

Svein Erik Førre

**Registerdataene i lys av
industristatistikken**

Notater

Innhold

1. Innledning.....	5
2. Omfanget av «99-bedrifter»	5
3. En sammenligning av sysselsettingstall	6
4. En sammenligning av antall bedriftsenheter	8
5. Hvor mange finner en makker ?	10
6. En sammenligning av jobbreallokeringsrater	13
6.1. En korreksjonsprosedyre	13
6.2. En sammenligning med Industristatistikken	15
7. Konklusjon	17
Referanser	18
Vedlegg A	19
Vedlegg B.....	20
De siste utgitte publikasjonene i serien Notater fra Forskningsavdelingen.....	21

1. Innledning¹

Dette notatet er oppstått som et biprodukt av en analyse av bedrifiers etterspørsel etter arbeidskraft med ulik utdanningsbakgrunn, utført av Salvanes og Førre (1997). Utgangspunktet for denne analysen var to av de administrative registrene som finnes i registerdatabasen i SSB: Arbeidsgiver-/arbeidstakerregisteret (A/a-registret) og Utdanningsregisteret. I A/a-registeret er hvert enkelt individs bedriftstilknytning angitt ved to ulike kjennetegn. Ved å aggregere med utgangspunkt i disse bedriftsspesifikke kjennetegnene, har vi kunnet nytte den individbaserte informasjonen til å skaffe oss bedriftsspesifikk informasjon. Vi har dermed kunnet analysere relative endringer i alders- og utdanningssammensetningen innen bedrifter. Det potensielt svakeste punktet i vår analyse har vært kvaliteten på koblingen mellom individ og bedrift over tid. Siden kvaliteten på denne koblingen er sentral for validiteten av våre resultater, var det derfor viktig å undersøke denne nærmere. De erfaringene som ble høstet er i dette notatet forsøkt satt inn i en mer helhetlig ramme. Disse erfaringene vil være nyttige siden kvaliteten på denne koblingen vil være av betydning for ethvert studium av bedriftsadfærd med basis i Registerdataene. Videre vil kvaliteten på denne koblingen være viktig for relevansen av ethvert studium der informasjon fra Registerdataene er blitt koblet med informasjon fra Industristatistikken på bedriftsnivå.

Informasjon om den enkelte arbeidstakers bedriftstilknytning er som nevnt, identifisert ved to ulike bedriftsspesifikke kjennetegn: Arbeidsgivernummeret i kombinasjon med et bedriftsspesifikt undernummer eller ved et bedriftsnummer samsvarende med bedriftsnummeret i Industristatistikken. Mens arbeidsgivernummeret endres ved eierskifte vil bedriftsnummeret være bestemt utfra virksomhetens geografiske beliggenhet. Siden vi i vår analyse av jobbskaping ønsket å skille endringer i eierforhold fra etablering og nedlegging, var vi interessert i å anvende bedriftsnummeret. Det var imidlertid i utgangspunktet ekstra usikkerhet knyttet til kvaliteten på bedriftsnummeret. Mens det arbeidsgiverbaserte identifikasjonsnummeret i en viss forstand var å betrakte som «opprinnelig», var bedriftsnummeret blitt generert i etterhånd ved at en på basis av arbeidsgivernummeret og informasjon om den næringen en arbeidstaker var sysselsatt i, fikk identifiserte den relevante bedriften i Industristatistikken. Ved siden av mulighetene for feil ved en slik ekstra tilordningsprosedyre, lyktes det for enkelte av bedriftene i Registerdataene ikke å finne den relevante bedriften i Industristatistikken. I disse tilfellene ble bedriftsnummeret blitt satt lik 99 00000 + bedriftens 5-siffer næringskode. Disse bedriftene vil heretter bli referert til som «99-bedrifter».

I avsnitt 2 vil omfanget av «99-bedrifter» bli undersøkt nærmere. I avsnitt 3 vil Industristatistikken tjene som en mal for en sammenligning av aggregerte tall for sysselsetting. I avsnitt 4 vil likeledes antall bedriftsenheter bli gjenstand for sammenligning. I avsnitt 5 vil vi undersøke hvor mange av bedriftene i Registerdataene som vi klarer å koble med en tilsvarende bedrift i Industristatistikken. Til sist vil vi i avsnitt 6 sammenligne jobbreallokeringsrater generert på basis av Registerdataene med tilsvarende rater generert på basis av Industristatistikken.

