



Ole Villund

**Partielt frafall av yrkesdata i
Arbeidstakerregisteret**

Notater

Innhold

1	Innledning.....	2
2	Bakgrunn for yrkesklassifisering.....	2
3	Frafall av yrkesdata.....	2
3.1	Manglende innlevering.....	2
3.2	Uspesifikke titler.....	3
3.3	Andre problematiske titler.....	3
3.4	Yrkeskoder levert av arbeidsgiver.....	3
4	Fob 2001 og sysselsettingsstatistikk.....	3
4.1	Prioritering.....	4
4.2	Noen problemområder i forhold til modellering.....	4
4.3	Konklusjon for sysselsettingsstatistikken.....	5
5	Justering for frafall av yrke i Arbeidstakerregisteret.....	5
5.1	Data.....	5
5.2	Frafallsmodeller.....	7
5.3	Imputering.....	11
6	Planer.....	16
	De sist utgitte publikasjonene i serien Notater.....	17

1 Innledning

Dette notatet har bakgrunn i notatet "Yrke i Arbeidstakerregisteret" (2003, Ole Villund ved Seksjon for arbeidsmarked), og en innleveringsoppgave i forbindelse med kurs i statistiske metoder SM05-2003. Det presenteres metoder for å justere for partielt frafall av yrke i Arbeidstakerregisteret, og anbefalinger for produksjon av statistikk. Mye av metodikken bygger på kursnotat av Li Chun Zhang 2003. Målgruppen er brukere av yrkesvariabelen ved Seksjon for arbeidsmarked, og andre interesserte.

I hovedsak diskuteres yrke i Arbeidstakerregisteret, men det er også et avsnitt om de nye registerbaserte sysselsettingsstatistikken. Den baserer seg på flere andre registre enn Arbeidstakerregisteret og ble første gang publisert i forbindelse med Folke- og boligtellingsen 2001 (FoB).

FoB inneholdt ikke yrkesstatistikk som følge av frafall og usikker kvalitet av yrke i Arbeidstakerregisteret på det tidspunktet. Videre er det slik at sysselsatte fra andre registre mangler yrkesdata. Dette gjelder f.eks. kortvarige jobber som finnes i Lønns- og trekkoppgaveregisteret (LTO), men ikke i Arbeidstakerregisteret.

2 Bakgrunn for yrkesklassifisering

Arbeidstakerregisteret administreres av Rikstrygdeverket (RTV) og inneholder over 2 millioner records. Enheten er arbeidstakerforhold, som er identifisert av person + bedrift + ansatt dato. Dette innebærer at det for hver person kan være flere records, og også dubletter på person + bedrift, med ulike ansettelsesdatoer. Et register over yrkesdata vil inneholde dubletter på arbeidstakerforhold (definert utfra person + bedrift + ansatt dato) fordi en person som endrer arbeidsoppgaver kan ha flere yrkeskoder i samme ansettelsesforhold.

Fra og med 2001 er det krav om at innmeldinger til Arbeidstakerregisteret skal inneholde yrkesdata for alle arbeidstakerforhold. Bedrifter med elektronisk innlevering skal levere yrkeskode, andre kan føre inn yrkestittel som tekst eller skrive yrkeskode. Arbeidsgivere kan finne riktig yrkeskode i Yrkeskatalogen som papirversjon eller på Internet, og få hjelp ved henvendelse til Statistisk sentralbyrå. De som benytter stillingskoder som brukes i offentlig sektor kan oppgi disse istedenfor yrkeskode. Statistisk sentralbyrå mottar ukentlig data fra Arbeidstakerregisteret og koder yrke i henhold til Standard for yrkesklassifisering. Yrkeskodene skal benyttes til statistikkformål og tilbakemelding til arbeidsgiver for kontroll. Yrkeskodingen baseres både direkte på det arbeidsgiver oppgir og ved bruk av tilleggsvariabler.

3 Frafall av yrkesdata

Et lavt og helt tilfeldig frafall i Arbeidstakerregisteret har liten statistisk betydning unntatt for svært små grupper. Analyser av Arbeidstakerregisteret og sammenlikninger med AKU gir grunn til å tro at frafallet ikke er tilfeldig og at det i tillegg er andre klassifiseringsproblemer. Med frafall i denne sammenheng menes alle ubrukbare inndata, det vil si leverte data hvor arbeidstakerforhold ikke kan yrkesklassifiseres.

Arbeidstakerregisteret er et forvaltningsmessig verktøy, og det er et krav fra RTV at alle arbeidsgivere skal levere yrkesdata på alle arbeidstakere. Frafall er derfor et administrativt problem i den grad RTV benytter yrkesdata på personnivå. Generelt sett kan frafall i yrkesdata være en indikatorer for kvaliteten av data fra Arbeidstakerregisteret. I den grad frafallet skyldes at visse grupper av arbeidsgivere synes det er vanskelig å finne riktig yrkeskode, vil det kunne bety at også de leverte yrkeskode kan være usikker hos liknende arbeidsgivere. Vi har for tiden ikke noe mål på kvaliteten av de yrkeskoder som er levert av arbeidsgiver uten omkoding ved Statistisk sentralbyrå.

3.1 Manglende innlevering

Det mangler fortsatt endel yrkesdata, pr. uke 13-2003, er det 12% av arbeidstakerforholdene som mangler en brukbar yrkeskode. Det meste er fra bedrifter som ikke har levert noe yrkesdata i det hele tatt, færre bedrifter har levert yrkesdata for kun deler av arbeidsstokken. Det er grunn til å tro at det er ikke-ignorerbare skjevheter i frafallet. Man har gode opplysninger om personer og bedrifter fra Arbeidstakerregisteret og Bedriftregisteret, som kan brukes for å justere for dette. Det er flere årsaker til mangler:

- Bedrifter som ikke leverer meldinger.
- Meldinger/årskontroller som ikke blir punchet.
- Maskinelle rutiner som ikke fungerer, IT-baserte lønssystemer og andre tekniske problemer.

Årsakene til manglende levering kan være:

- Utilstrekkelig kommunikasjon mellom RTV og arbeidsgivere. Dette handler ikke nødvendigvis om manglende informasjon, men hvordan informasjonen blir nyttiggjort.
- Reelle problemer med å finne riktig yrkeskode. Dette skyldes både egenskaper i bedriftene og egenskaper ved yrkesstandarden, yrkeskatalogen, søkemuligheter.

- Noen bedrifter kan oppleve dette som en økning av oppgavebyrden og prioriterer det ned. Det er også forståelig med bedrifter som føler at de har fått tilbake svært upassende yrkeskoder, og at det er unødig merarbeid ved årskontrollen pga. feilklassifiseringer fra vår side.

Årsakene til manglende registrering ved trygdekontor kan være interne forhold som vi ikke spekulerer om. Det kan være verdt å merke seg at det er 1 trygdekontor i hvert fylke som har ansvaret for registreringen av alle meldinger fra arbeidsgivere i fylket. Trygdekontorspesifikke forhold kan derfor påvirke den geografiske fordelingen av frafall. Det samme kan frafall hos dominerende bedrifter på små steder.

3.2 Uspesifikke titler

Bruken av tekst som ikke beskriver arbeidsoppgaver – f.eks. *vikar*, *assistent* og *konsulent* – skaper problemer for yrkeskodning, særlig i næringer med uensartede arbeidsoppgaver. Næringsstandarden er blitt noe endret i 2003, blant annet med mer detaljering i enkelte tertiærnæringer. Fortsatt er mange arbeidstakerforhold slik at de ikke kan yrkesklassifiseres utfra en generell tekst + næringskode.

Arbeidstakerforhold som blir levert med for uspesifikk tekst fikk ved årskontrollen i Arbeidstakerregisteret koden "0000100" med teksten "Angi yrkestittel mer fullstendig". Dette utfra Arbeidstakerregisterets forvaltningsmessige karakter, der alle arbeidstakerforhold skal være oppført med en regulær kode. For rene statistikkformål kunne noen av disse arbeidstakerforhold vært klassifisert på et mer aggregert nivå.

3.3 Andre problematiske titler

Bruken av nye ord og engelske versjoner er økende. Dette kan til dels avhjelpest ved å lage nye koder, dels å hjelpe arbeidsgivere med oversettelser. Når arbeidsgiver leverer slik tekst uten å kontakte oss, kan man ofte få en negativ reaksjon på koden/teksten som sendes tilbake også i de tilfeller der yrkeskoden på 4 siffer-nivå faglig sett er korrekt.

Bruken av titler på uventede steder er langt mer problematisk, f.eks. at man kaller en som selger bensin for *stasjonsbetjent*, en tittel som normalt brukes innen om helt andre arbeidsoppgaver i jernbanenæringen. Forutsetningene for automatisk tekstbasert yrkeskodning er at teksten forteller noe om reelle arbeidsoppgaver. Når teksten ikke lenger er adekvat for dette formål, må andre metoder benyttes. En konsekvens er at stadig færre titler kan kodes entydig utfra tekst, en metode som i utgangspunktet er sikrest, raskest og billigst.

Misvisende tekstbruk er vanskelig å oppdage, og kan gi betydelige systematiske feil. Mye er blitt rettet opp med manuell kontroll, og man kan i ettertid si at det kunne ha vært en like god ressursbruk å kode disse manuelt med en gang. Kodeprogrammene er derfor justert slik at en viss mengde flere arbeidstakerforhold kodes manuelt.

