å klassifisere lønnstakere, sysselsatte generelt og ikke-sysselsatte ved at både samsvar og nivå blir bra innenfor ulike delpopulasjoner som fylke, kommune, aldersgruppe, næring m.m. Dette gir god grunn til å tro at klassifiseringsrutinen kan benyttes i FoB2000 til å lage tabeller over sysselsatte (og lønnstakere) på kommunenivå og for ulike grupper innenfor kommunen. For selvstendige er klassifikasjonsrutinen ikke så god for små delpopulasjoner, men dette kan muligens bli bedre for dataene for år 2000, ettersom det i Seksjon 260 arbeides med å få bedre selvstendigregistre.



## 7. Alternative klassifikasjonsrutiner

I tillegg til den foreslåtte klassifikasjonsrutinen for å bestemme sysselsettingsstatus (sysselsettingsdefinisjonen), er en rekke alternative måter undersøkt. Vi skal presentere enkelte av disse utfra problemstillingen: vil mer avanserte rutiner gi klart bedre resultat når det gjelder samsvar, dvs andel personer der registerbasert sysselsettingsstatus stemmer overens med AKU-basert status. Kapittelet består av tre deler. Først vil vi se på hvordan lønn kunne vært brukt mer aktivt i den foreslåtte rutinen. Dernest vil vi se på hvordan vi kan slakke på kravet til datering og hvor fruktbart det er. Det har vært reist tvil om dateringen i enkelte registre, og vi vil se på en metode der vi delvis ser bort fra datering, og der vi også utvider antall trinn i forhold til den foreslåtte metoden. Til slutt vil vi se på en metode som tar utgangspunkt i maksimalt samsvar innenfor delgrupper ved hjelp av lønn.

Bortsett fra i 7.1, er metodene i dette kapittelet prøvd ut for todelt sysselsettingsstatus for enkelhetsskyld, dvs at vi kun ser på sysselsatt vs. ikke-sysselsatt. Dette betyr at vi undervurderer den foreslåtte metoden ved at vi der har vært nødt til å godta lavere samsvar på todelt status for å få mer samsvar på tredelt nivå. Det viser seg imidlertid at de alternative klassifikasjonsrutinene ikke gjør det vesentlig bedre enn den foreslåtte.

### 7.1 Lønn som indikator på sysselsetting

I den utarbeidete klassifikasjonsrutinen presentert i kapittel 5 ble sum lønn av alle arbeidstakerforhold benyttet som indikator på om man er sysselsatt i trinn 3. Blant personene med viktigste arbeidstakerforhold på referansetidspunktet registrert enten kun i AT eller kun i LTO ble de med høyest sum kontantlønn klassifisert som sysselsatt og arbeidstakere. Begrunnelsen for å benytte lønn som tilleggskriterium var at personer med høy lønn i større grad viser seg å være AKU-sysselsatte enn de med lav lønn.

En mulig utvidelse av klassifikasjonsrutinen som ble undersøkt, var å innføre en lønnsgranse for personer i trinn 1. Det vil si at i tillegg til at viktigste arbeidstakerforhold skulle være registrert både i AT og LTO på referansetidspunkt, måtte personene ha en samlet lønn fra alle arbeidstakerforhold, over en gitt grense. Bakgrunnen var å redusere andelen feilklassifiserte blant personer med et koplet registerforhold.

Tabell 7.1 fordeler alle personene som blir klassifisert som selvstendig og arbeidstaker i trinn 1, etter hvilken sysselsettingstatus de har i AKU og etter samlet kontantlønn. Utfra tabellen kan vi se effekten av å innføre en lønnsgranse. La oss si at vi setter som tilleggskriterium i trinn 1 at personer må ha en samlet kontantlønn som overstiger 5 000 kroner. Dette gir som resultat at vi fjerner 91 personer som i AKU oppgir at de ikke er sysselsatt. Imidlertid vil vi også miste 59 personer som er sysselsatt i følge AKU. Vi konkluderer dermed med at siden andelen riktig registersysselsatte forbedres i liten grad, er det ingen grunn til å komplisere rutinen ved å benytte en lønnsgranse som tilleggskriterium på trinn 1.

Tabell.7.1 AKU-status for ulike lønnsgrupper blant de som er klassifisert som registersysselsatt på trinn 1.

|                  | Ikke AKU-<br>sysselsatt | AKU-sysselsatt<br>selvstendig | AKU-sysselsatt<br>arbeidstaker | I alt |
|------------------|-------------------------|-------------------------------|--------------------------------|-------|
| under 5 000      | 91                      | 10                            | 49                             | 150   |
| 5 001 - 10 000   | 75                      | 14                            | 112                            | 201   |
| 10 001 - 20 000  | 178                     | 25                            | 327                            | 530   |
| 20 001 - 30 000  | 120                     | 22                            | 334                            | 476   |
| 30 001 - 40 000  | 86                      | 30                            | 365                            | 481   |
| 40 001 - 50 000  | 81                      | 18                            | 356                            | 455   |
| 50 001 - 60 000  | 57                      | 17                            | 398                            | 472   |
| 60 001 - 70 000  | 46                      | 15                            | 448                            | 509   |
| 70 001 - 80 000  | 30                      | 11                            | 614                            | 655   |
| 80 001 - 90 000  | 34                      | 8                             | 729                            | 771   |
| 90 001 - 100 000 | 35                      | 11                            | 752                            | 798   |
| over 100 000     | 218                     | 100                           | 21304                          | 21622 |
| I alt            | 1051                    | 281                           | 25788                          | 27120 |

## 7.2 Tidsavstand

Hittil har vi stort sett benyttet oss av kriteriene at et registerforhold må være aktivt som viktigste forhold i referanseuka for å kvalifisere til at personen blir registersysselsatt. Imidlertid har vi opplysninger om nøyaktige datoer både for referanseuka (AKU-uka) og registerforholdenes start- og sluttdatoer, og vi så i kapittel 4 at det er beregnet tidsavstand for alle forhold, der tidsavstand er korteste avstand mellom referanseuka og tidsrommet der registerforholdet er aktivt (som viktigste forhold). Figur 4.1 illustrerer situasjonen for en tenkt person, og der blir tidsavstanden lik 0. En person som intervjues midt i juni, og hvor eneste arbeidsforhold er aktivt fra januar til midten av juni, vil få en tidsavstand på ca. 30 dager. Vi har tidligere i kapittelet sett at det å ha et koplet forhold er det mest sentrale kriterium for å klassifiseres som sysselsatt. Vi vil derfor holde oss til dette kriteriet her, men der vi tidligere i kapittelet krevde at forholdet skal være aktivt som et viktigste-forhold i referanseuka, vil vi nå se på alternative krav til tidsavstanden enn at den skal være 0.

Vi kan se på bruken av tidsavstand som et vindu vi legger rundt referanseuka. Dersom det komplette forholdet er aktivt innenfor vinduet, klassifiserer vi personen som sysselsatt. Det trangeste vinduet er når vinduet akkurat dekker referanseuka og tidsavstanden er null dager fra referanseuka, og da er vi i situasjonen fra trinn I i 5.1. Vi kan imidlertid velge et større vindu/større tidsavstand. Vinduet som har størrelse 40 er f.eks det vinduet som strekker seg 40 dager i hver retning fra start og stopp av referanseuka. Dette tilsvarer tidsavstand på 0-40 dager, og dersom vi velger en slik tidsavstand, vil personen ovenfor med forhold som avsluttes i juni og intervju i juli, klassifiseres som sysselsatt.