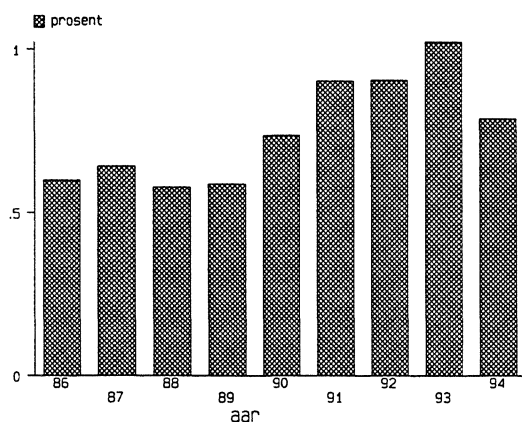
2. Omfanget av «99-bedrifter»

I figur 1 ser vi at «99-bedriftene» utgjør fra i overkant av 0.5 prosent av observasjonene i begynnelsen av perioden til oppunder 1 prosent av bedriftene i slutten av perioden. Figur 2 viser andelen «99-bedrifter» innen hver av 2-siffernæring². Vi ser at næring 32 har den høyeste andelen (2 prosent) mens næring 33 og 34 har de laveste andelene (0.5 prosent).

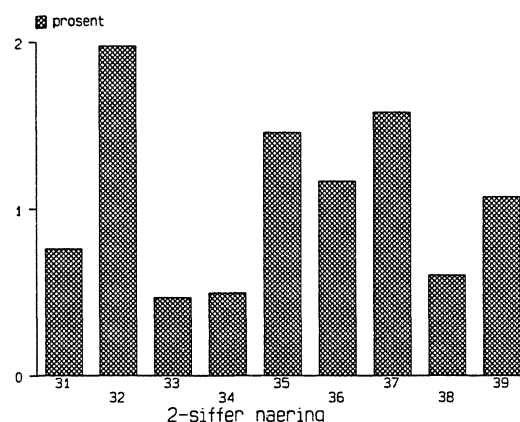
¹ Dette arbeidet er støttet av midler fra Norges Forskningsråd (prosjektnummer: 102791/531).

² Se vedlegg B for en definisjon av de ulike 2-siffer næringer.

Figur 1. Andelen bedrifter per år med bedriftsnummeret satt lik 99+næring



Figur 2. Andelen bedrifter i ulike tosiffrer næringer med bedriftsnummeret satt lik 99+næring



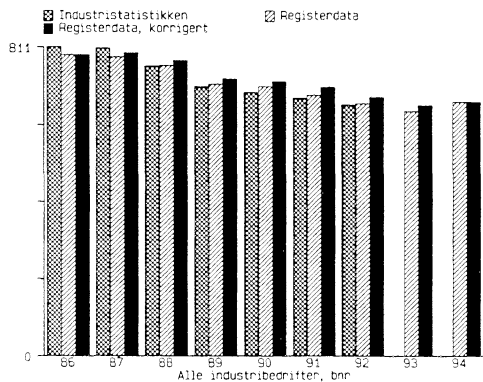
«99-bedriften» er i all hovedsak en liten bedrift: 50 prosent av disse bedriftene har 2 eller færre sysselsatte mens 95 prosent har mindre enn 20 sysselsatte. Totalt utgjør sysselsettingen i disse bedriftene kun 0.2 prosent av den totale sysselsettingen. Problemet med «99-bedrifter» synes derfor å være av begrenset karakter særlig hvis man ser bort fra bedrifter med mindre enn 5 sysselsatte.

3. En sammenligning av sysselsettingstall

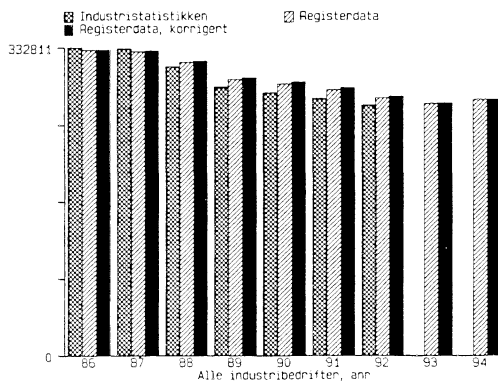
En analyse av nettoendringer i bedriftens arbeidstokk innenfor ulike utdanningsnivåer fordrer en aggregering av den individspesifikke informasjonen. Denne aggregeringen kan foretas på basis av en av to bedriftsspesifikke kjennetegn: (1) arbeidsgivernummeret og et bedriftsspesifikke undernummeret (*anr*) eller (2) et bedriftsnummer som samsvarer med bedriftsnummeret i Industristatistikken (*bnr*). Tilfeldige feil i koblingen mellom individ og bedrift vil etter aggregering både kunne resultere i et feilaktig nivå og i en for stor variasjon i antall sysselsatte i bedriftene over tid. Dersom en aggregerer på basis av «*bnr*», vil «99-bedriftene» resultere i ytterligere variasjon i sysselsetting over tid. Problemet oppstår i de bedriftene som i enkelte år faller inn under gruppen av «99-bedrifter» men i andre år har fått tilordnet et bedriftsnummer. Disse bedriftene fremstår med en «hullete» tidserie som feilaktig vil gi inntrykk av påfølgende nedleggelser og nyetableringer.