Et annet problem er titler som tildels brukes om liknende arbeidsoppgaver, men som skal ha yrkeskoder på ulike yrkesfelt. F.eks. skal *salgskonsulent* og *selger* skal til yrkesfelt 3 og 5, mens enkelte andre selgere skal til 4. Det er grunn til å spørre om arbeidsgiver faktisk bruker yrkestitlene på den måte som er forutsatt. Det er ikke alltid innlysende utfra yrkestittelen hva som er de reelle arbeidsoppgavene, i forhold til de arbeidsoppgaver som er nevnt i Standard for yrkesklassifisering.

3.4 Yrkeskoder levert av arbeidsgiver

Stadig flere arbeidsgivere leverer yrkeskoder istedenfor tekst. Denne koden kan være basert på yrkeskode fra Statistisk sentralbyrå utfra tidligere levert tekst, eller en yrkeskode arbeidsgiver har kommet fram til. Det er ifølge RTV stadig flere arbeidsgivere som leverer data til Arbeidstakerregisteret på diskett. Pr. mai 2003 er det omlag 2500 foretak med noe over 600.000 arbeidstakerforhold som leverer diskett, og som derfor skulle levert yrkeskode. Når arbeidsgiver leverer yrkeskode er den maskinelle behandling enklere, men det foreligger ingen direkte holdepunkter for arbeidstakerens reelle arbeidsoppgaver. Det må derfor kontrolleres mot avledede variabler, men her gjenstår mye arbeid. Dette må basere seg i stor grad på sammenlikning med tekstbasert yrkesklassifisering i Arbeidstakerregisteret og AKU. For tiden kan vi derfor ikke si noe nærmere om kvaliteten på leverte koder.

4 Fob 2001 og sysselsettingsstatistikk

I den nye registerbaserte sysselsettingsstatistikken beskrives sysselsatte som ikke finnes i Arbeidstakerregisteret. De største gruppene er selvstendig næringsdrivende, og mange sysselsettingsforhold som defineres utfra fra lønns- og trekkoppgaveregisteret (LTO). Mange kjennemerker hentes eller klassifiseres utfra kobling mot andre registre, men yrkesdata mangler fullstendig for disse gruppene. Pr 1. oktober 2001, manglet det yrkesdata på rundt 1 million sysselsatte. Dette skyldes både mangler i data som sendes via databaserte lønssystemer og fra arbeidsgiverne direkte, samt de med stillingskoder som ennå ikke var yrkeskodet. I tillegg som nevnt de som er klassifisert som sysselsatte utenom Arbeidstakerregisteret. Det ble derfor ikke levert yrkesvariabel til FoB 2001.

For sysselsatte utenom Arbeidstakerregisteret kan man tenke seg å nytte kobling til andre registerdata som f.eks. stillingskoder i offentlige register, personens utdanning og foretakets næring. I tillegg kan også foretakets størrelse ved antall sysselsatte og personens lønn eller inntekt være en type informasjon som kan utnyttes til å finne et sannsynlig yrke.

4.1 Prioritering

I AKU publiseres yrkesfelt (1 siffer) etter fylke, kjønn og alder. Videre publiseres på 4 siffer for yrker over en viss størrelse, etter kjønn. All publisert yrkesstatistikk er estimerte årsgjennomsnitt. Ujusterte kvartalsnivåer viser store svingninger, og endringstall pr. kvartal er ikke vurdert publisert. Det kan tenkes at svingningene bl.a. har sammenheng med trekkemetoden som er klyngetrekking av familier. Man vet at det finnes intrafamiliære korrelasjoner mellom yrkesrelaterte kjennemerker og i yrke selv. Et eksempel på det siste er "legefamilier" og liknende, hvor yrke går i arv eller parforhold som oppstår i studietiden.

For publisering basert på Arbeidstakerregisteret og eventuelt andre registre er det aktuelt å publisere detaljerte yrkesdata (4- og i noen tilfeller 7-siffer nivå), og på kommunenivå f.eks. yrkesfelt (1-siffer). Siden detaljerte yrkesdata prioriteres, vil dette gi andre vurderinger av metodikken enn om formålet var å gi presise totaler eller endringstall.

Som følge av de nevnte prioriteringer anbefaler vi at det forsøkes flere ulike metoder ved justering og imputering. Ulemper ved dette er omfattende arbeid med utredning og programmering, samt at en ikke kan gi et samlet mål på kvaliteten. Alternativet er å lage en enkelt multivariabel modell.

4.2 Noen problemområder i forhold til modellering

I definisjonen av et yrke grupperes liknende arbeidsoppgaver. En modell for et yrke som lages utfra faglige og skjønsmessig valg må benytte svært detaljerte inndelinger i f.eks. utdanning og næring for å kunne anta at arbeidsoppgavene tilsvarer et bestemt yrke. I praksis kan derfor noen av gruppene bli for små og usikre. Det er da naturlig å tenke seg at hvis man aggregerer på en av de yrkeskarakteriserende variablene, så kan man aggregere tilsvarende på yrke også. Dette er ikke uproblematisk, som framgår av et par eksempler.

Eksempel 4-1: utdanning kan aggregeres fra 6 til 3 siffer, og yrke fra 4 til 2 siffer

utdanning	tekst	yrke
457118	Maler	7141
457119	Murer	7121
457129	Tømrer	7125

Eksempel 4-2: aggregering av utdanning gir forskjellige yrkesfelt

utdanning	tekst	yrke
469901	Ambulanse	5139
469902	Audiograf	3221
469907	Renhold	9132

Det samme gjør seg gjeldende for endel andre utdanninger, kanskje særlig innen tekniske yrker. På grunn av yrkesstandardens oppbygning, er dette heller ikke enkelt å omgå dette problemet ved å gruppere ikke-hierarkisk slik det er mulig med f.eks. fag i NUS 2000. For eksempel kunne man gruppere tekniske utdanninger finnes ved å velge utdanningskodens 2.-3.siffer = 55. For å finne tekniske yrker må en gå på tvers av samtlige siffer i yrkeskoden, f.eks. 21xx, 31xx, 72xx, 745x.

Variabler som ikke karakteriserer yrke direkte er også tenkt brukt i imputeringsopplegget. Det er da interessant å se på aggregering av yrke i forhold til prediktive, men ikke-klassifiserende variabler, som f.eks. kjønnsfordeling på ulike nivåer i enkelte yrkesgrupper. Kjønn er bestemmende for yrke (4-siffer), men i liten grad på yrkesfelt (1-siffer).

Tabell 4-3: Utdrag fra yrkesfelt og yrke etter kjønn. AKU 2002. Tusentall og prosent.

Yrkeskoder (2 ulike nivåer)	I alt		Menn		Kvinner	
2 AKADEMISKE YRKER	255	100 %	147	58 %	108	42 %
2130 Systemutviklere og programmerere	37	100 %	31	84 %	6	16 %
2230 Spesialsykepleiere og jordmødre	17	100 %	1	6 %	16	94 %
3 HØYSKOLEYRKER	523	100 %	249	48 %	275	53 %
3111 Bygningsingeniører og teknikere	16	100 %	14	88 %	2	13 %
3231 Sykepleiere	50	100 %	4	8 %	46	92 %

4.3 Konklusjon for sysselsettingsstatistikken

I AKU er det publisert tall for yrkesfelt (1 siffer) etter fylke, og yrke (4 siffer) for de største yrkene. Det er ikke noe mål å komme fram til andre yrkesandeler for alle sysselsatte på de allerede publiserte nivåene. Hensikten med å lage registerbasert yrkesstatistikk er å gi tall på alle yrker (4 siffer), dvs. også for mange små yrker som ikke kan publiserer fra AKU. Videre er det interesse for kommunetall.

Da det ikke er hensikten å lage et nytt estimat for de aggregerte tallene tar vi ikke utgangspunkt i en felles, multivariat modell. Vi satser på å lage en sammensatt metodikk og forsøker å favorisere kvaliteten på små grupper. Yrke predikeres og imputeres ved hjelp av ulike metoder avhengig av kilden til sysselsettingsklassifiseringen, opphavsregistrenes kvalitet, egenskaper ved bedrifter, m.m. Kvaliteten kan eventuelt dokumenteres separat for hver gruppe. Dette er i tråd med dokumentasjonen av andre variabler i sysselsettingsstatistikken.

5 Justering for frafall av yrke i Arbeidstakerregisteret

Resten av notatet diskuterer konkrete metoder for å justere for frafall, estimere yrkesnivåer, og kvantifisere usikkerheten til disse nivåer. Videre gis resultater av anvendelse av disse metodene på aktuelle data som kun har opphav i Arbeidstakerregisteret og ikke andre kilder til sysselsettingsstatus. Anbefalinger som gis gjelder derfor yrkesestimater for *arbeidstakere*. Det er verdt å merke seg at slike estimater ikke bare er interessant for statistikkpublikasjoner fra Arbeidstakerregisteret og sysselsettingsstatistikk, men også for mange oppdragsgivere med ønsker om spesielle analyser.

Interessevariablene som presenteres i dette notatet er antallet og andelene av de 10 største yrkene. 'Yrke' er definert ved 4-siffer nivå i henhold til Standard for yrkesklassifisering 1999 (NOS C521).