Tabell 7.2 Andel og antall AKU-sysselsatte blant personer med koblet forhold etter hvor stor tidsavstand det er mellom referanseuka og registerforholdet (som viktigste arbeidsforhold)

| Personer med koblet arbeidsforhold | Tidsavstand i dager |       |        |         |         |         |          |           |           |          |
|------------------------------------|---------------------|-------|--------|---------|---------|---------|----------|-----------|-----------|----------|
|                                    | 0                   | (0,7) | [7,14) | [14,21) | [21,28) | [28,56) | [56,112) | [112,224) | [224,366) | missing* |
| Andel AKU-sysselsatte              | 96 %                | 45 %  | 41 %   | 40 %    | 34 %    | 36 %    | 37 %     | 34 %      | 46 %      | 29 %     |
| Antall AKU-sysselsatte             | 25856               | 98    | 93     | 72      | 38      | 211     | 336      | 353       | 139       | 4744     |

\* Forhold som enten ikke har eksisterende dato for når det komplette forholdet er viktigste forhold, eller der vedkommende ikke har koblet forhold

Vi vil se på de personene vi legger til som sysselsatte når vi øker vinduet, og vi ser i tabell 7.2 at blant de som har aktivt forhold med tidsavstand 0, er det (heldigvis) mer enn 50 prosent AKU-sysselsatte. Imidlertid, allerede for de som har aktivt forhold først når avstanden til referanseuka er mellom 1 og 7 dager, er det flere ikke-sysselsatte enn sysselsatte. Som et resultat av dette gir det ikke bedre resultat å slakke på kravet til tidsavstand. Ettersom vi slakker mer og mer på kravet til tidsavstand, ser vi at andelen AKU-sysselsatte blir stadig lavere blant de nye som blir registersysselsatte, unntatt når vi slakker kravet fra 224 dager til hele året,<sup>12</sup> men da er også de nye registersysselsatte i større grad ikke-sysselsatte enn sysselsatte iflg. AKU.

### 7.2.1 Asymmetrisk vindu

Dersom stoppdatoene er mer upresise enn startdatoene kan vi tenke oss å godta mye bredere vindu etter referanseuka enn før, og motsatt dersom det er startdatoene som er mest upresise. Hensikten med å øke vinduet er å senke kravet til å bli registersysselsatt og dersom det skal ha noen hensikt må det være fordi man kan fange opp noen flere AKU-sysselsatte. Vi konsentrerer oss derfor nå om AKU-sysselsatte med koblet arbeidsforhold som ikke er aktivt i referanseuka. Det finnes da to mulige situasjoner, illustrert i figur 7.1.

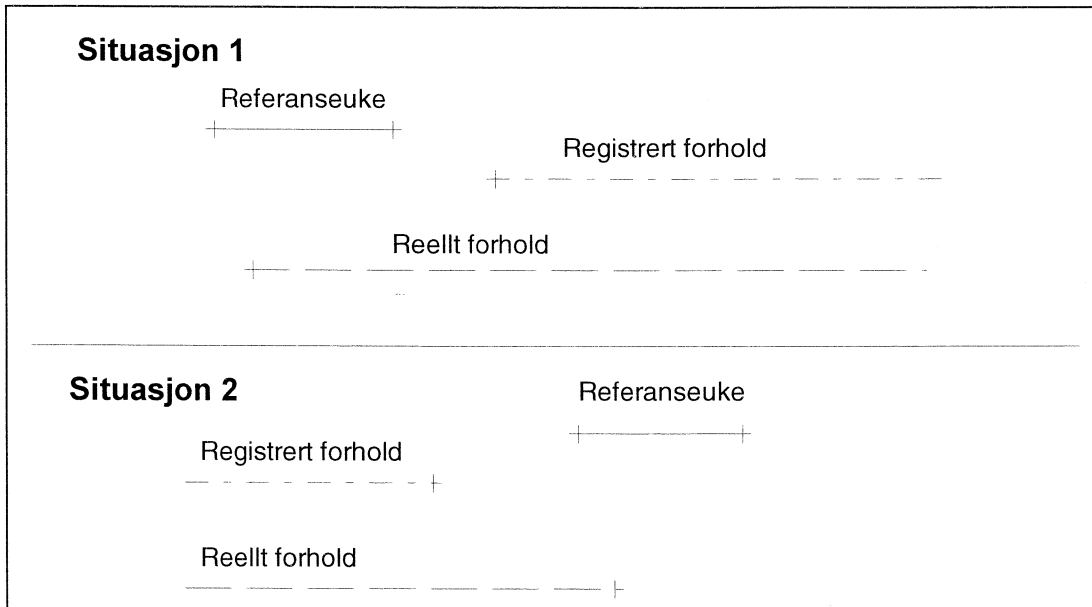
I situasjon 1) i figur 7.1 blir forholdet registrert innmeldt på et for sent tidspunkt. Ved å kun øke vindusbredden (fra 0) på høyre side (dvs etter) referanseuka, vil man kunne fange opp disse. Dette er prøvd ut, men det viser seg at de nye personene vi klassifiserer som sysselsatte dersom vi utvider vinduet kun på høyre side istedet for å kreve treff i referanseuka, er mer ikke-sysselsatte enn sysselsatte, dvs at vi fremdeles bør kreve treff i referanseuka.

I situasjon 2) registreres arbeidsforholdet å ha opphørt før det faktisk har det. Det er vanskelig å motivere hvorfor det skulle eksistere en slik tendens i registreringssystemet, og det viser seg også her at vinduet bør være referanseuka, dvs at vi krever treff i referanseuka.

Analogt med at starttidspunktet kan være forskjøvet mot et for sent tidspunkt (tilfelle 1) kunne vi tenke oss at det samme kan gjelde stopptidspunktet, dvs at registerforholdet skulle vart i et kortere tidsrom enn det vi har registrert. Dersom dette er tilfellet vil vi få færre personer med aktive arbeidsforhold, mens vårt hovedproblem er det motsatte: vi ikke fanger opp nok av de faktisk sysselsatte.

<sup>12</sup> Tendensen når tidsavstandskravet slakkes opp til et helt år, kunne kanskje skyldes at en del arbeidsgivere som har glemt å rapportere inn et forhold, registrerer personen som sysselsatt f.eks kun i desember for å dekke over forsinkelsen. Dette ville det kunne være interessant å undersøke nærmere bl.a fordi det kan være relevant for korttidsstatistikken

Figur 7.1 Illustrasjon av hvilke mulige avvik fra referanseuka et registerforhold kan ha som gjør at det kan være relevant å tillate slakkere krav til tidstavstand mellom referanseuka og registerforholdet hhv. etter og før referanseuka



Vi har ovenfor sett at det i utgangspunktet er fornuftig å kreve at et koplekt forhold må treffe referanseuka for å kvalifisere til statusen sysselsatt. Senere skal vi imidlertid se at vi likevel kan slakke på kravet til datering dersom vi innfører tilleggskrav om høy lønn.

### 7.3 Enkel utvidelse av den foreslåtte klassifikasjonsrutinen fra tre til fem trinn

Vi vil nedenfor presentere en utvidelse av klassifikasjonsrutinen fra kap 5 der vi modifierer trinn 3 slik at vi totalt får fem trinn. Metoden vi ser på er en naturlig utvidelse spesielt i den situasjonen at vi ikke har tiltro til dateringer verken når de stammer fra LTO eller selvstendigfilene. Mens vi i vår foreslåtte metode krevde korrekt datering, og i 7.2 så på varierende grad av å slakke på kravet, tenker vi oss nå at vi bortsett fra for koplete forhold ikke har noe krav til datering i det hele tatt. Vi vil både se på hva en slik utvidelse innebærer, hvor komplisert den er, og vi vil se på hvilke avveininger som må gjøres pga at vi har fem trinn samt at vi for mange personer ikke har dateringer. Til slutt vil vi vurdere idéen med fem trinn utfra hvor gode resultater det gir. Det viser seg til slutt at resultatene ikke er veldig mye bedre enn den foreslåtte metoden, heller ikke dersom vi benytter dateringer.