For å kunne vurdere omfanget av disse problemene vil Industristatistikken i det følgende tjene som en mal for sammenligning. Figur 3a og 3b viser totaltall for sysselsetting i industrien for hvert år i perioden 1986 til 1994 etter aggregering mhp. henholdsvis «*bnr*» og «*anr*». Hvert år er representert med tre søyler som viser tall basert på Industristatistikken og Registerdataene i en opprinnelig og en korrigert versjon. Den korrigerede søylen er et resultat av en korreksjonsprosedyre beskrevet i avsnitt 6.1 og vil ikke bli lagt vekt på her. Vi ser av figuren at avviket i total sysselsetting er relativt lite mellom Industristatistikken og Registerdataene. I absoluttverdi overskrider avviket aldri mer enn 2.8 prosent dersom dataene er aggregert på basis av «*bnr*» og 3.2 prosent dersom ved bruk av «*anr*». Registerdatene gir, med unntak av de to første årene, høyere sysselsettingstall for hele perioden (for årene 93 og 94 har vi ingen tall fra Industristatistikken).

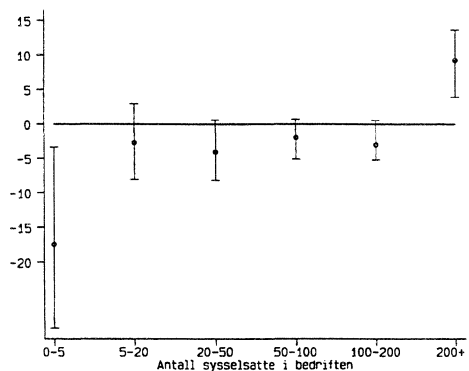
Figur 3a. Sysselsetting i industrien over tid. Industristatistikken og Registerdataene («bnr»)



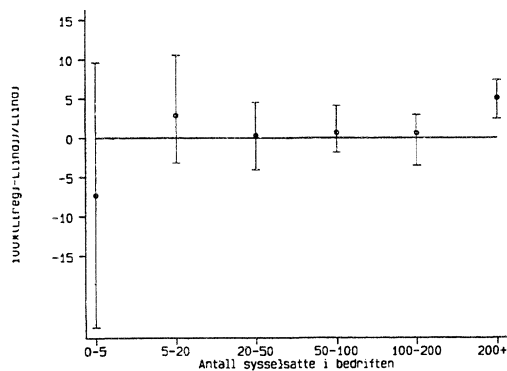
Figur 3b. Sysselsetting i industrien over tid. Industristatistikken og Registerdataene («anr»)



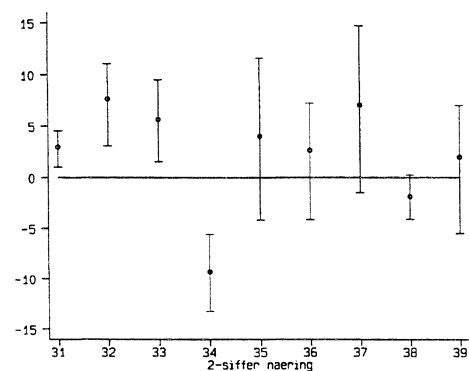
Figur 3c. Avvik i sysselsetting innen ulike størrelseskategorier mellom Registerdataene («bnr») og Industristatistikken



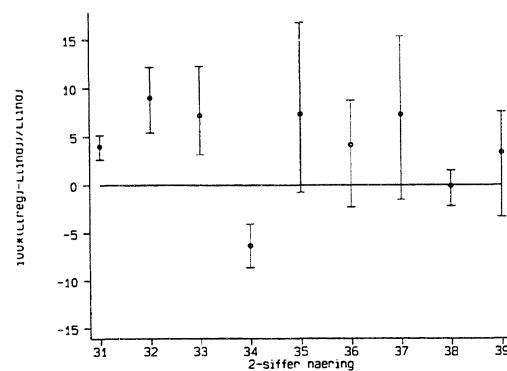
Figur 3d. Avvik i sysselsetting innen ulike størrelseskategorier mellom Registerdataene («anr») og Industristatistikken