5.1 Data

Inndata er en periodefil fra Arbeidstakerregisteret, som blir koblet med yrkesdata fra det såkalte yrkesregisteret. Yrkesregisteret er samlede og prioritert yrkesklassifisering pr. arbeidstakerforhold i ukefiler fra Arbeidstakerregisteret. Både innholdet i yrkesregister og selve koblingen med resten av Arbeidstakerregisteret byr på endel problemstillinger som kan ha noe betydning for resultatene, men som ikke omtales i så detaljert her som i notatet "Yrke i Arbeidstakerregisteret". Tabellen nedenfor viser fordelingen av de ulike koblingstypene som er gjort. Årsaken til at det ikke benyttes en entydig kobling er bl.a. feil i dateringer, bytte av organisasjonsnummer, endring i næringskoder. Vi velger allikevel å bruke en gradvis kobling, da det er svært få arbeidstakere som har flere arbeidstakerforhold med helt ulik yrkeskode, blant de som har koblingstyper av lavere kvalitet. For personer som har samme yrket i ulike jobber, spiller ikke koblingstypen noen rolle.

5.1.1 Arbeidstakerregister

Datagrunnlaget som brukes består av 2 070 489 records fra Arbeidstakerregisteret periodefil pr. uke 35, 2003 som er koblet mot alle tilgjengelige gyldige yrkesdata. Aktive hoved- og biarbeidsforhold er definert ved bosatte (regst16=1) og gyldige (og typarb16=1,2). Samlet frafall av yrke, av alle ulike årsaker er 8.7%.

5.1.2 Sammenlikning med andre yrkesdata

For en ekstra sammenlikning er det laget en egen tabell fra AKU over yrke for *ansatte* (definert ved sstat=111,121) pr.2.kvartal 2003,. Vanlig publisert yrkesstatistikk i AKU inkluderer selvstendig næringsdrivende. Da disse ikke forekommer i Arbeidstakerregisteret, holdes de utenfor i denne sammenlikningen. Grunnen til at AKU-ansatte fortsatt er høyere enn Arbeidstakerregisternivået er at det er endel sysselsettingsforhold som fanges opp i AKU, men som verken er selvstendige eller innmeldes til Arbeidstakerregisteret, f.eks. kortvarige jobber. Det betyr at forskjellene i yrkesandeler *ikke er jevnt fordelt etter yrke*, da småjobber vil være vanligere i visse yrker enn andre. En ytterligere sammenlikning kunne derfor gjøres ved å selekttere etter samme type arbeidstid i begge datamasser.

Nedenfor vises tabeller over datagrunnlaget som brukes i de videre forsøkene.

Tabell 5-1: Fordeling av koblingstyper mellom Arbeidstakerregisteret og yrkesdata.

	Prosent				Antall arbeidstakerforhold				
	Kun person	+Næring	+Bedrift	Full	I alt	Kun person	+Næring	+Bedrift	Full
I alt	12	1	17	69	2070489	258431	21367	351964	1438727
IKKE OPPGITT	100	0	0	0	161560	161560	0	0	0
0 Andre yrker	3	6	0	90	13178	429	848	17	11884
1 Lederyrker	3	1	20	77	127098	3429	1083	25306	97280
2 Akademiske yrker	4	2	14	81	182263	7317	2915	24773	147258
3 Høyskoleyrker	4	1	19	76	408374	16128	5660	75746	310840
4 Kontor- og kundeservice	4	1	16	79	178969	7789	1488	28244	141448
5 Salgs- og serviceyrker	7	1	31	61	495428	35404	5016	152635	302373
6 Bønder, fiskere o.l.	5	1	5	89	23605	1292	201	1107	21005
7 Håndverkere o.l.	3	1	8	89	174530	5148	1500	13122	154760
8 Operatører, sjåførere m	4	1	7	89	169916	6798	1342	11139	150637
9 Andre yrker	10	1	15	75	135566	13137	1314	19874	101241
A Jordbruk og skogbruk	11	1	4	84	25815	2821	179	1030	21785
B Fiske	11	1	5	84	5367	567	51	250	4499
C Olje, gass, bergverk	38	2	2	58	31692	12069	699	593	18331
D Industri	7	0	5	87	260781	19295	919	13636	226931
E Kraft- og vannforsyning	11	0	65	24	15059	1586	14	9847	3612
F Bygg og anlegg	10	1	3	86	126873	12899	1160	3630	109184
G Handel,butikk,rep.	8	1	12	79	316726	25777	2865	37426	250658
H Hotell,restaurant,bar	9	1	26	65	72209	6175	831	18556	46647
I Transport/kommunikasjon	18	1	11	70	150867	27014	1809	16402	105642
J Bank,finans,forsikring	15	1	6	78	44376	6668	464	2580	34664
K Eiendom, forretn.tjen.	9	1	21	69	208965	19382	1919	43035	144629
L Offentlig forvaltning	13	2	1	84	147328	19114	3592	1137	123485
M Undervisning	8	1	1	90	186601	14260	2390	1131	168820
N Helse- og sosialtj.	21	1	47	31	398942	82600	3665	189351	123326
O Andre sos. og pers.tj.	9	1	17	73	75475	6832	757	13014	54872
P Lønnet husarbeid	42	2	10	46	2419	1028	50	239	1102
Q Internasj. org.	28	0	6	66	147	41	0	9	97
Z Uoppgitt	36	0	12	52	847	303	3	98	443

Tabell 5-2: Koblingstyper mellom Arbeidstakerregisteret og yrkesdata, etter dublett-type.

	I alt	Kun person	+Næring	+Bedrift	Full
I alt	2070489	258431	21367	351964	1438727
Antall records pr. person					
1	1822087	218035	16459	309500	1278093
2	233084	37809	4478	39748	151049
3+	15318	2587	430	2716	9585
Person har samme yrke	1927200	252801	20008	325242	1329149
Har flere ulike yrker	143289	5630	1359	26722	109578

5.1.3 Oppsummering om datagrunnlaget med hensyn til kobling av data

Det er visse problemer med å koble yrke til arbeidstakerforhold på en entydig måte, med de årsakene som er nevnt i forrige avsnitt som bytte av bedriftsnummer, datering, osv. Det er allikevel slik at det er ganske få arbeidstakerforhold hvor en person har flere ulike yrker, og en kobling av dårlig kvalitet. Hvis det var slik at mange arbeidstakere hadde flere arbeidstakerforhold hver seg med helt ulike yrker, kunne man vurdere å bruke en strengere kobling og derved betrakte flere arbeidstakerforhold som uten yrkesdata (frafall). Omvendt kan man også betrakte en "løs" kobling som en form for imputering av yrke pr. person.

Tabell 5-3: Oversikt over frafall av yrkesdata i Arbeidstakerregisteret, etter arbeidssted og næring.

	I alt	Ant. Frafall	Ant. Svar	Frafall %	Svar %
I alt	2070489	161560	1908929	8	92
01 ØSTFOLD	100633	17405	83228	17	83
02 AKERSHUS	202340	23861	178479	12	88
03 OSLO	362893	28519	334374	8	92
04 HEDMARK	71909	4733	67176	7	93
05 OPPLAND	72932	2468	70464	3	97
06 BUSKERUD	101684	4737	96947	5	95
07 VESTFOLD	87560	3705	83855	4	96
08 TELEMARK	65975	2295	63680	3	97
09 AUST-AGDER	38195	1623	36572	4	96
10 VEST-AGDER	65663	2743	62920	4	96
11 ROGALAND	178450	15188	163262	9	91
12 HORDALAND	199015	11787	187228	6	94
14 SOGN OG FJORDANE	45247	2616	42631	6	94
15 MØRE OG ROMSDAL	99756	5794	93962	6	94
16 SØR-TRØNDELAGE	124359	4913	119446	4	96
17 NORD-TRØNDELAGE	48569	4642	43927	10	90
18 NORDLAND	94604	7149	87455	8	92
19 TROMS	66349	9989	56360	15	85
20 FINNMARK	29063	2573	26490	9	91
Annen kommune	14923	4765	10158	32	68
Uoppgitt kommune	370	55	315	15	85
A Jordbruk og skogbruk	25815	1621	24194	6	94
B Fiske	5367	386	4981	7	93
C Olje, gass, bergverk	31692	10775	20917	34	66
D Industri	260781	14314	246467	5	95
E Kraft- og vannforsyning	15059	1031	14028	7	93
F Bygg og anlegg	126873	9330	117543	7	93
G Handel, butikk, rep.	316726	15283	301443	5	95
H Hotell, restaurant, bar	72209	3099	69110	4	96
I Transport/kommunikasjon	150867	21121	129746	14	86
J Bank, finans, forsikring	44376	5684	38692	13	87
K Eiendom, forretn. tjen.	208965	11078	197887	5	95
L Offentlig forvaltning	147328	10943	136385	7	93
M Undervisning	186601	7796	178805	4	96
N Helse- og sosialtj.	398942	44850	354092	11	89
O Andre sos. og pers. tj.	75475	3645	71830	5	95
P Lønnet husarbeid	2419	434	1985	18	82
Q Internasj. org.	147	30	117	20	80
Z Uoppgitt næring	847	140	707	17	83

Ser at frafallet ikke er jevnt i forhold til næring, og kan ha sin årsak i enkelte store foretak. Vi har konkret kjennskap til at enkelte foretak har hatt problemer med å levere yrkesdata. Vi går derfor utfra at mye av frafallet vil bli rettet opp når innrapporteringen fra disse kommer i gang.

5.2 Frafallsmodeller

Frafall er avhengig av en rekke forhold som nevnt i avsnitt 3, og siden det er arbeidsgivers ansvar å levere yrkesdata for alle ansatte, er det naturlig å tenke seg at det er egenskaper ved bedriftene som påvirker frafallet. Videre er det jo slik at hver bedrift vil jo ha en mer eller mindre ensartet arbeidsstokk, slik det kan også være nyttig å bruke egenskaper ved arbeidstakerne selv for å justere for frafall av yrke. Slike egenskaper kan være kjønn, alder og utdanning. Effekten av dette vil avhenge av hvor stor sammenheng det er mellom den enkeltes bakgrunn og nåværende arbeidsoppgaver som ligger til grunn for yrkeskoden. Vi vet at det er betydelig variasjon i dette.