Når vi ser helt bort fra dateringer for de fleste typer forhold, kan det synes merkelig utfra at vi hittil nettopp har krevd at forholdet skal være aktivt i referanseuka for å bli tatt i betraktning. Imidlertid viser det seg å være en del sysselsatte som har feildateringer, selv etter de datojusteringene som er beskrevet i kapittel 3. I og med at vi ikke har datering, må vi prioritere mellom ulike forhold utfra andre kriterier når vi skal knytte riktig næring og andre jobbspesifikke variable til en person som har flere jobbforhold. I utgangspunktet velger vi det forholdet som er årsaken til at personen blir klassifisert som sysselsatt, men dersom personen har flere forhold må andre metoder benyttes, noe vi imidlertid ikke ser på her.

Klassifikasjonsrutinen kan skisseres som følger:

Trinn 1: som i den foreslåtte rutinen, dvs de med koplekt forhold som viktigste forhold i referanseuka, blir klassifisert som sysselsatte.

Trinn 2: alle med selvstendigforhold uansett datering blir klassifisert som sysselsatte.<sup>13</sup>

Trinn 3: Alle med koplet forhold og høy samlet lønn blir klassifisert som sysselsatte.

Trinn 4: Alle med ukoplet AT-forhold og høy samlet lønn blir klassifisert som sysselsatte (at de har positiv lønn betyr at de har et LTO-forhold også).<sup>14</sup>

Trinn 5: Blant alle som ennå ikke er klassifisert som sysselsatte, blir de med høyest samlet lønn klassifisert som sysselsatte, slik at totalt antall som er klassifisert som sysselsatt blir likt antall sysselsatte iflg. AKU.

Rekkefølgen av trinn 3 og trinn 4 er litt vilkårlig, mens vi betrakter personene med ukoplet LTO-forhold på siste trinn fordi at denne gruppen er så stor at vi ikke risikerer å gå tom for personer når vi skal oppnå AKU-nivået.

På trinn 3 og 4 er det ikke så opplagt hvordan vi skal sette lønns grensen. Vi må avveie mellom to forhold. For det første ønsker vi at flest mulig klassifiseres likt som den statusen de har i AKU (dvs høyest mulig samsvar), men vi ønsker også å komme opp på AKU-nivået til slutt. Dersom kravet til samsvar har vært satt for høyt på trinn 3-4, må vi for å få nok registersysselsatte totalt, sette lønns grensen så lavt på trinn 5 at samsvaret blir dårlig her. Maksimalt samsvar for de personene vi vurderer på trinn 3 og 4 er ikke nødvendigvis noen god idé når vi skal opp på AKU-nivået til slutt.

Dersom vi på trinn 3 og trinn 4 prioriterer at flest mulig av de vi klassifiserer som registersysselsatte også er AKU-sysselsatte, men med sideblikk til at kravet ikke skal være for strengt, ville et naturlig utgangspunkt f.eks vært at i alle lønnsgrupper der man blir registersysselsatt, så skal minst P prosent faktisk være det, vi kaller regelen "P-prosentregelen". Her må parameteren P settes etter skjønn slik at vi får en gunstig avveining mellom nivå og samsvar. Problemet med metoden er at en person må sitte og inndele i naturlige lønnsgrupper, og dessuten skjønnsmessig vurdere hva man skal gjøre dersom andelen flere ganger mellom over og under P prosent når vi ser på ulike lønnsgrupper etter stigende lønn. Det er vanskelig å få et dataprogram til å skille mellom at andelen tilfeldigvis er over eller under P prosent for f.eks 60 000-70 000 kr, og den lønnsverdien som iflg hovedtrenden skiller mellom "over P prosent" og "under P prosent", f.eks 35 000-40 000 kr. Et annet problem med P-prosentregelen er at det har betydning for utfallet hvordan vi inndeles i lønnsgrupper, for 35 000-45 000 kan gi et annet bilde enn dersom vi ser på 30 000-40 000 og 40 000-50 000.

For å få en mer automatisk rutine, som både er enkel å forstå og å gjennomføre, kan vi sette som krav at av alle de som vi klassifiserer som sysselsatte på et trinn, skal minst f.eks 67 prosent av dem være riktig klassifisert. Vi begynner med å klassifisere de med høyest lønn, og vurderer fortløpende hvor mange registersysselsatte som er riktig klassifisert. Tallet skal starte på omtrent 100 prosent og så synke gradvis når lønnsnivået synker. Imidlertid kan vi oppleve at personen med høyest lønn tilfeldigvis er ikke-sysselsatt og da er andelen riktige null prosent. For å unngå at vi da stopper med en gang og sier at ingen kan bli registersysselsatte, må vi vente til f.eks 10 personer er blitt klassifisert før vi måler andelen riktige, for da har vi så mange personer at tilfeldige utslag ved noen få enkeltpersoner ikke kan stoppe hele klassifikasjonen. En svakhet ved metoden er også at tallet 67 prosent må settes utfra en skjønnsmessig vurdering av hvor mye vi kan oppnå. Jo bedre registergrunnlaget er, jo større kan tallet settes. På trinn 3 og 4 er registerinformasjonen såpass usikker at et såpass lavt tall som 67 prosent er bra. Med metoden nevnt ovenfor og fem trinn, får vi flg. sluttresultatet som vist i tabell 7.3.

---

<sup>13</sup> Grunnen til at vi ikke har med krav om lav lønn slik som forslaget i kap. 5, skyldes utelukkende at vi kun ser på todelt sysselsettingsstatus. Ettersom vi da ikke trenger å begrense antall selvstendige, kan vi kun betrakte hvor stor andel som er AKU-sysselsatt, noe som leder til at vi klassifiseres alle som sysselsatte.

<sup>14</sup> Ettersom vi betrakter datering fra AT som pålitelig, kunne vi betraktet ukoplet forhold i referanseuka før vi så på de som ikke har datering, men dette er så få personer at vi heller prioriterer å begrense antall trinn.

Tabell 7.3 Sammenlikning mellom AKU-basert sysselsettingsstatus og registerbasert der vi har fem trinn og til en viss grad ser bort fra dateringer.

| Sysselsettingsstatus ifølge registrene: | Sysselsettingsstatus ifølge AKU |            |        |
|---|---------------------------------|------------|--------|
|   | Ikke sysselsatt                 | Sysselsatt | Totalt |
| Ikke sysselsatt                         | 12 715                          | 2 149      | 14 864 |
| Sysselsatt                              | 2 149                           | 29 791     | 31 940 |
| Totalt                                  | 14 864                          | 31 940     | 46 804 |

Dette gir samsvar på 90,82 prosent, noe som er litt bedre enn for den foreslåtte metoden. Dersom vi modifiserer den foreslåtte metoden slik at vi ikke begrenser antall selvstendige, blir samsvaret for den metoden 90.5 prosent. Dette betyr at vi kun får en ganske begrenset forbedring dersom vi utvider fra tre til fem trinn. Dersom vi tar hensyn til datering for alle typer forhold, slik vi har gjort med den foreslåtte rutinen, gir en liknende fem-trinnsmetode et liknende resultat som det vi ser i tabell 7.3.

## 7.4 Metode basert på maksimalt samsvar

Den foreslåtte klassifikasjonsrutinen i kap. 5 er basert på en avveining mellom ulike hensyn, både enkelhet, samsvar og riktig nivå, og der fokus har vært på en skrittvis klassifikasjon der man etter et trinn kan vurdere hva man skal gjøre videre basert på resultatene fra de tidligere trinnene. Vi vil nedenfor se at den foreslåtte rutinen ikke har mye lavere samsvar i forhold til en mer avansert rutine basert på maksimalt samsvar, og videre viser det seg at selv med todelt sysselsettingsstatus som vi ser på, blir rutinen ganske komplisert.