Figur 3e. Avvik i sysselsetting på næringsnivå mellom Registerdataene («bnr») og Industristatistikken



Figur 3f. Avvik i sysselsetting på næringsnivå mellom Registerdataene («anr») og Industristatistikken



Figur 3e og 3f viser det prosentvise avviket i sysselsetting på næringsnivå (2-siffer, ISIC) mellom Registerdataene og Industristatistikken. I figurene er det gjennomsnittlige avviket angitt sammen med de maksimale og minimale verdiene over perioden. Vi ser av begge figurene at næring 34 har færre sysselsatte mens næringene 31, 32 og 33 har flere sysselsatte i Registerdataene enn Industristatistikken. Store avvik i sysselsetting på næringsnivå vil være et resultat av forskjeller i klassifisering av bedrifters næringsstilhørighet.

Figur 3c og 3d viser prosentvis avvik i sysselsetting mellom Registerdatene og Industristatistikken innenfor ulike grupper av bedrifter. Bedriftene er tilordnet en av seks størrelsesgrupper på basis av antall sysselsatte i bedriften. Periodegjennomsnittet er i figurene angitt sammen med de maksimale avvikene i hver retning. Forskjellen mellom figur 3c og 3d skyldes at vi identifiserer bedriften med henholdsvis «bnr» og «anr». I Figur 3c der «bnr» danner utgangspunktet for aggregering er «99-bedriftene» utelatt. Dette resulterer i en reduksjon i antall sysselsatte, særlig gjør dette seg gjeldende i gruppen av de minste bedriftene. Dersom vi sammenligner Figur 3c med Figur 3d ser vi at bedriftene med færre enn 5 sysselsatte får tilordnet omlag 8 prosent færre sysselsatte i Registerdataene dersom vi aggregerer ved «anr» og helt ned mot 17 prosent færre dersom vi aggregerer på basis av «bnr», en refleksjon av at særlig mange av «99-bedriftene» befinner seg i denne gruppen. I gruppen av de største bedriftene, med mer enn 200 sysselsatte, er sysselsettingen 5 prosent høyere i Registerdataene.

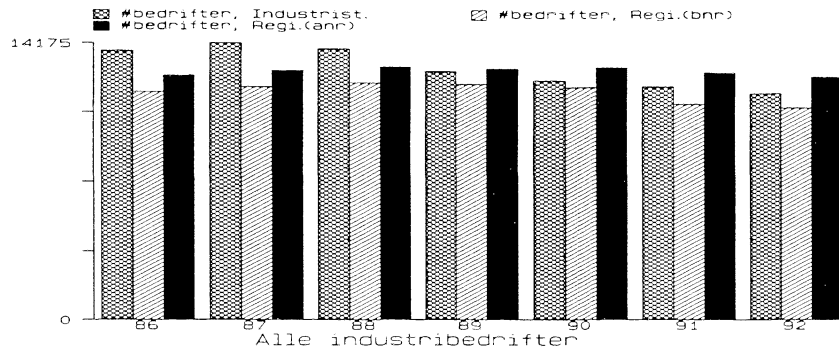
4. En sammenligning av antall bedriftsenheter

Etter å ha sammenlignet tall for sysselsetting skal vi nå undersøke samsvaret i antall bedriftsenheter i de to datasettene. Figur 4a viser det totale antall bedriftsenheter i Industristatistikken og i Registerdataene avhengig av om vi aggregerer på basis av «bnr» eller «anr». Dette største avviket oppstår ved bruk av «bnr» og utelatelse av «99-bedrifter». Aggregering på basis av «bnr» resulterer i et lavere antall bedrifter over hele perioden mens en aggregering på basis av «anr» gir et større grad av samsvar og et høyere antall bedrifter i Registerdataene enn i Industristatistikken i slutten av perioden.