5.2.1 Modell HTF

For å ha et grunnlag for sammenlikning av realistiske modeller lages en estimering ved direkte vektning basert på helt tilfeldig frafall. Denne basismodellen tar for gitt at frafallet ikke avhenger av noen kjente størrelser, noe som jo virker lite sannsynlig på bakgrunn av f.eks. tabell 5.1.2.

Definerer:

U	populasjonen = alle identifiserte enheter i Arbeidstakerregisteret
N	antall i U
s	enheter i Arbeidstakerregisteret med brukbar yrkeskode
n	antall i s
m	antall frafall, i dette tilfelle også: $m = N - n$
Y	interessevariabel: arbeidstakerforhold med yrkeskode/kjønn
$r_i = \begin{cases} 1 & i \in s \\ 0 & i \notin s \end{cases}$	responsindikator.
p_i	svarsannsynlighet = andel med brukbar yrkeskode i register
w_i	designvekt, her: $1/p_i$
y_i	interessevariabel - indikatorer

Estimerer:

$$\hat{Y} = \sum_{i \in s_r} w_i y_i \quad \text{Totalt antall i yrket er summen av vekten til hver person ganget med indikator (0 eller 1).}$$

$$\hat{Y} = \frac{\hat{Y}}{N} \quad \text{Andelen i yrket er det estimerte antallet med yrket delt på populasjonstotalen (antall i Arbeidstakerregisteret).}$$

Vi kan bruke den enkle formelen fordi her er N er like stor som estimert populasjonsstørrelse: $\hat{N} = \sum_{i \in s_r} w_i$

5.2.2 Ikke-informativ SHG-modell

En svarhomogen gruppe-modell tar for gitt at det er mulig å identifisere grupper i hvilke frafallet er like stort. I en ikke-informativ SHG-modell forutsetter man i tillegg at frafallet er avhengig bare av kjente egenskaper, ikke av interessevariabelen selv. Hvor vellykket dette er vil avhenge av i hvilken grad forutsetningen er riktig, og av hvor stor den faktiske variasjonen av frafallet er.

Estimerer svarsannsynligheten pr. enhet i gruppen:

$$\hat{p}_{i,g} = \frac{n_g}{n_g + m_g} \quad \text{hvor } n \text{ og } m \text{ er definert som ovenfor og gruppeindeks er } g \in \{1, 2, \dots, G\}$$

Siden jeg har definert populasjonen = det identifiserte register, setter jeg designvekten lik frafallsvekten:

$$w_{i,g} = 100\% \cdot \frac{1}{\hat{p}_{i,g}}$$

Estimat for populasjonstotalen:

$$\hat{Y} = \sum_{g=1}^G \sum_{i \in s_r} w_{i,g} y_{i,g} \quad \text{og estimat for andelen i yrket: } \hat{Y} = \frac{\hat{Y}}{N}$$

5.2.2.1 Valg av gruppering

Valg av gruppering må balanseres i forhold til ulike interesser:

- Grupperingen bør ha grunnlag i en intuitiv determinasjon eller ha en teoretisk forankring. Dette både for å kunne velge passende variabler og for å gi en koherent forklaring av resultatene.
- Gruppene skal være mest mulig homogene, men samtidig ikke for små.
- Tilleggsvariablene som brukes for å identifisere gruppen må være mest mulig komplette og uavhengige.

Det må nødvendigvis bli endel prøving og feiling, og den valgte grupperingen i disse forsøkene er neppe den ideelle. Vi må forsøke å dokumentere resultatene av forsøket og måle kvaliteten for estimatene, vurdere utslagene for statistikk, avvente kommentarer, og eventuelt revidere inndelingen.

Definerer bedriftskategori som $næringsshovedområde + \text{round}(\log_{10}(\text{ansatte}))$. Dette utgjør en bokstavkode for næring og et tall for størrelsesgruppe som tilsvarer 10-potenser. Antall ansatte pr. bedrift har en frekvensfordeling som er svært skjev, i våre data blir det med denne metoden omtrent 4 størrelsesgrupper i hver hovednæring. Dette for å få et håndterlig antall grupper med variabler som er tilgjengelige og som det er gode grunner til å anta vil beskrive ulikheter ved bedriftene både når det gjelder arbeidsoppgaver og administrative forhold. Neste tabell viser frafallet fordelt etter denne gruppering, og vi legger merke til skjevheten i frafall.

Tabell 5-4: Oversikt over frafall av yrkesdata i Arbeidstakerregisteret, etter bedriftsgruppering.

	I alt	Frafall	Svar	F.%	S.%												
I alt	2070489	161560	1908929	8	92	F 3	11607	1673	9934	14	86	L 4	16135	24	16111	0	100
A 0	7847	468	7379	6	94	G 0	30844	1630	29214	5	95	M 0	2108	132	1976	6	94
A 1	8253	541	7712	7	93	G 1	213064	9894	203170	5	95	M 1	45574	2074	43500	5	95
A 2	8853	602	8251	7	93	G 2	69785	3545	66240	5	95	M 2	120585	5336	115249	4	96
A 3	862	10	852	1	99	G 3	3033	214	2819	7	93	M 3	12911	224	12687	2	98
B 0	866	86	780	10	90	H 0	4234	262	3972	6	94	M 4	5423	30	5393	1	99
B 1	3903	296	3607	8	92	H 1	41420	1727	39693	4	96	N 0	11991	645	11346	5	95
B 2	598	4	594	1	99	H 2	25272	1103	24169	4	96	N 1	109534	9318	100216	9	91
C 0	405	28	377	7	93	H 3	1283	7	1276	1	99	N 2	184173	17962	166211	10	90
C 1	3067	142	2925	5	95	I 0	10493	555	9938	5	95	N 3	67685	12265	55420	18	82
C 2	9025	1161	7864	13	87	I 1	45006	2973	42033	7	93	N 4	25559	4660	20899	18	82
C 3	15865	6583	9282	41	59	I 2	62604	6821	55783	11	89	O 0	11248	571	10677	5	95
C 4	3330	2861	469	86	14	I 3	32764	10772	21992	33	67	O 1	39944	2004	37940	5	95
D 0	8021	551	7470	7	93	J 0	1294	65	1229	5	95	O 2	21206	1037	20169	5	95
D 1	66828	2330	64498	3	97	J 1	14146	762	13384	5	95	O 3	3077	33	3044	1	99
D 2	126861	6311	120550	5	95	J 2	16629	1417	15212	9	91	P 0	2237	432	1805	19	81
D 3	59071	5122	53949	9	91	J 3	12307	3440	8867	28	72	P 1	182	2	180	1	99
E 0	428	32	396	7	93	K 0	26227	1256	24971	5	95	Q 0	18	1	17	6	94
E 1	6353	469	5884	7	93	K 1	80285	2466	77819	3	97	Q 1	129	29	100	22	78
E 2	7494	526	6968	7	93	K 2	81310	5449	75861	7	93	Z 0	452	99	353	22	78
E 3	784	4	780	1	99	K 3	21143	1907	19236	9	91	Z 1	363	32	331	9	91
F 0	11372	670	10702	6	94	L 0	2352	188	2164	8	92	Z 2	32	9	23	28	72
F 1	63896	2767	61129	4	96	L 1	37408	3453	33955	9	91						
F 2	39998	4220	35778	11	89	L 2	68454	6691	61763	10	90						
						L 3	22979	587	22392	3	97						

Dette er altså en mer detaljert inndeling enn tabell 5-3. Ser at f.eks. gruppe C3 og C4 har stort frafall. Dette er store bedrifter i olje- og gassnæring, og her gjør som nevnt enkeltforetak store utslag. I3 er store bedrifter i transport og kommunikasjon, og her også vi har vi kjennskap til at enkelte store foretak som har hatt problemer med å levere yrkesdata.

Tabell 5-5: Yrke i antall og andeler, resultat av justering ved HTF-modell og ikke-informativ SHG-modell.

Yrke	AKU-ansatte	AKU-andel	HTF-antall	HTF-andel	II-Bedrift-antall	II-Bedrift-andel
5221 BUTIKKMEDARB.	156611.1	7.49	174181.3	8.41%	169253.2	8.17%
5132 HJELPEPLEIERE, O.L.	76584.4	3.66	75196.6	3.63%	77723.6	3.75%
5131 BARNE/UNGDOMSARB.	67099.4	3.21	58148.3	2.81%	57817.6	2.79%
3310 GRUNNSKOLELÆRERE	66924.5	3.20	70340.7	3.40%	67972.1	3.28%
9132 RENGJØRING (BEDR.)	57434.4	2.75	73685.7	3.56%	73052.9	3.53%
3231 SYKEPLEIERE	52471.1	2.51	42369.1	2.05%	45173.4	2.18%
5139 ANNET PLEIE O.L.	51032.1	2.44	59954.2	2.90%	62158.6	3.00%
3415 SALGSREPRESENTANTER	42883.8	2.05	28956.5	1.40%	28389.6	1.37%
4114 KONTORMEDARBEIDERE	35144.3	1.68	61007.4	2.95%	60695.3	2.93%
4113 SEKRETÆRER	34401.6	1.65	28555.2	1.38%	28213.3	1.36%

Det framgår av tabellen en viss forskjell mellom modellene, men også at forskjellen mellom estimatene og AKU er enda større. Dette forklares dels med jobber som inngår i AKU-ansatte og ikke i Arbeidstakerregisteret – og dels med ulikheter i selve yrkeskodingen.