### 7.4.1 Inndeling i disjunkte grupper

Vi inndeler hele utvalget i ikke-overlappende (disjunkte) kategorier etter hva slags arbeidsmarkedsregisterforhold hver av personene har. Vi betrakter inndeling kun etter nærmeste type arbeidsforhold (som er viktigst), avstand til referanseuka (AKU-uka) og forventet arbeidstid, og de kategoriene vi kommer fram til er beskrevet i tabell 7.4.

Tabell 7.4 En inndeling i disjunkte (ikke-overlappende) grupper der man både benytter type forhold som er viktigst nærmest referanseuka, avstanden til referanseuka og forventet arbeidstid

| gruppe  | Nærmeste viktigste forhold    | avstand $d$ til referanseuka | Forventet arb.tid (t/uke) |
|---------|-------------------------------|------------------------------|---------------------------|
| 1.....  | Koplet forhold                | $d = 0$ dager                | 4 -19                     |
| 2.....  | ----- “ -----                 | --- “ ---                    | 20 - 29                   |
| 3.....  | ----- “ -----                 | --- “ ---                    | 30 +                      |
| 4.....  | ----- “ -----                 | $0 < d \leq 10$ dager        | 4 -19                     |
| 5.....  | ----- “ -----                 | --- “ ---                    | 20 - 29                   |
| 6.....  | ----- “ -----                 | --- “ ---                    | 30 +                      |
| 7.....  | ----- “ -----                 | $d > 10$ dager               | 4 -19                     |
| 8.....  | ----- “ -----                 | --- “ ---                    | 20 - 29                   |
| 9.....  | ----- “ -----                 | --- “ ---                    | 30 +                      |
| 10..... | Selvstendig jordbruk, skog... |                              |                           |
| 11..... | Selvstendig andre næringer    |                              |                           |
| 12..... | Ukoplet ATreg-forhold         | $d = 0$ dager                |                           |
| 13..... | Ukoplet ATreg-forhold         | $d < 0$ dager                |                           |
| 14..... | Ukoplet LTO-forhold           | $d = 0$ dager                |                           |
| 15..... | Ukoplet LTO-forhold           | $d < 0$ dager                |                           |

Vi har som utgangspunkt at vi krysser type forhold med en tredelt “avstand-fra-AKU-uke”-variabel, samt med forventet arbeidstid (som også er tredelt). Dette gir  $5 \times 3 \times 3 = 45$  kategorier, men forventet arbeidstid er kun definert i Atreg, og i tillegg ønsker vi å begrense antall grupper: vi slår sammen de ulike forventet arbeidstid for ukoplet Atreg-forhold og slår sammen “avstand fra AKU” til hhv. en og to verdier for hhv. selsvtendige og ukoplet Atreg/ukoplet LTO. Dette gir 15 kategorier som både er disjunkte og som fanger opp alle personene vi ønsker å se på.

Dersom man ikke deler inn i disjunkte grupper, men derimot inndeler f.eks slik som i fem-trinnsmetoden fra 7.2, kan vi ikke betrakte hver gruppe separat når man skal finne lønnsgrænse som gir maksimalt samsvar. Kategoriernes innhold bestemmes av i hvilken rekkefølge de betraktes, og metoden vil bare gi maksimalt samsvar for denne rekkefølgen.

#### **7.4.2 Idealisert metode**

Innenfor hver av de disjunkte kategoriene beskrevet ovenfor, bestemmer vi en lønnsgrænse<sup>15</sup> som klassifiserer alle med høyere lønn som registersysselsatte, mens de under grensen blir ikke-sysselsatte. Grensen bestemmes i utgangspunktet slik at samsvaret for  $2 \times 2$ -tabell (ikke-sysselsatt/sysselsatt) blir størst mulig, dvs at andelen personer som er klassifisert riktig i forhold til AKU, enten som sysselsatt eller ikke-sysselsatt, blir størst mulig. Denne metoden vil være analog med det vi ville fått med en passende spesifisert regresjonsmodell.

#### **7.4.3 Modifisert metode**

Det er to aspekter som gjør at metoden ovenfor må modifiseres. Det ene er at personer med ikke-positiv lønn må behandles spesielt og det andre er at vi ønsker å treffe AKU-nivået for antall sysselsatte.

Det kan oppstå et teknisk problem dersom lønnsgrænse er null eller manglende verdi. Mange personer innenfor flere av disse 15 gruppene har hver av disse verdiene, og vi vil unngå at personer med samme verdi klassifiseres ulikt. Vi betrakter verdi null og “manglende verdi” å være kvalitativt det samme. Vi må ha en strategi dersom den lønnsgrænsen som gir maksimalt samsvar skulle være null eller “manglende verdi”, for da vil det gi markert forskjellig resultat avhengig av om de med lønn eksakt null eller “manglende verdi” blir registersysselsatte eller ikke. I en slik situasjon velger vi å la disse personene være ikke-registersysselsatte.

At vi ønsker å nå AKU-tallet best mulig, viser seg i praksis å bety at vi må ta med flere personer som registersysselsatte enn vi ellers ville gjort. For å ta med de ekstrapersonene som skader samsvaret minst mulig, senker vi innenfor hver av de 15 gruppene lønnsgrænsen slik at samsvaret minker med  $k$  prosent i forhold til det maksimale samsvaret. Vi kjører gjennom opplegget med ulik parameter  $k$  og velger den  $k$  som gir best treff av AKU-nivået.

#### **7.4.4 Resultat**

Resultatet med fremgangsmåten skissert ovenfor der vi senker samsvaret med  $k=0,7$  prosent er gitt i tabell 7.5, og samsvaret blir 91,02 prosent, som altså er andelen i de grå feltene. Dersom vi deler opp sysselsatte i selvstendig og lønns/arbeidstaker, så kan vi se fra tabell 7.6 at samsvaret nå er 89,38 prosent, men metoden er ikke laget med tanke på å få maksimalt samsvar for tredelt sysselsettingsstatus, og videre får vi ikke riktig selvstendignivå. Det er imidlertid vanskelig å treffe AKU-nivået eksakt. Vi overestimerer totalt antall sysselsatte i forhold til AKU med ca. 1 prosent, men dette er det nærmeste vi klarer å komme gjennom å justere parameteren  $k$ . Dette er naturligvis en svakhet ved metoden.

---

<sup>15</sup> Lønn er her sum av alle personens kontantlønner for personer hvis viktigste forhold er arbeidstaker-/lønnstakerforhold, og næringsinntekt (Likningsregisteret) for selvstendigforhold fra hhv. jordbruk/skogbruk/fiske og andre næringer.