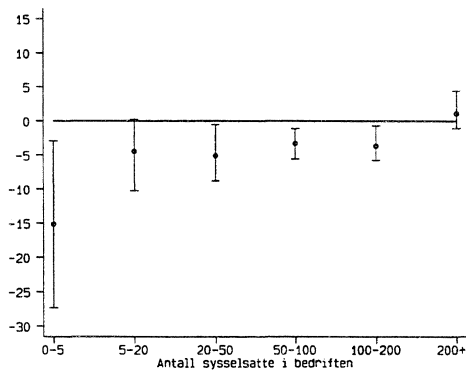
I Figur 4b og 4c er bedriftene igjen gruppert etter størrelse. Figurene viser det prosentvise avviket i antall bedrifter innenfor de ulike størrelseskategoriene. Periodegjennomsnittet er angitt sammen med det maksimale og minimale avviket over perioden. Vi ser at avviket er størst for gruppen av de minste bedriftene. Antallet bedrifter er her mye lavere i Registerdataene. Det er imidlertid kun i gruppen av de største bedriftene at antallet sysselsatte er høyere i Registerdataene.

Figur 4d og 4e viser det prosentvise avviket på næringsnivå (2-siffer, ISIC) mellom Registerdataene og Industristatistikken. Vi ser at retningen på avviket i antall bedrifter varierer sterkt innenfor de ulike næringene. Næring 34 har både et lavere antall bedrifter og som vist i Figur 3e et lavere antall sysselsatte. Dette skyldes etter alt å dømme forskjeller i næringsklassifisering i de to datasettene som har resultert i at flere av de bedriftene som oppgitt produksjon i næring 34 i Industristatistikken ikke er klassifisert i denne næringen i Registerdatene. For mange av de andre næringene er det imidlertid ikke en slik entydig sammenheng mellom avvik i antall sysselsatte og antall bedrifter. Næring 35 og 36 har i Registerdataene eksempelvis et høyere antall sysselsatte, men et lavere antall bedrifter. I tillegg til forskjeller i bedrifters næringsklassifisering kan dette skyldes av at enkelte bedrifter opptrer med høyere antall sysselsatte i Registerdataene enn i Industristatistikken.

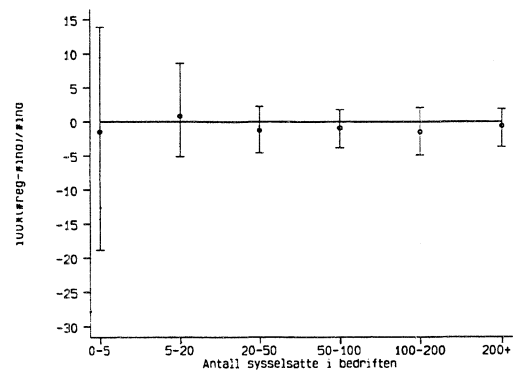
Figur 4a. Antall bedrifter i industrien over tid



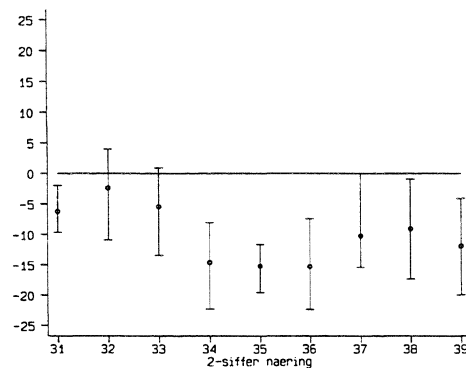
Figur 4b. Prosentvis forskjell i antall bedrifter innen ulike størrelse-kategorier, basert på «bnr»



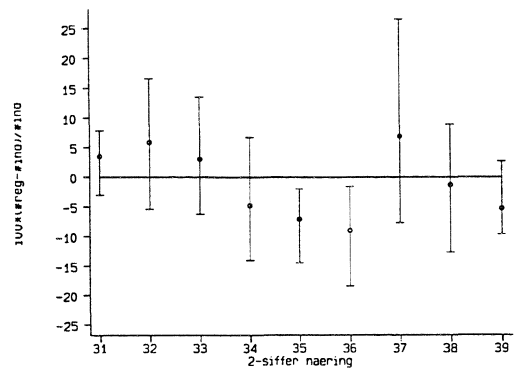
Figur 4c. Prosentvis forskjell i antall bedrifter innen ulike størrelse-kategorier, basert på «anr»



Figur 4d. Prosentvis forskjell i antall bedrifter på næringsnivå (2-siffer, ISIC), basert på «bnr»