5.2.2.2 Kvalitet

For å si noe om kvalitetsforskjeller tar jeg utgangspunkt i variansen til estimatet av gjennomsnittet (andelene) av interessevariabelen. Estimert gjennomsnitt er vanligvis:

$$\hat{Y} = \frac{\sum_{i \in s_r} a_i y_i}{\sum_{i \in s_r} a_i}$$

Der a som nevnt er trekk sannsynligheten. Siden register = populasjon er $a = 1$ og andelen estimeres enklere:

$$\hat{Y} = \frac{\sum_{i \in s_r} y_i}{n}$$

Variansen av estimatene av andelene er for HTF-modellen, forutsatt konstant varians:

$$v_0 = \frac{\sum_{i \in s_r} a_i^2}{(\sum_{i \in s_r} a_i)^2} \cdot \text{var}(y_i)$$

som med konstant a kan forenkles til:

$$v_0 = \frac{\text{var}(y)}{n}$$

For de andre modeller er variansen avhengig av designvekten på følgende måte:

$$v_1 = \frac{\sum_{i \in s_r} w_i^2}{(\sum_{i \in s_r} w_i)^2} \cdot \text{var}(y_i) = \frac{\sigma_w^2 + w_i^2}{n \cdot w_i^2} \cdot \text{var}(y_i) = \text{var}(y_i) \cdot (1 + c_w)$$

hvor variansen av vektene $\sigma^2 = \frac{\sum_{i \in s_r} w_i - \bar{w}}{n}$ der $\bar{w} = \frac{\sum_{i \in s_r} w_i}{n}$ og variasjonskoeffisienten $c_w = \frac{\sigma_w}{\bar{w}}$.

Kan definere en rate for å måle effekten av å bruke andre modeller enn rent tilfeldig frafall:

$$\gamma = \frac{v_1}{v_0} = \frac{1 + c_w^2}{1 + c_a^2}$$

Siden $\text{var}(y)$ forekommer både i teller og nevner, "forsvinner" den fra formelen. Dette betyr i praksis at denne raten avhengig av modellen og ikke av interessevariabelen.

Spesialtilfeller:

C_a blir i vårt tilfelle null, fordi det er å regne som en fulltelling og a ikke varierer.

C_w blir ved HTF-modellen null, ettersom alle vektene er like.

Raten gamma for den ikke-informative SHG-modell med bedriftsgruppering blir da kun avhengig av de justerte vektene. Resultatet viser en betydelig variansøkning ved å benytte denne modellen.

\bar{w}	1.098
σ_w	0.257
$c_w = \frac{\sigma_w}{\bar{w}}$	0.234

Denne modellen gir liknende estimater som den forrige, men man får en økning i variansen. Som et forsøk estimeres samme parametere utfra en annen stratifisering. Her benyttes færre strata ved å slå sammen bedriftskategorier med liknende frafall.

Tabell 5-6: Ny stratifisering etter bedrifts næring, størrelse og frafallsandel i gruppen.

Ny gruppe	Antall gamle	Sum arb.taker	Svar %	Ny vekt	Grupper fra tabell 5.2.1
s_0	10	64234	98.6 %	1.014610877	L 4 E 3 H 3 M 4 B 2 O 3 P 1 A 3 M 3 L 3
s_1	5	277701	96.3 %	1.038880243	K 1 D 1 H 1 F 1 H 2
s_2	3	169226	95.5 %	1.046711283	M 2 M 1 C 1
s_3	3	260497	95.3 %	1.049079779	G 1 K 0 O 2
s_4	5	249132	95.0 %	1.052806843	D 2 O 1 J 0 O 0 G 2
s_5	14	244496	93.8 %	1.066480557	G 0 I 0 N 0 J 1 Q 0 F 0 A 0 H 0 M 0 A 1 I 1 K 2 A 2 D 0
s_6	11	209565	91.6 %	1.091922281	C 0 E 2 G 3 E 1 E 0 B 1 L 0 N 1 J 2 D 3 Z 1
s_7	3	242724	90.4 %	1.106298028	K 3 L 1 N 2
s_8	6	192554	89.3 %	1.120138218	L 2 B 0 F 2 I 2 C 2 F 3
s_9	10	160360	74.3 %	1.345189162	N 3 N 4 P 0 Z 0 Q 1 J 3 Z 2 I 3 C 3 C 4
i alt	70	2070489			

5.2.2.3 Vurdering

Den nye stratifiseringen gir ikke særlig andre estimater, noe som ikke er overraskende med tanke på de høye svarandelene totalt. Imidlertid er det betraktelig lavere variasjonskoeffisient som vist i tabellen under. Allikevel vil denne metoden ha en ulempe ved at stratifiseringen ikke er konstant, siden den bygger på punktvis frafallsandel og ikke en mer invariant egenskap ved bedriften. Strengt tatt er jo ikke næringskode og størrelse helt konstant, men dog mer stabile kjennetegn ved en bedrift.

\bar{w}	1.0895
σ_w	0.079258
$c_w = \frac{\sigma_w}{\bar{w}}$	0.0727

De ikke-informative modellene er hovedsakelig vist for å gi et sammenlikningsgrunnlag med metodene i de påfølgende avsnitt. Det er ikke noen spesiell grunn til å anbefale denne metoden slik data foreligger for tiden.

5.3 Imputering

Da det for registerbasert yrkesstatistikk er interessant med tall for detaljerte yrker, og for små grupper, f.eks. kommune, ville det være en stor fordel med komplette mikrodata. Komplettering av data må slik det ser ut i Arbeidstakerregisteret for tiden, skje ved hjelp av imputering. Da yrke regnes som rent partielt frafall (ingen enhetsfrfall) kan man i stor grad basere seg på samme tilgjengelig informasjon for hvert arbeidstakerforhold uten yrkeskode som arbeidstakerforhold med komplette yrkesdata. I tillegg kobles på annen registerinformasjon som f.eks. arbeidstakers utdanning.

5.3.1 Tilleggsdata

5.3.1.1 Utdanning

Kobler på utdanning fra "Befolkningens høyeste utdanning", den siste filen som foreligger er pr. nov. 2001. Observerer ca. 2% arbeidstakere som mangler utdanningskode, og analyserer frafallet nærmere. Her er vist rent frafall, ikke andre typer av ubrukelige koder. Frafall av utdanningskode vil utgjøre et problem i den grad det er korrelert med yrke. Ser at de største manglene er innen de yngste arbeidstakere, og for yrkene innen primærnæringer og yrker uten krav til utdanning, samt for arbeidstakerforhold som mangler yrkeskode.

Tabell 5-7: Partielt frafall av utdanningskode, etter kjønn og aldersgrupper; og etter yrkesfelt

	I alt	Har utd.	Frafall	Utd.%	Frafall %
I Alt	2070489	2038657	31832	98	2
K 10	575	2	573	0	100
K 20	114776	104581	10195	91	9
K 30	226870	224355	2515	99	1
K 40	256495	255496	999	100	0
K 50	239605	239163	442	100	0
K 60	155017	154912	105	100	0
K 70	18472	18463	9	100	0
K 80	837	837	.	100	.
K 90	26	26	.	100	.
M 10	1007	.	1007	.	100
M 20	111562	101450	10112	91	9
M 30	250730	247429	3301	99	1
M 40	271010	269348	1662	99	1
M 50	238144	237475	669	100	0
M 60	164081	163853	228	100	0
M 70	19751	19736	15	100	0
M 80	1469	1469	.	100	.
M 90	62	62	.	100	.
I Alt	2070489	2038657	31832	98	2
Mangler	161562	157253	4309	97	3
0 Andre yrker	13178	13161	17	100	0
1 Lederyrker	127098	126761	337	100	0
2 Akademiske yrker	182263	181213	1050	99	1
3 Høyskoleyrker	408374	406640	1734	100	0
4 Kontor/kundes	178969	177577	1392	99	1
5 Salg/service	495428	485958	9470	98	2
6 Bønder, fiskere	23605	22372	1233	95	5
7 Håndverkere o.l.	174530	173294	1236	99	1
8 Operat/sjåfører	169916	168731	1185	99	1
9 Andre yrker	135566	125697	9869	93	7

De to neste tabellene viser en nærmere oversikt over utdanningsnivået blant arbeidstakere i data som brukes i forsøket. For endel yrker må det benyttes svært detaljerte utdanningskoder for å kunne imputere yrke deterministisk, og samtidig få en viss kontroll med mikrokonsistens.

Tabell 5-8: Utdanningsnivå etter kjønn/aldersgruppering. Prosent av kjønn/aldersgruppering.

	K 10	K 20	K 30	K 40	K 50	K 60	K 70	K 80	K 90	M 20	M 30	M 40	M 50	M 60	M 70	M 80	M 90
0 Ingen utdanning og førskole	.	2	2	1	1	1	1	1	.	2	2	2	1	1	1	1	2
1 Barneskoleutdanning	.	0	0	0	0	0	0	.	.	0	0	0	0	0	0	0	.
2 Ungdomsskoleutdanning	50	10	4	6	13	19	30	40	46	10	5	7	12	17	22	26	10
3 Videregående, grunnutd.	50	41	16	28	39	43	43	40	38	47	17	20	27	28	29	36	47
4 Videregående, avsluttende	.	43	33	25	14	10	7	8	4	38	42	36	25	21	16	16	23
5 Påbygging til videregående	.	0	2	3	3	2	2	1	.	1	4	4	4	3	3	1	2
6 Univ./høgsk.utd. lavere	.	4	37	29	27	22	15	9	12	1	24	21	21	20	16	11	10
7 Univ./høgsk.utd. høyere	.	0	6	5	3	3	3	1	.	0	6	8	9	10	12	8	8
8 Forskerutdanning	.	.	0	0	0	0	0	.	.	.	0	1	1	1	3	1	.