Tabell 7.5 Samsvar mellom AKU-sysselsettingsstatus og registerbasert status med klassifikasjonsrutine basert på maksimalt samsvar med kategoriene i tabell 7.4. Sysselsettingsstatus er todelt

| Sysselsettingsstatus ifølge registrene: | Sysselsettingsstatus ifølge AKU |            |        |
|---|---------------------------------|------------|--------|
|   | Ikke sysselsatt                 | Sysselsatt | Totalt |
| Ikke sysselsatt                         | 12 606                          | 1 942      | 14 548 |
| Sysselsatt                              | 2 258                           | 29 998     | 32 256 |
| Totalt                                  | 14 864                          | 31 940     | 46 804 |

Tabell 7.6 Samsvar mellom AKU-sysselsettingsstatus og registerbasert status med klassifikasjonsrutine basert på maksimalt samsvar med kategoriene i tabell 7.4. Sysselsettingsstatus er tredelt ved at sysselsatt er delt i to deler.

| Sysselsettingsstatus ifølge registrene: | Sysselsettingsstatus ifølge AKU |             |              | Totalt |
|---|---------------------------------|-------------|--------------|--------|
|   | Ikke sysselsatt                 | Selvstendig | Arbeidstaker |        |
| Ikke sysselsatt                         | 12 554                          | 263         | 1 649        | 14 466 |
| Selvstendig                             | 406                             | 2 143       | 523          | 3 072  |
| Arbeidstaker                            | 1 904                           | 223         | 27 139       | 29 266 |
| Totalt                                  | 14 864                          | 2 629       | 29 311       | 46 804 |

### 7.3.5 En vurdering av metoden

Selv om vi skulle klare å justere metoden slik at vi treffer AKU-nivået eksakt, så vil dette ta ekstra tid og gjøre metoden mer komplisert sett fra et produksjonssynspunkt. Fra før er metoden relativt komplisert i den forstand at programmet er litt vanskelig å forstå, spesielt med modifisert metode. Sistnevnte bidrar også til at metoden begrepsmessig blir litt komplisert. Metoden tar ikke hensyn til at personer med mye registrert informasjon er å foretrekke dersom dette ikke går ut over samsvaret i noen særlig grad. Sist, men ikke minst, blir ikke samsvaret mer enn ca. ett prosentpoeng bedre.



## 8. Analyse av personer uten registrerte arbeidsforhold

Utvalget som lå til grunn for utarbeidelsen av klassifikasjonsrutinen bestod av personer som hadde svart på Arbeidkraftsundersøkelsen (AKU), og som vi fant informasjon om i registrene i løpet av året. Vi så med andre ord bort fra frafallet i AKU-utvalget, og de personene som hadde svart på AKU men som ikke var å finne i registrene. Problemet med en slik seleksjon ligger i muligheten for at vi systematisk kan ha utelatt enkeltgrupper, fordi disse er sterkt overrepresentert i den delen av AKU-utvalget vi har valgt bort. Sagt på en annen måte kan vi ved vår utvalgsmetode ha forbigått personer fra grupper som registrene har dårlig informasjon om. Dette fører igjen til at en undersøkelse av hvor god klassifikasjonsrutinen er ved hjelp av samsvarsanalyser, gir en overvurdering av kvaliteten ved at vi ser bort fra “problemgruppene” i analysen.

For å få en oversikt over sammensetningen av den delen av populasjonen vi har valgt å utelate skal vi i dette kapitlet se nærmere på de personene som har svart på AKU, og som vi ikke finner informasjon om i registrene. Dette vil også kunne bidra til å få et bilde av om det er enkelt grupper som er av dårlig kvalitet i registrene.

### 8.1 Hovedtrekk

Tabell 8.1 fordeler nettoutvalget i AKU etter om det finnes opplysninger om arbeidsforhold / lønnsforhold i registrene i løpet av året, og om personene ifølge AKU er sysselsatt eller ikke-sysselsatt. Som nevnt i kapittel 4.2 utgjøres gruppe B av personer hvor vi både har register- og AKU-informasjon. For personer i gruppe A har vi derimot kun AKU-informasjon. Tabellen viser som forventet at den største delen av gruppe A utgjøres av de som i AKU oppgir at de ikke er sysselsatte, nærmere bestemt 71 prosent. I den videre analysen er vi derimot interessert i å beskrive de 307 som har oppgitt i AKU at de er sysselsatt. Siden vi ikke har registerinformasjon om denne gruppen vil vi ved hjelp av klassifikasjonsrutinen feilaktig definere disse som ikke sysselsatte.

Tabell 8.1 Nettopopulasjonen i AKU fordelt på gruppe og status som sysselsatt.

| Yrkesstatus     | Gruppe A (kun AKU) |         | Gruppe B (AKU/reg.) |         |
|-----------------|--------------------|---------|---------------------|---------|
|                 | Antall             | Prosent | Antall              | Prosent |
| Sysselsatt      | 307                | 10      | 31 956              | 63      |
| Ikke sysselsatt | 2 243              | 72      | 14 865              | 29      |
| Uoppgitt        | 580                | 18      | 3 939               | 8       |
| I alt           | 3 130              | 100     | 50 760              | 100     |

I tabell 8.2 har vi trukket ut de personene som har oppgitt at de er sysselsatt ifølge AKU, og fordelt disse på statuskoder. Av det totale antall AKU-sysselsatte er det 1 prosent som ikke er å finne i registrene. Dette utgjør 307 personer. For å få et bilde av hvor mange personer dette kan representere av totalpopulasjonen, beregner vi et anslag ved hjelp av oppblåsningsfaktoren i AKU. Oppblåst utgjør de 307 AKU sysselsatte som vi feilklassifiserer 20 267 personer.

Tabell 8.2 AKU sysselsatte fordelt etter om de i løpet av året har et eller annet registerforhold, og statuskode. Absolutte tall, oppblåste tall, og prosent

| Statuskode    | Gruppe A |          |         | Gruppe B |           |         | I alt  |           |         |
|---------------|----------|----------|---------|----------|-----------|---------|--------|-----------|---------|
|               | Antall   | Oppblåst | Prosent | Antall   | Oppblåst  | Prosent | Antall | Oppblåst  | Prosent |
| Vernepliktige | 26       | 1 746    | 7       | 364      | 23 862    | 93      | 390    | 25 608    | 100     |
| Selvstendige  | 114      | 7 647    | 4       | 2 632    | 17 814    | 96      | 2 746  | 25 461    | 100     |
| Ansatte       | 116      | 7 703    | 0       | 28 504   | 1 758 934 | 100     | 28 620 | 1 766 637 | 100     |
| Fam. arb.     | 43       | 2 657    | 10      | 367      | 22 883    | 90      | 410    | 25 540    | 100     |
| Uoppgitt      | 8        | 514      | 8       | 89       | 5 954     | 92      | 97     | 6 468     | 100     |

De vernepliktige (statuskode 102) fanges i utgangspunktet ikke opp i registrene. Vi ser allikevel at 93 prosent av disse har et lønnsforhold eller arbeidstakerforhold i løpet av året. En mulige årsak til dette kan være at personer som svarer i AKU at de er vernepliktige, i løpet av året har hatt annet arbeid som registreres i registrene. Dette kan enten være deltidsarbeid under vernepliktsperioden, eller at de har vært i arbeid før, eventuelt etter perioden er avsluttet. I tillegg kan enkelte vernepliktige få en lønns- og trekkoppgave fordi de utfører spesielle arbeidsoppgaver under millitærtjenesten.

Familiearbeiderne (statuskode 113 og 123) er også en gruppe som man skulle anta i liten grad er oppført i registrene. Tabellen bekrefter dette til en viss grad. Av de familiearbeiderne som oppgir at de er sysselsatt i AKU, er 10 prosent ikke i å finne i registrene. Utfra denne analysen kan vi ikke si noe om hvorvidt de resterende egentlig kan defineres som familiearbeidere, eller om de har andre yrkesforhold som gjør at de er oppført i registrene.

Som vi har vært inne på før i dette notatet, kan vi ved hjelp av registrene bare identifisere de selvstendige ved hjelp av pensjonsgivende inntekt fra Likhingsregisteret. Dette kan være noe av årsaken til at vi av gruppen selvstendige (statuskode 111 og 121) kun finner igjen 93 prosent i registrene. Det vil si at 114 (oppblåst 7 647) personer som ifølge AKU er selvstendige, feilaktig blir klassifisert som ikke-sysselsatt ifølge klassifikasjonsrutinen fordi vi mangler registerinformasjon.

Den største delen av de som er definert som sysselsatt har statuskode 112 og 122, det vil si at de er ansatt. Prosentmessig ser vi at personene i denne gruppen i stor grad fanges opp av registrene. Det er imidlertid verdt å merke seg at antallet som befinner seg i gruppe A er 116 personer, oppblåst 7 703.