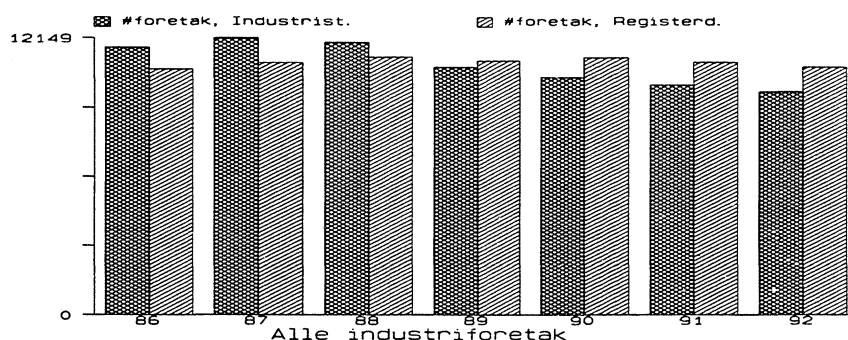


Figur 4e. Prosentvis forskjell i antall bedrifter på næringsnivå (2-siffer, ISIC), basert på «anr»

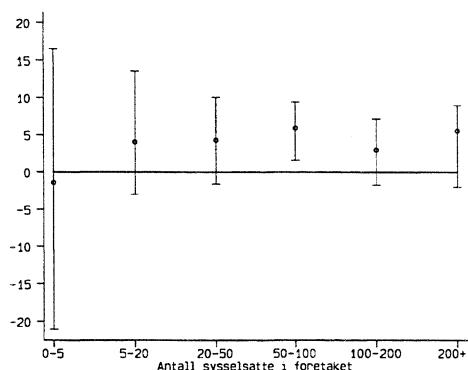


I Figur 5a sammenligner vi det totale antall foretaksheter. Antallet foretak er lavere i Registerdataene først i perioden mens at det motsatte er tilfelle fra 89 og utover. I Figur 5b er det prosentvise avviket i sysselsetting på foretaksnivå angitt for foretak av ulik størrelse. Det er noe flere foretak i Registerdataene innenfor alle størrelseskategorier med unntak av den minste størrelsesgruppen. Figur 5c forskjeller i antall foretak innenfor ulike 2-siffernæringer og er tilnærmet identisk med figur 4e. Den eneste store forskjellen er at næring 35 som hadde et lavere antall bedrifter i Registerdataene enn i Industristatistikken viser seg å ha et høyere antall foretak.

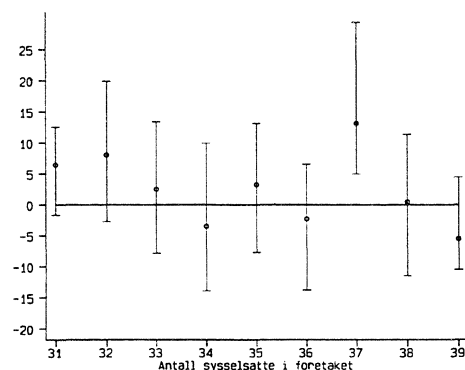
Figur 5a. Antall foretaksheter i industrien over tid



Figur 5b. Prosentvis avvik i sysselsetting innen ulike foretaksstørrelseskategorier



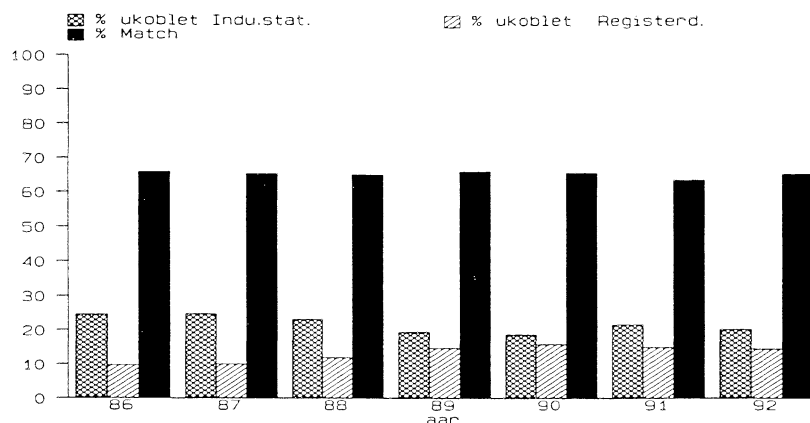
Figur 5c. Prosentvis forskjell i antall foretak på næringsnivå (2-siffer, ISIC), basert på «anr»