Tabell 5-9: Utdanningsnivå etter yrkesfelt. Prosent av yrkesfelt.

	IKKE OPPGITT	0 Andre yrker	1 Leder-yrker	2 Akademi-ske yrker	3 Høyskol-eyrker	4 Kontor- og kundese-rvick	5 Salgs- og service-yrker	6 Bønder, fiskere o.l.	7 Håndver-ker o.l.	8 Operatø-rer, sjåfør-er, m	9 Andre yrker
0 Ingen utdanning og førskole	4	0	1	2	1	2	3	6	2	2	13
1 Barneskoleutdanning	0	.	0	0	0	0	0	0	0	0	1
2 Ungdomsskoleutdanning	10	1	5	1	3	11	13	17	10	20	23
3 Videregående, grunnutd.	26	8	18	5	13	39	41	37	30	36	37
4 Videregående, avsluttende	26	36	22	10	19	31	33	31	51	35	21
5 Påbygging til videregående	3	6	4	2	5	4	2	1	3	2	1
6 Univ./høgsk.utd. lavere	23	29	37	39	55	12	8	6	3	4	5
7 Univ./høgsk.utd. høyere	7	19	11	35	5	1	0	1	0	1	0
8 Forskerutdanning	0	0	1	5	0	0	0	0	0	0	0

5.3.1.2 Næring

Næringskode for bedrift hentes fra bedriftsregisteret og er klassifisert utfra bedriftens hovedsaklige virksomhet, altså hva bedriften produserer av varer eller tjenester. Bedriftens næringskode kan være en annen en foretakets, og vi går utfra at det er bedriftens næringskode som er mest bestemmende for arbeidstakerens arbeidsoppgaver. Når det gjelder frafall kunne en tenkt seg at det var foretakets egenskaper som var bestemmende, i den grad innsending av yrkesdata skjer foretaksvis. Av praktiske grunner benyttes samme variabel gjennom hele denne analysen, og under vises en nominell kontroll av næringskoder i de aktuelle data. Frafall utgjør et ubetydelig problem i disse data, og vi kjenner ikke til omfanget av feilkoding. Man vet at det er endel bedrifter som har næringskode som ikke samsvarer helt med deres virksomhet, og det foregår en løpende revisjon av registerdata. Bl.a. er det slik at siden næringskoden hentes ved hjelp av organisasjonsnummer, er det viktig at arbeidstakerne er knyttet til riktig enhet, ikke bare at enheten har riktig næringskode.

Tabell 5-10: Frafall av næringskoder, etter bedriftsstørrelse. Antall arbeidstakerforhold.

Størrelsesgruppe	I alt	Kode	Frafall
I alt	2070489	2069699	790
0 (små)	132437	132030	407
1	779355	779004	351
2	842879	842847	32
3	265371	265371	.
4 (store)	50447	50447	.

5.3.1.3 Andre bedriftsegenskaper

Enhetens størrelsen målt ved antall ansatte har betydning for organisering av arbeidet og derfor for arbeidsoppgaver som igjen får konsekvenser for yrkesfordelingen i en bedrift eller foretak. Dette skyldes støttefunksjoner som vokser med økende størrelse, for eksempel yrker innen administrasjon og ledelse, samt lager, transport osv. Man antar også at det er høyere grad av spesialisering innen ledelsen av store bedrifter, f.eks. egne personaldirektører, osv. I små bedrifter vil arbeidsoppgavene til den enkelte være mer varierte, noe vi vet har skapt usikkerhet om hvilken yrkeskode de bør velge. Hovedregelen er at de arbeidsoppgavene som man bruker lengst tid på skal være bestemmende for yrkeskoden.

Foretakenes ansattetall øker og minsker som følge av både organisatoriske omgrupperinger, så vel som reelle ansettelser/oppsigelser. Foretak kan derfor bytte størrelsesgruppe fra en periode til en annen, og dette vil vel skje i større utstrekning enn endringer i næringskoder. Andre egenskaper er organisasjonsform (A/S, osv.), og hvorvidt det er en eller flere bedrifter i foretaket. Dette er egenskaper som først og fremst kan ha betydning for ledelses- og administrasjonsyrker.

5.3.1.4 Andre egenskaper ved arbeidstakerforholdet

Det har vært foreslått å bruke lønn i modellering av yrke. Uten å ha undersøkt dette nærmere kan vi peke på to forhold:

- En betydelig del av arbeidstakerforholdene i Arbeidstakerregisteret er ikke koblet til Lønns- og trekkoppgaveregisteret (LTO) slik denne koblingen foretas i forbindelse med sysselsettingsstatistikken. Å forsøke enda en koblingsmetode her vil bringe inn nye usikkerhetsmomenter. Man må også merke seg at det i LTO ikke er direkte kobling til bedrift (kun foretak), og at det er betydelige svakheter i dateringene.
- Vi tror at lønn har endel å si for yrkesfelt (1-siffer nivå), men kan være mindre nyttig for å skille yrker på et detaljert nivå. Dette er ikke undersøkt nærmere, og det kan tenkes at det for enkelte yrkesgrupper er mer nyttig enn antatt.

Det vil bli mer aktuelt å bruke lønn for å predikere yrke for selvstendig næringsdrivende. For denne gruppen vil man ha et mer entydig datagrunnlag for inntekt, da disse defineres bl.a. utfra selvangivelsesregisteret.

Det vil også bli aktuelt å bruke yrkeskoder som er samlet inn i forbindelse med lønnsstatistikken. Dette er en stor utvalgsundersøkelse hvor bedrifter leverer yrkes- og lønnsdata på personnivå.

5.3.2 Imputeringsmetodikk

Vi kan imputere alle som mangler yrke i Arbeidstakerregisteret ved en imputeringsmodell i flere trinn. Valg av variabler og detaljeringsnivåer må baseres på yrkesfaglig skjønn i tillegg til en ren kvantitativ vurdering av effekt og kvalitet. Det betyr at vi ikke kan gi en uttømmende begrunnelse for alle valgene her.

Beskrivelse av metoden:

- Deterministisk imputering utfra en ikke-informativ SHG-modell.
- Estimert gruppegjennomsnitt (urealistisk verdi) for alle innen hver SHG.
- Imputerer i flere trinn, med de mest detaljerte ("beste") først, og grovere inndelinger i videre trinn.

Ideelt sett skulle man identifisert det vi kan kalle 'ikke-informative yrkeshomogene grupper', dvs. i et stratum har alle samme yrke, og stratumstilhørighet er definert av andre variabler enn yrke. For en slik gruppe vil alle arbeidstakerforhold få realistiske verdier ('1' for medlemmer og '0' for ikke-medlemmer). For de øvrige gruppene vil verdiene ikke gi mening på mikronivå, men gi liknende makrotall som andre metoder.

Noen egenskaper ved den valgte metoden:

- Det er enkelt å lage makrotall utfra den komplette yrkesvariabelen.
- Ingen ekstra estimeringsusikkerhet slik som ved stokastiske metoder.
- Metoden er mulig å forklare på en intuitiv måte, altså at den virker rimelig for brukerne av yrkesdata.
- Verdiene må konverteres for å kunne brukes på mikronivå.
- Programmeringen blir mer omfattende og dokumentasjonen blir mindre sammenhengende enn ved en mer enhetlig metodikk.
- Man underslår endel variasjon i data, og metoden kan være mer sårbar for skjevheter.
- Variansen av estimatet må justeres for effekten av imputering for å gi et realistisk mål.

Estimerer andelen av yrke Y innen gruppen g :

$$\hat{Y}_g = \frac{\sum_{i \in S_r, g} y_{i,g}}{n_g}$$

Imputerer yrke for arbeidstakerforhold i som mangler yrke og som er medlem av gruppen g :

$$Y_{i \notin S_r, i \in g} = \hat{Y}_g$$

Sum for imputert yrke:

$$Y^* = \sum_{i=g}^G \sum_{i \notin S_r} Y_{i,g}$$

Estimat for populasjonsandelen blir da:

$$\hat{Y} = \frac{Y_{S_r} + Y^*}{N}$$

For å kvantifisere effekten av imputering beregnes en justering av varians av estimatet. Dette er altså en økning av den usikkerheten som allerede finnes pga. rent frafall.

La v^* være variansestimater for den imputerte estimator og svarandelen $\bar{r} = \frac{n}{n+m}$

Da er et forenklet frafallsjustert variansestimater gitt ved:

$$v_1 = \frac{v^*}{\bar{r}} \text{ for deterministisk imputering}$$

5.3.3 Variabler og grupperingsnivåer

Tabell 5-11: Valgte inndelinger i forsøk med deterministisk imputering av yrke.