Disse resultatene viser som forventet at det særlig er blant de vernepliktige, familiearbeidere og selvstendige hvor avvikene mellom AKU og registrene er størst. Naturlig nok utgjør antall ansatte også en stor del av gruppe A, men det prosentmessig avviket i forhold til registrene er lite. Det kan allikevel være hensiktsmessig å se litt nærmere på hva som karakteriserer de ansatte som ikke fanges opp av registrene.

Vi skal videre gå litt nærmere inn bak tallene i tabell 7.2, ved å se på hvordan personene fordeler seg på andre kjennetegn som kjønn, alder, næring, men først vil vi ta for oss hvor mange det er i gruppe A som har pensjonsgivende inntekt.

## 8.2 En utdypning av resultatene i analysen av personer uten registrerte arbeidsforhold

Vi skal i dette delkapittelet utdype analysen ovenfor ved å se på hvilke andre kjennetegn som beskriver personene i gruppe A. Antall personer er i teksten oppgitt som 'det faktiske antall AKU sysselsatte' / 'oppblåste tall'.

### Pensjonsgivende inntekt

En pekepinn på om personene i gruppe A kan ha oppgitt feil svar i AKU, eller om det skyldes manglende registerføringer, kan vi få ved se hvor mange av personene som har pensjonsgivende inntekt. Tabell 8.3 gir en oversikt over dette.

Tabell 8.3 AKU sysselsatte i gruppe A, henholdsvis med og uten pensjonsgivende inntekt (PGI).

| Statuskode    | Gruppe A med PGI |          |         | Gruppe A uten PGI |          |         | I alt  |          |         |
|---------------|------------------|----------|---------|-------------------|----------|---------|--------|----------|---------|
|               | Antall           | Oppblåst | Prosent | Antall            | Oppblåst | Prosent | Antall | Oppblåst | Prosent |
| Vernepliktige | 1                | 75       | 4       | 25                | 1 671    | 96      | 26     | 1746     | 100     |
| Selvstendige  | 18               | 1 232    | 16      | 96                | 6 415    | 84      | 114    | 7647     | 100     |
| Ansatte       | 12               | 794      | 10      | 104               | 6 909    | 90      | 116    | 7703     | 100     |
| Fam.arb.      | 2                | 101      | 5       | 41                | 2 557    | 95      | 43     | 2658     | 100     |
| Uoppgitt      | 1                | 74       | 13      | 7                 | 440      | 88      | 8      | 514      | 100     |

Denne viser at hele 89 prosent av de personene som ifølge AKU er sysselsatt, men ikke er å finne i registrene, heller ikke pensjonsgivende inntekt i 1992 ifølge likningsregisteret. Dette kan skyldes flere faktorer.

- 1) Personen kan ha oppgitt gale opplysninger om sin status til AKU.
- 2) Svart arbeid som ikke fanges opp av registrene.
- 3) På grunn av defenisjonsforskjeller mellom AKU og registrene, kan en person være i arbeid i en så kort periode, eventuelt motta en så lav lønn, at arbeidsforholdet ikke fanges opp av registrene.

### Alder

Gruppen 16-19 år er overrepresentert i gruppe A. Av de totalt 1 167 / 71 491 AKU sysselsatte i denne aldersgruppen er det 5 prosent (58 / 3 576) som ikke er i registrene. Ser vi nærmere på hvilken statuskode disse har, finner vi at de for en stor del er ansatte (37 / 2 287) og familiearbeidere (12 / 732).

Tabell 8.4 Personer sysselsatt ifølge AKU fordelt på alder. Absolutte tall og prosent.

| Alder    | Gruppe A (kun AKU) |         | Gruppe B (AKU/reg.) |         | I alt  |         |
|----------|--------------------|---------|---------------------|---------|--------|---------|
|          | Antall             | Prosent | Antall              | Prosent | Antall | Prosent |
| 16-19 år | 58                 | 5       | 1 109               | 95      | 1 167  | 100     |
| 20-29 år | 67                 | 1       | 7 047               | 99      | 7 114  | 100     |
| 30-49 år | 119                | 1       | 16 430              | 99      | 16 549 | 100     |
| 50-66 år | 62                 | 1       | 6 938               | 99      | 7 000  | 100     |
| 66 år -  | 1                  | 0       | 432                 | 100     | 433    | 100     |

## Kjønn

Kvinner utgjør en større andel av gruppe A enn de gjør av gruppe B. Forskjellene kommer tydeligere frem når vi tar i betraktning hvorvidt disse er ansatt, selvstendige, eller familiearbeidere. Av de sistnevnte gruppene er det henholdsvis 9 prosent (55 / 3593) og 13 prosent (30 / 1 850) av de sysselsatte kvinnene som ikke fanges opp i registrene, mot 3 prosent og 7 prosent for menn.

Tabell 8.5 Personer sysselsatt ifølge AKU fordelt på kjønn. Absolutte tall og prosent

| Kjønn   | Gruppe A (kun AKU) |         | Gruppe B (AKU/reg) |         | I alt  |         |
|---------|--------------------|---------|--------------------|---------|--------|---------|
|         | Antall             | Prosent | Antall             | Prosent | Antall | Prosent |
| Menn    | 146                | 0.84    | 17 274             | 99.16   | 17 420 | 100     |
| Kvinner | 161                | 1.08    | 14 682             | 98.92   | 14 843 | 100     |

## Næringsgrupper

Av de som ifølge AKU er sysselsatt i «jordbruk, skogbruk fiske og fangst» er det 4 prosent (68 / 4 341), som vi ikke finner igjen i registrene. Mer detaljerte studier av næringen viser at av disse utgjør sysselsatte innenfor jordbruksyrker 94 prosent (64 personer), hvorav 45 prosent av disse er oppgitt som familiearbeidere og 45 prosent som selvstendige.

Sysselsatte i industrien fanges derimot i større grad opp av registrene. Prosentmessig utgjør gruppe A her ikke mer enn 0,27 prosent. Når det gjelder varehandelen er andelen i gruppe A særlig stor blant de selvstendig næringsdrivende. AKU sysselsatte i denne næringsgruppen utgjør totalt 348 / 23 075 personer. Av disse er 7 prosent ikke i registrene (25 / 1 663).

Tabell 8.6 Personer sysselsatt ifølge AKU fordelt på næring. Absolutte tall og prosent

| Næring                  | Gruppe A (kun AKU) |         | Gruppe B (AKU/reg) |         | I alt  |         |
|-------------------------|--------------------|---------|--------------------|---------|--------|---------|
|                         | Antall             | Prosent | Antall             | Prosent | Antall | Prosent |
| Uoppgitt                | 5                  | 6       | 73                 | 94      | 78     | 100     |
| Jordbruk                | 68                 | 4       | 1 682              | 96      | 1 750  | 100     |
| Olje                    | .                  | .       | 409                | .       | .      | .       |
| Industri                | 13                 | 0       | 4 880              | 100     | 4 893  | 100     |
| Kraft                   | .                  | .       | 348                | .       | .      | .       |
| Bygg                    | 15                 | 1       | 1 947              | 99      | 1 962  | 100     |
| Vare                    | 55                 | 1       | 5 561              | 99      | 5 616  | 100     |
| Transport               | 19                 | 1       | 2 463              | 99      | 2 482  | 100     |
| Bank                    | 7                  | 0       | 2 478              | 100     | 2 485  | 100     |
| Offentlig tjenesteyting | 125                | 1       | 12 115             | 99      | 12 240 | 100     |

## Yrkesgrupper

Fordi antall AKU-sysselsatte i gruppe A er såpass lite (307), får vi et noe begrenset antall observasjoner på hver enkelt yrkesgruppe. Dette gjør det vanskelig å trekke noen bastante konklusjoner. Tallene bekrefter imidlertid resultatene vi så under næringsgrupperingen, ved at vi også her finner en liten overvekt av jordbruksyrker i gruppe A. I tillegg skiller gruppen for lønnet husarbeid seg ut, ved at 20 prosent av de AKU-sysselsatte ikke er registrert i registrene (20 / 2 003).