5. Hvor mange finner en makker?

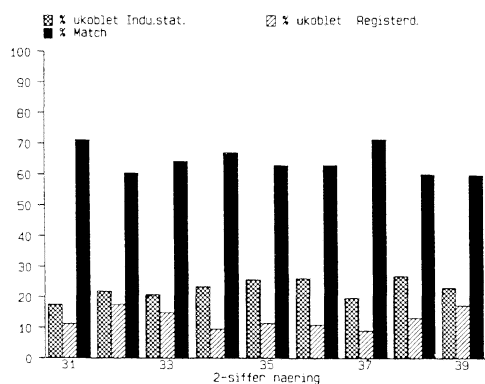
Etter å ha sett på samsvaret mellom antall bedrifter i Industristatistikken og Registerdataene skal vi nå, ved å ta utgangspunkt i «bnr», undersøke hvor mange bedrifter som finner en «makker» i Industristatistikken? Figur 6 viser at andelen som finner makker er relativt stabil over tid, gjennomsnittlig 64 prosent av bedriftene finner en «makker». Andelen bedrifter som blir stående igjen uten «makker» er noe høyere i Industristatistikken, men denne forskjellen avtar over perioden. Samtidig vet vi at antallet bedrifter som er identifisert ved et «bnr» er lavere i Registerdataene enn i Industristatistikken.

Figur 6. Koblingsgrad over tid

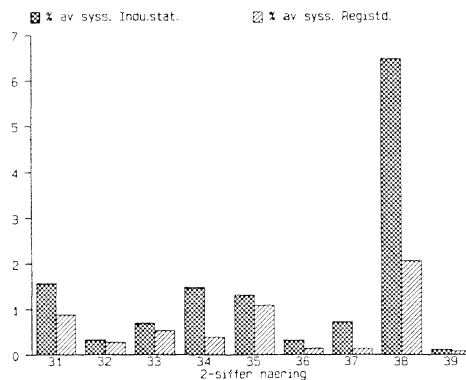


Figur 7a viser at det ikke er store forskjeller i koblingsgrad mellom de ulike næringene. I Figur 7b ser vi på andelen av sysselsettingen til de ikke-koblede bedriftene. Vi ser at næring 38 skiller seg ut. Bedriftene i Industristatistikken som ikke finner noen «makker» sysselsetter nesten 7 prosent av næringens totale arbeidsstokk.

Figur 7a. Kobling etter næring (2-siffer, ISIC)

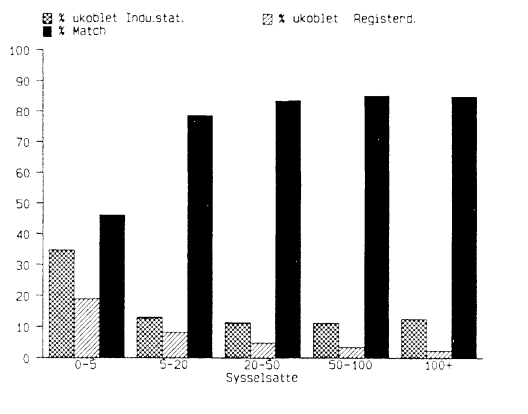


Figur 7b. Sysselsettingsandel til ikke-koblede bedrifter etter næringsnivå (2-siffer, ISIC)

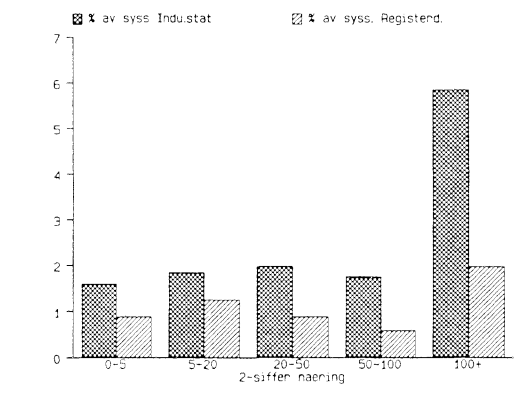


Koblingsgraden er også avhengig av størrelsen på bedriftene. I Figur 8a ser vi at i gruppen av bedrifter med mindre enn 5 ansatte, finner kun i underkant av 50 prosent finne en makker, mens koblingen er oppunder 90 prosent for de største bedriftene. Ved å betrakte Figur 8a og 8b i sammenheng ser vi at de 10 prosent av bedriftene med mer enn 100 sysselsatte i Industristatistikken uten «makker», sysselsetter 6 prosent av arbeidstyrken, mens de minste bedriftene uten «makker» som utgjør over 30 prosent samlet innehar 1-2 prosent arbeidstyrken i Industrien.