Trinn	Variabel	Aggregering	Antall verdier	Antall grupper
1	Næring	5 (næringsundergruppe)	581	19741
	Utdanning	3 (nivå + faggruppe)	220	
	Antall ansatte	Log-grupper	5	
2	Næring	2 (næring)	45	370
	Utdanning	1 (nivå)	9	
3	Bedriftstype	Næringshovedområde ¹ + størrelsesgruppe	70	70

Tabellen viser forholdet mellom aggregeringsgrad og antall grupper. På sikt kunne man analysert dette nærmere og eventuelt kommet fram til en bedre balanse eller mer raffinerte metoder. Eksempel på en analyse vil være å måle oppsplitting av yrker innen de ulike gruppene (eks. ved mikrokonsistens) for å identifisere inndelinger som er mer effektive.

5.3.4 Resultater

Etter å ha kjørt imputeringsopplegget på de 10 største yrkene, vises resultater av estimering sammenliknet med svargruppen. Detaljerte tabeller brukes for å vurdere om estimeringen fører til spesielle resultater for små grupper. Dette fordi det vil være interessant å lage yrkesstatistikk på et svært detaljert nivå. Vi må derfor kontrollere om den forsøkte metodikken gir store utslag, og i hvilken grad vi kan forklare dette utfra de foreliggende data.

Tabell 5-12: Antall imputert i de ulike trinn, samt tall for øvrige arbeidstakerforhold i samme grupper.

	imputeringstrinn			
	I alt	1	2	3
I alt	2070489	1810379	241389	18721
Frafall, imputert	161560	143340	17023	1197
Levert, ikke-imp.	1908929	1667039	224366	17524

¹ Dette er bokstavkoder, ikke "1-siffer næring"

Tabellen viser at de fleste blir imputert på et detaljert nivå og at det er en rimelig sammenheng mellom størrelsene på giver- og mottakergruppene. Det må bemerkes at bildet vil se annerledes ut for de resterende (342) yrkene, og at en vil forvente mer usikkerhet i små grupper.

Tabell 5-13: Estimerte antall i vanlige yrker, etter kjønn og 10-årige aldersgrupper.

	5221 BUTIKKMEDARB.		5132 HJELPEPL., O.L.		5131 BARN/UNG.ARB.		3310 GRUNNSK.LÆR.		9132 RENGJØR.(BED.)		3231 SYKEPLEIERE		5139 PLEIE O.L.		3415 SALGSREPR.		4114 KONTORMED.		4113 SEKRETÆRER		
	Svar	Estimat	Svar	Estimat	Svar	Estimat	Svar	Estimat	Svar	Estimat	Svar	Estimat	Svar	Estimat	Svar	Estimat	Svar	Estimat	Svar	Estimat	
	Kvinner	10	76	90.9	0	0.18	4	4.36	0	0.36	70	77.85	0	0.22	4	4.34	0	0.48	12	12.5	1
	20	37146	39048.16	4130	4661.48	5471	5900.38	539	587.19	7841	8255.63	763	871.08	7699	8259.11	474	503.96	2134	2286.24	462	502.44
	30	24063	25180.87	10046	11077.62	10872	11666.69	11401	11899.7	8824	9479.83	11212	12451.68	9596	10568.28	2513	2616.97	7978	8345.24	3878	4004.46
	40	17576	18497.48	14470	15733.48	13939	14865.51	10924	11441.24	12229	13070.53	10802	12030.63	10068	11044.34	1926	2050.94	12277	12793.46	6724	6932.41
	50	15254	16023.2	20846	22551.26	12670	13477.7	13095	13754.76	14001	14986.2	8387	9400.26	9333	10362.29	911	984.82	12706	13279.27	6797	7012.95
	60	12033	12717.74	11660	12790.32	5140	5593.78	9396	9889.08	11402	12138.46	4270	4842.43	6642	7428.34	486	535.59	9934	10371.75	5534	5701.24
	70	1856	1983.54	1261	1487.77	355	410.1	533	582.98	1812	1962.08	525	618.66	1142	1300.36	37	42.26	1197	1276.46	528	553.27
	80	86	96.09	34	45.33	0	2.44	3	3.5	108	120.16	15	20.34	44	50.44	2	2.22	73	77.01	15	16.25
	90	3	3.11	0	0.28	0	0.06	0	0	3	3.21	0	0	1	1.37	0	0	1	1.06	1	1.06
Menn	10	85	101.51	0	0.16	1	1.18	0	0.01	21	27.58	0	0.23	1	1.35	3	3.92	10	10.77	2	2.21
	20	21177	22899.48	451	541.76	1175	1324.94	248	290.21	2357	2610.01	26	54.71	1549	1713.48	845	886.88	1033	1161.58	115	153.21
	30	13146	13831.52	1754	1937.8	1847	1976.54	5061	5226.18	3448	3736.92	1240	1472.52	3610	3924.77	5745	5926.47	2547	2789.17	550	638.25
	40	7905	8429.17	2152	2356.68	1182	1285.81	3437	3571.3	3090	3450.25	1006	1217.9	2773	3070.66	6298	6528.27	2228	2509.55	524	635.22
	50	5482	5890.8	1750	1941.24	644	731.21	5065	5263.42	1670	2036.99	635	793.42	1833	2090.42	4242	4424.96	2043	2362.38	590	728.37
	60	3913	4238.69	697	802.44	267	329.12	4810	5053.64	839	1177.91	162	212.24	862	1024.13	2827	2951.13	1665	1942.66	523	645.3
	70	701	770.06	74	92.29	40	55.56	338	363.44	197	270.45	20	26.96	114	138.28	351	367.43	346	397.12	79	101.29
	80	83	94.91	4	6.34	3	5.04	2	4.1	24	31.46	0	0.75	5	7.74	35	36.58	58	65.95	4	6.93
	90	5	8.54	0	0.02	1	1.04	0	0	0	0.28	0	0	0	0.01	2	2.58	5	5.7	0	0.12

Tabell 5-14: Estimerte antall i vanlige yrker, etter fylke.

	5221 BUTIKKMEDARB.		5132 HJELPEPL., O.L.		5131 BARN/UNG.ARB.		3310 GRUNNSK.LÆR.		9132 RENGJØR.(BED.)		3231 SYKEPLEIERE		5139 PLEIE O.L.		3415 SALGSREPR.		4114 KONTORMED.		4113 SEKRETÆRER	
	Svar	Estimat	Svar	Estimat	Svar	Estimat	Svar	Estimat	Svar	Estimat	Svar	Estimat	Svar	Estimat	Svar	Estimat	Svar	Estimat	Svar	Estimat
	01 ØSTFOLD	7589	8819.78	3358	4325.72	2613	2989.54	3090	3569.85	2815	3444.63	1968	2240.45	2893	3442.17	972	1107.88	3138	3512.9	973
02 AKERSHUS	17258	18331.89	4787	6040.32	6065	6787.08	6624	6983.39	5906	6516.87	2382	3379.35	4567	5711.68	4736	4958.34	5478	5964.96	2884	3038.94
03 OSLO	22138	23876.05	5205	6592.71	5298	6241.41	5432	5710.82	9129	10088.16	4290	5954.53	4937	6251.94	9608	9860.87	10089	10852.94	7942	8200.73
04 HEDMARK	5538	6010.45	3742	3858.45	2079	2213.09	2756	2819.52	2856	3125.04	1604	1645.5	2425	2529.71	477	509.92	1746	1867.21	661	700.76
05 OPPLAND	6198	6437.57	4012	4074.92	2088	2137.74	2804	2827.4	3130	3227.43	1687	1712.16	2464	2531.47	481	504.21	1907	1956.17	756	776.5
06 BUSKERUD	8712	9207.96	4271	4415.45	3336	3445.89	2791	3286.22	3424	3639.49	1915	1983.42	5099	5251.58	1254	1304.12	2775	2890.63	1194	1236.18
07 VESTFOLD	8328	8584.66	3432	3517.32	2610	2751.89	2655	3115.53	3055	3206.25	1844	1887.67	2850	2948.85	1009	1028.87	2655	2723	904	934.55
08 TELEMARK	5618	5784.96	3176	3257.34	1961	2020.97	2384	2421.13	2294	2435.53	1502	1527.19	2142	2222.71	433	451.27	2110	2160.29	658	678.77
09 AUST-AGDER	3374	3528.24	1941	1975.13	1122	1149.83	1610	1622.61	1309	1378.24	1138	1152.47	1095	1125.33	288	297.25	1225	1250.51	412	423.84
10 VEST-AGDER	5872	6021.37	2706	2827.51	1780	1828.58	2496	2540.81	2267	2376.17	1669	1710.33	1265	1340.27	509	524.91	1901	1955.09	829	856.68
11 ROGALAND	14740	15275.52	5550	5832.89	4617	4820.61	6068	6211.55	5655	6068.17	3521	3675.37	4260	4570.28	1508	1610.99	4267	4521.55	1862	1990.64
12 HORDALAND	15316	16020.72	7040	7327.66	5141	5346.97	6566	6636.1	6569	6898.82	4426	4583.38	6663	6986	1968	2044.25	5817	6034.4	1964	2055.98
14 SOGN OG FJORD.	3116	3267.81	1991	2096.91	1371	1458.31	1981	2015.12	1940	2018.29	996	1057.85	1556	1642.3	235	247.98	1247	1297.54	294	317.61
15 MØRE OG ROMSLG.	7999	8292.88	4204	4427.18	2768	2885.61	3691	3774.26	3345	3570.99	2459	2604.16	3174	3353.81	731	762.96	2530	2645.05	1365	1414.09
16 SØR-TRØNDELAG	10041	10282.46	4515	4632.15	3111	3222.91	3896	3958.5	4300	4509.75	3058	3133.38	3040	3153.73	1293	1328.14	3593	3683.82	1375	1410.42
17 NORD-TRØNDELAG	4130	4315.89	2045	2335.52	1438	1594.7	2044	2087.16	1745	1965.52	1413	1518.76	1138	1314.53	186	207.76	1235	1336.33	421	467.85
18 NORDLAND	7143	7619.6	4365	4774.85	3107	3257.98	3957	4023.55	3885	4139.07	1679	2120.26	2963	3263.6	392	424.27	2260	2431.08	920	964.87
19 TROMS	4967	5600.56	2073	2633.53	2079	2267.27	2601	2770.05	2404	2747.06	812	1353.13	1926	2419.76	469	528.08	1423	1662.56	676	758.01
20 FINNMARK	2406	2495.73	912	1074.35	990	1169.7	1375	1525.47	1474	1610.65	605	663.42	807	912.49	104	109.29	815	893.11	215	238.92

Generelt kan man si at effekten er størst i de yngste og eldste (små grupper), i Østfold (høyest frafall) og Troms (ganske høyt frafall, og lavt antall arbeidstakerforhold i alt).