## Fylke

Som tabellen 8.7 viser er det for fylke ingen klare tendenser når det gjelder avvik mellom AKU og registrene, når vi tar med alle statuskoder. En videre oppsplitting i ansatte og selvstendige gir imidlertid som resultat at av de selvstendige i fylkene Oslo og Akershus, er det henholdsvis 9 prosent (16 / 1 070) og 7 prosent (17 / 1 162) som ikke fanges opp i registrene. En slik oppsplitting fører også her med seg problemer ved at antallet observasjoner i de forskjellige kategoriene blir få.

Tabell 8.7 Personer sysselsatt ifølge AKU fordelt på fylke. Absolutte tall og prosent

| Fylke            | Gruppe A (kun AKU) |         | Gruppe B (AKU/reg) |         | I alt  |         |
|------------------|--------------------|---------|--------------------|---------|--------|---------|
|                  | Antall             | Prosent | Antall             | Prosent | Antall | Prosent |
| Uoppgitt         | .                  | .       | 10                 | 100     | 10     | 100     |
| Østfold          | 11                 | 1       | 1 732              | 99      | 1 743  | 100     |
| Akershus         | 31                 | 1       | 3 520              | 99      | 3 551  | 100     |
| Oslo             | 37                 | 1       | 3 313              | 99      | 3 350  | 100     |
| Hedemark         | 16                 | 1       | 1 082              | 99      | 1 098  | 100     |
| Oppland          | 27                 | 2       | 1 665              | 98      | 1 692  | 100     |
| Buskerud         | 21                 | 1       | 1 750              | 99      | 1 771  | 100     |
| Vestfold         | 12                 | 1       | 1 485              | 99      | 1 497  | 100     |
| Telemark         | 13                 | 1       | 1 193              | 99      | 1 206  | 100     |
| Aust Agder       | 1                  | 0       | 272                | 100     | 273    | 100     |
| Vest Agder       | 16                 | 1       | 1 079              | 99      | 1 095  | 100     |
| Rogaland         | 20                 | 1       | 3 068              | 99      | 3 088  | 100     |
| Hordaland        | 36                 | 1       | 3 189              | 99      | 3 225  | 100     |
| Sogn og Fjordane | 6                  | 1       | 721                | 99      | 727    | 100     |
| Møre og Romsdal  | 17                 | 1       | 1 766              | 99      | 1 783  | 100     |
| Sør Trøndelag    | 16                 | 1       | 2 061              | 99      | 2 077  | 100     |
| Nord Trøndelag   | 5                  | 1       | 722                | 99      | 727    | 100     |
| Nordland         | 12                 | 1       | 1 670              | 99      | 1 682  | 100     |
| Troms            | 7                  | 1       | 1 057              | 99      | 1 064  | 100     |
| Finmark          | 3                  | 1       | 551                | 99      | 554    | 100     |

## Avtalt arbeidstid

Det er kun de som definerer seg selv som ansatte, som blir spurt om avtalt arbeidstid i AKU. Av disse er det i gruppe A kun 57 prosent som har avtale om fast arbeidstid (66 / 4 295). Tilsvarende tall for gruppe B er 73 prosent. Fordeler vi de personene med avtale om fast arbeidstid på arbeidstid, ser vi en klar tendens til at de personene som havner i gruppe A har kortere arbeidstid enn i gruppe B. Her er imidlertid tallmaterialet meget spinkelt, ved at vi kun har 66 personer i gruppe A som har angitt avtalt arbeidstid i AKU.

Tabell 8.8 Personer sysselsatt ifølge AKU fordelt på om de har avtalt arbeidstid.  
Absolutte tall og prosent

| Avtalt arbeidstid           | Gruppe A (kun AKU) |         | Gruppe B (AKU/reg) |         |
|-----------------------------|--------------------|---------|--------------------|---------|
|                             | Antall             | Prosent | Antall             | Prosent |
| Uoppgitt                    | 2                  | 2       | 62                 | 0       |
| Avtale om fast arb. tid     | 66                 | 57      | 21 226             | 74      |
| Avtale om variabel arb. tid | 12                 | 10      | 4 430              | 16      |
| Ingen avtale                | 32                 | 28      | 2 657              | 9       |
| Tilfeldig arb. forhold      | 4                  | 3       | 129                | 0       |
| I alt                       | 116                | 100     | 28504              | 100     |

### 8.3 En kort oppsummering

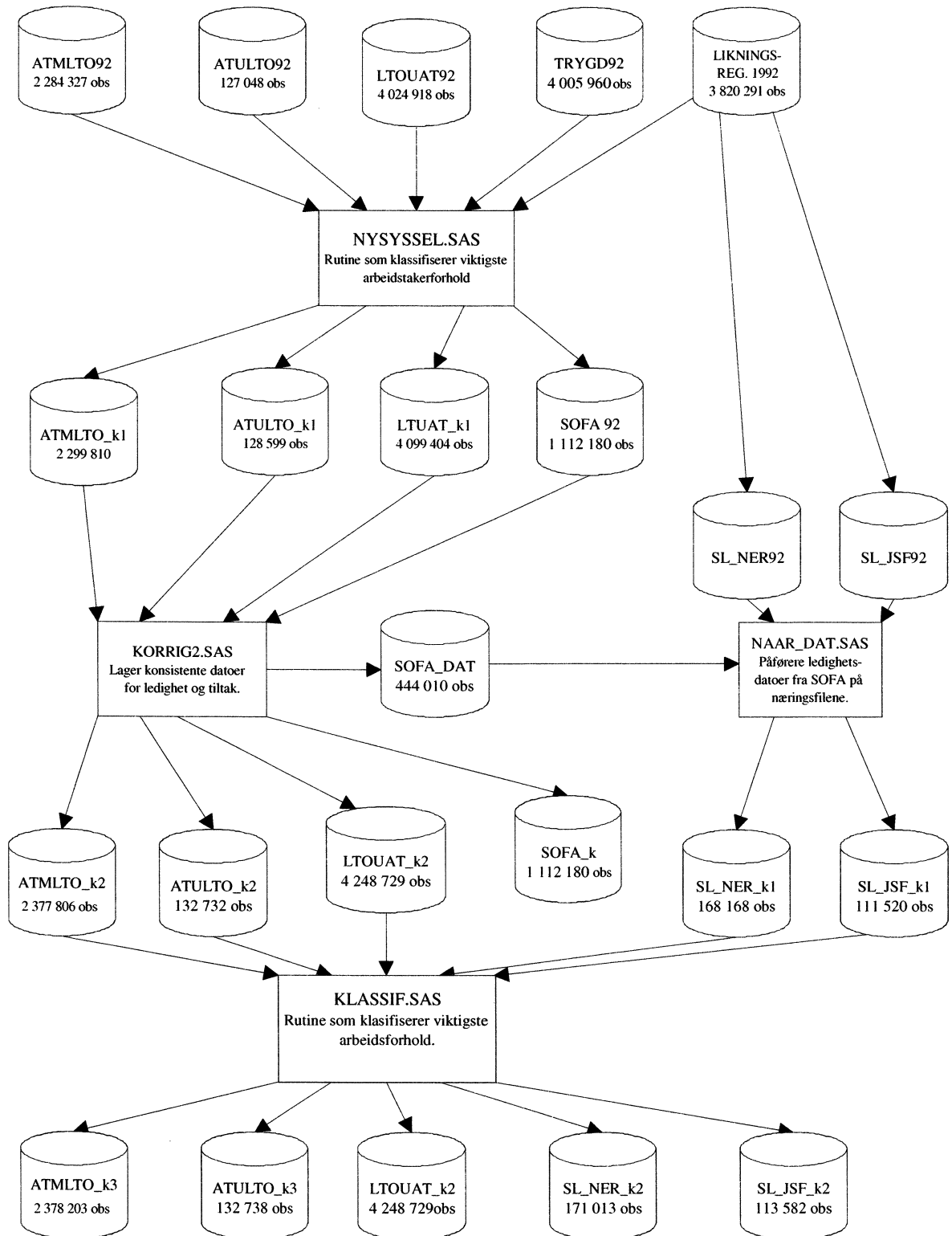
Antall personer i gruppen som har oppgitt i AKU at de er sysselsatt, men som vi ikke finner informasjon om i registrene i løpet av året, er såpass få at vi skal være forsiktige med å trekke bastante konklusjoner. Når vi splitter tallmaterialet opp på endel utvalgte kjennetegn, ser vi allikevel et velkjent mønster fra foregående kvalitetsanalyser av registrene. Eksempler på dette er at gruppene selvstendige og familiarbeidere utgjør en relativt stor andel av de vi kun har AKU- informasjon om. Når det gjelder næring er jordbruk overrepresentert. Samtidig har vi at for personer i alderen under 20 år, er andelen av AKU-sysselsatte som vi ikke finner registerinformasjon om, større sammenlignet med andre aldersgrupper.

Som en totalvurdering anser vi imidlertid at andelen av personer i de ovennevnte gruppene som ikke fanges opp i registrene, som relativt lav. I forhold til den utvalgte klassifikasjonsrutinen betyr dette at vi forløpig har valgt å ikke utarbeide noen spesiell prosedyre for enkeltgrupper. Hvis vi ved gjennomføring av tilsvarende analyser for andre årganger skulle få resultater som antyder at kvaliteten på registerinformasjonen for spesielle delpopulasjoner er dårligere enn det vi her har funnet, vil vi overveie å innføre tillegsrutiner som kan fange opp skjevhetene.

# Vedlegg A

Figur A1. Prosedyre for klassifisering av viktigste arbeidsforhold.

Program og datasett ligger på UNIX-områdene: /ssb/micmac/a9/sfp/prog  
/ssb/micmac/a9/sfp/data



## Referanser

Dale Olsen, H. (1998?), *Klassifisering av status som sysselsatt og ikke-sysselsatt*, upublisert.

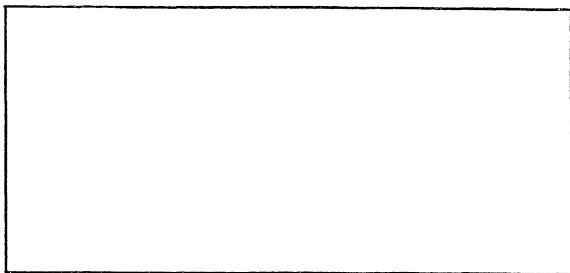
Utne, H. & Vassnes, E. (1995), *Kopling av A/A og LTO-register. Dokumentasjon av kvalitet og konsistens i begrep*. Notater 95:2. Statistisk sentralbyrå, Oslo



## De sist utgitte publikasjonene i serien Notater

- 98/34 A. Sundvoll og H.M. Teigum: Samordnet levekårsundersøkelse 1997 - tverrsnittundersøkelsen: Dokumentasjonsrapport. 130s.
- 98/35 K. J. Einarsen, A. B. Skara og C. Strand: Faktaark for FylkesKOSTRA-utdanning. 1. tertial 1998. Sør-Trøndelag fylkeskommune: Nøkkeltall med indikatorer for Prioriteringer, Dekningsgrad, Produktivitet. 39s.
- 98/36 P. Bakken og J.A. Osnæs: Kvartalsvis ordrestatistikk. 53s.
- 98/39 I. Melby og R. Aaberge: Sammenligning og fordeling av husholdsinntekt blant barn og unge. 31s.
- 98/40 A.A. Ritland: Evaluering av Reform 94. En spørreskjemaundersøkelse: Dokumentasjonsrapport. 43s.
- 98/41 D. Roll-Hansen, L. Solheim og L.C. Zhang: Kopiering ved universiteter og høyskoler. 88s.
- 98/42 M.V. Dysterud og P. Schønning: Etterprøvbare miljømål for byer og tettsteder: Et metodeprosjekt for utvikling og prøving av miljøindikatorer. 40s.
- 98/43 J. Epland: Inntekt etter skatt: Revisjon av inntektsregnskapet i inntekts- og formuesundersøkelsen for husholdninger. 40s.
- 98/44 E. Sørensen: Produksjonsindeks for industrien. 48s.
- 98/45 L. Aaram og Ø. Skullerud: Statistikk over emballasjeavfall: Utprøving av metode og foreløpige resultater. 32s.
- 98/46 L-C. Zhang: Empirisk imputering: En ny metode for å behandle tilfeldig partielt frafall. 20s.
- 98/47 L. Dalen, P.M. Bergh, J-A. Sigstad Lie og A. Vedø: Energibruk i næringsbygg 1995-1997. 69s.
- 98/48 B. Strand og H. Utne: FoB2000: Rapport fra seminar 12. mars 1998 om arbeidsmarkedsdelen i Folke- og boligtellingsen 2000. 33s.
- 98/49 N.Ø. Mæhle og K. Nyborg: Energibruk og utslipp til luft i norsk produksjon: Direkte og indirekte virkninger. 23s.
- 98/50 T. Eidem og J. Lajord: FD-Trygd. Dokumentasjonsrapport: Utdanning 1992-1993. 87s.
- 98/51 A. Bjerkestrand og S. Fjeld: Regnskapsstatistikk for aksjeselskaper 1996: Dokumentasjon. 34s.
- 98/52 G. Haakonsen, S. Holtskog og B. Tornsjø: Energibruk og utslipp til luft i Oslo, Drammen, Bergen og Trondheim 1995. 57s.
- 98/53 E. Holmøy: Hvordan generelle likevekts-effekter bidrar til prisfølsomheten i den norske el-etterspørselen: Dokumentasjon av beregningsrutiner. 33s.
- 98/54 F.R. Aune, T. Bye, M.I. Hansen og T.A. Johnsen: Kraftpris og skyggepris på CO<sub>2</sub> - utslipp i Norge til 2027. 13s.
- 98/55 S. Blom: Holdning til innvandrere og innvandringspolitikk: Spørsmål i SSBs omnibus i mai/juni 1998. 34s.
- 98/56 K. Bjønnes og B.R. Joneid: FD - Trygd. Dokumentasjonsrapport: Foreløpig uførestønad, 1992-1993. 34s.
- 98/57 T. Bye: Fleksibel gjennomføring av en klimaavtale. 27s.
- 98/58 K.J. Einarsen (red.): Arbeidsutvalgets evaluering av faktaark for FylkesKOSTRA-utdanning: 1. tertial 1998. Sør-Trøndelag fylkeskommune. 33s.
- 98/59 I. Øyangen: Inntekts- og formueundersøkelsen 1997: Dokumentasjonsrapport. 23s.
- 98/60 B. Olsen og I. Tuveng: Utvalgsundersøkelsen om sykefravær, 1-3 dager for 3. kvartal 1997: Dokumentasjon. 19s.
- 98/62 A.G. Hustoft: Forslag til ny regional inndeling: Etablering av publiseringsnivå mellom fylke og kommune. 61s.

## Notater



Tillatelse nr.  
159 000/502

**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