5.3.5 Vurdering

Det foreslåtte estimeringsopplegget ved imputering kan brukes for å lage detaljert statistikk over yrke i Arbeidstakerregisteret. De største fordelene er:

- Komplette mikrodata, kan gi tall for små grupper.
- Antall arbeidstakerforhold stemmer overens med annen publisert statistikk fra Arbeidstakerregisteret.
- Usikkerhet som skyldes selve estimeringen er lav.

Siden frafallet er ganske lavt i de foreliggende data, er den teoretisk varians lav og det vil også være en lav tilleggusikkerhet ved imputeringen. Det er grunn til å tro at andre forhold er mer bekymringsfulle når det gjelder kvaliteten på yrkesdata i Arbeidstakerregisteret. Dette er i hovedsak klassifiseringsproblemer, altså validitetsproblematikk ikke estimeringsproblematikk. Samsvarsanalyser med AKU viser endel avvik i yrkeskodningen på personnivå, og stikkprøver av mikrodata i Arbeidstakerregisteret viser noen få prosent som har feil eller usikker yrkeskode. Det blir til slutt et skjønsspørsmål hva man mener er en akseptabel kvalitet.

Tabellen nedenfor viser et kvalitetsmål for estimeringsmetoden som er foreslått. Det betyr at vi kan bestemme en grense for akseptabel kvalitet, og fastsette publiseringsnivå utfra denne. Variasjonskoeffisient til estimatet er justert på den forenklete måten som anvist for imputering i pkt. 5.3.2.

Tabell 5-15: Estimerte antall i vanlige yrker, og kvalitetsmål.

	Andel	Antall	Variasjonskoeffisient
5221 BUTIKKMEDARB.	8.21 %	169 905.77	0.84 %
5132 HJELPEPLEIERE, O.L.	3.67 %	76 026.44	1.93 %
5131 BARNE/UNGDOMSARB.	2.78 %	57 631.46	2.56 %
3310 GRUNNSKOLELÆRERE	3.28 %	67 931.10	2.17 %
9132 RENGJØRING (BEDR.)	3.55 %	73 435.79	2.00 %
3231 SYKEPLEIERE	2.13 %	44 014.05	3.37 %
5139 ANNET PLEIE O.L.	2.95 %	60 989.71	2.42 %
3415 SALGSREPRESENTANTER	1.35 %	27 865.47	5.34 %
4114 KONTORMEDARBEIDERE	2.88 %	59 687.87	2.47 %
4113 SEKRETÆRER	1.33 %	27 636.11	5.39 %

Andre forhold som må forklares i forbindelse med publisering er avvik mellom Arbeidstakerregisteret i forhold til publisert yrkesstatistikk i AKU, jf. pkt 5.1.2. Tabellen nedenfor viser andelene for de største yrkene, når selvstendig næringsdrivende holdes utenfor i AKU-data 2.kvartal 2003.

Tabell 5-16: Andeler av vanlige yrker, AKU-ansatte (upubliserte tall), og AA-estimat.

Yrke	AKU	AA
	Andel	Andel
5221 BUTIKKMEDARB.	7.49	8.21
5132 HJELPEPLEIERE, O.L.	3.66	3.67
5131 BARNE/UNGDOMSARB.	3.21	2.78
3310 GRUNNSKOLELÆRERE	3.20	3.28
9132 RENGJØRING (BEDR.)	2.75	3.55
3231 SYKEPLEIERE	2.51	2.13
5139 ANNET PLEIE O.L.	2.44	2.95
3415 SALGSREPRESENTANTER	2.05	1.35
4114 KONTORMEDARBEIDERE	1.68	2.88
4113 SEKRETÆRER	1.65	1.33

6 Planer

Følgende arbeidsoppgaver vil ha høy prioritet i arbeidet med yrkesvariabelen i Arbeidstakerregisteret:

- Kontakt med arbeidsgivere for å øke leveringsgraden. Her vil vi særlig prioritere foretak med mange ansatte og som har maskinell innrapportering til Arbeidstakerregisteret, da dette vil gi størst effekt på leveringsgraden.
- Forsette revisjon av systemer for manuell- og automatisk yrkeskodning, for å sikre mest mulig faglig korrekte yrkeskoder.
- Bruke yrke aktivt i interne analyser, f.eks. i forhold til sykefravær, for å få nye innfallsvinkler til å bedømme kvaliteten.
- Fortsatt holde et høyt servicenivå ovenfor arbeidsgivere i forbindelse med årskontrollen 2004, med rask hjelp til å finne riktig yrkeskode. Dette utfra et ønske om større grad av levering av koder framfor tekst.

De sist utgitte publikasjonene i serien Notater

- 2003/50 E. Holmøy: Velferdsregnskap - et mulig teoretisk rammeverk. 35s.
- 2003/51 C. Wiecek: Undersøkelse om fremtidsplaner, familie og samliv. Dokumentasjonsrapport. 59s.
- 2003/52 KOSTRA: Arbeidsgrupperapporter 2003. 153s.
- 2003/53 A. Haglund: Rapport fra arbeidsgruppa om forslag til arbeidsdeling mellom Brønnøysundregistrene (BR) og Statistisk sentralbyrå (SSB). 40s.
- 2003/54 E. Eng Eibak: Forventningsindikator - konsumprisene. Mai - november 2003. 19s.
- 2003/55 G. Daugstad: Levekår for ungdom i større byer. 80s.
- 2003/56 A. Vedø og D. Rafat: Sammenligning av utvalgsplaner i AKU. 17s.
- 2003/57 L. Belsby: Frafall og vekter i Tidsbruksundersøkelsen 2000-2001. 20s.
- 2003/58 L. Belsby: Vekter i Forbruksundersøkelsen. 28s.
- 2003/59 M. Mogstad og L.C. Zhang: På veien fra familie- til husholdningsregister. En metode for prediksjon av samboere uten barn .53s
- 2003/60 A. Vedø og D. Rafat: Redigering av husholdningsfilen fra Kvalitetsundersøkelsen. 13s.
- 2003/61 M. Mogstad: Analyse av fattigdom basert på register- og folketellingsdata. 75s.
- 2003/62 T. Eika og J.A. Jørgensen: Makroøkonomiske virkninger av høye strømpriser i 2003. En analyse med den makroøkonometriske modellen KVARTS. 16s
- 2003/63 B. Mathisen: Flyktninger og arbeidsmarkedet 4. kvartal 2001. 32s.
- 2003/64 E. Røed Larsen og D.E. Sommervoll: Til himmls eller utfor stupet? En katalogisering av forklaringer på stigende boligpriser. 31s.
- 2003/65 P.E. Tønjum: Tilbakemelding/ dokumentasjon av prosjektet: Avstemming av KNR mot nye årstall ifølge tallrevisjonen. 43s.
- 2003/66 B.A. Holth: Arbeids- og bedriftsundersøkelsen 2003. Dokumentasjon. 67s.
- 2003/67 H. Tønseth: Kommuneale helseforskjeller -de finnes, men kan de måles? 15s.
- 2003/68 T.M. Normann: Omnibusundersøkelsen mai/juni 2003. Dokumentasjonsrapport. 50s.
- 2003/69 KOSTRA (Kommune- Stat- Rapportering) Rutinebeskrivelse og dokumentasjon. 60s.
- 2003/70 E. Holmøy og B. Strøm: Fordeling av tjenesteproduksjon mellom offentlig og privat sektor i MSG-6. 25s.
- 2003/71 J.K. Dagsvik: Hvordan skal arbeidstilbudseffekter tallfestes? en oversikt over den mikrobaserde arbeidstilbudsforskningen i Statistisk sentralbyrå. 67s.
- 2003/72 A. Steinkellner: Inntektsstatistikk for personer og familier 1999-2001. Dokumentasjon av datagrunnlag og produksjonsprosess. 43s.
- 2003/73 F. Tverå, I. Sagelvmo: Beregning av næringene fiske eget bruk, fiske og fangst og fiskeoppdrett i nasjonalregnskapet. 19s.
- 2003/74 K.H. Grini: Lønnsstatistikk privat sektor 1997-2001. Dokumentasjon av utvalg og beregning av vekter. 36s.
- 2003/75 A.H. Foss: Grafisk revisjon av nøkkeltallene i KOSTRA. 16s.
- 2003/76 K. Hansen: Ideelle organisasjoner i nasjonalregnskapet. 30s.
- 2003/77 E.E. Eibak: Undersøking om foreldrebetaling i barnehagar, august 2003. 46s.
- 2003/78 A.H. Foss: Kvaliteten i husholdningsdelen i Folke- og bolig tellingen 2001. 31s.