Figur 8a. Kobling etter størrelse

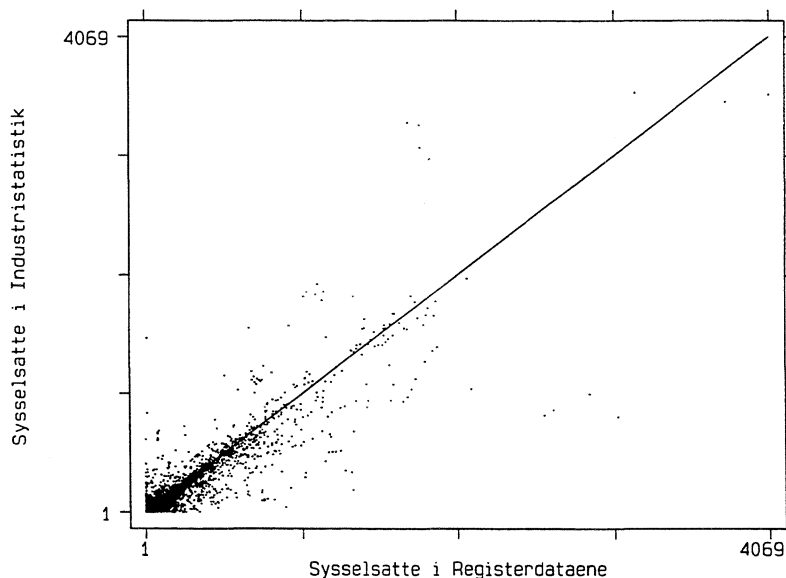


Figur 8b. Ikke-koblede bedrifters sysselsettingsandel etter størrelse



Vi viste i de foregående avsnittene at det var stor variasjon i antall sysselsatte og i antall bedrifter innenfor de ulike næringene. Et spørsmål som her ble reist var i hvilken grad dette dreide seg om ulik næringsklassifisering av bedrifter, eller om de samme bedriftene fremstod med ulik sysselsetting i de to datasettene. Vi kan nå belyse dette spørsmålet litt nærmere ved å ta utgangspunkt i de bedriftene som har klart å finne en kobling i de to datasettene. I Figur 9 under er den enkelte bedrifts sysselsetting i Registerdataene plottet mot den tilsvarende sysselsettingen som følger fra Industristatistikken. Korrelasjonen mellom sysselsettingen i de to datasettene er på 0.92, og av figuren ser vi at sysselsettingen i flere av bedriftene avviker sterkt fra diagonalen.

Figur 9. Samvariasjon i antall sysselsatte på bedriftsnivå mellom Industristatistikken og Registerdataene



6. En sammenligning av jobbreallokeringsrater

Registerdataene dannet som nevnt innledningsvis utgangspunktet for en analyse av jobbskapning på bedriftsnivå. Grunnlaget for denne analysen var årlige jobbkreasjonsrater³ beregnet for ulike utdanningsgrupper i bedriftene. Tidligere har jobbkreasjonsrater blitt beregnet for den *totale* sysselsettingen i bedriftene på basis av Industristatistikken (se f.eks. Klette og Førre, 1995). En viktig kvalitetskontroll for koblingen mellom individ og bedrift i Registerdataene, er derfor å ta utgangspunkt i den totale sysselsettingen i bedriftene og sammenligne jobbkreasjonsrater generert på basis av Registerdataene med tilsvarende ratene generert på basis av Industristatistikken. Dersom Industristatistikken utgjør malen, vil avviket til ratene generert på basis av Registerdataene være et mål på kvalitetsforringelsen ved å nytte Registerdataene, siden feil og mangler ved koblingen mellom individ og bedrift over tid vil resultere i en for høy jobbreallokeringsrate.

Et alternativ til Industristatistikken er å sammenligne de opprinnelige ratene med rater som er basert på en «korrigert» tidsserie. Korreksjonen av den opprinnelige tidsserien kan skje ved en algoritme som forsøksvis korrigerer manglende eller «feilaktig» informasjon utfra visse kriterier. I dette notatet vil den korrigerte tidsserien kun utgjøre et grunnlag for sammenligning.

Vi vil nå redegjøre for en slik korreksjonsprosedyre for deretter å vende oppmerksomheten mot jobbkreasjonsrater generert på basis av Industristatistikken.

6.1. En korreksjonsprosedyre

Etter aggregering av Registerdataene til bedriftsnivå vil tidserien til enkelte bedrifter kunne fremstå både med manglende observasjoner og ekstremt avvikende observasjoner. I det følgende vil det bli redegjort for en algoritme som identifiserer og deretter glatter over «hull» og uregelmessigheter i datasettet. En ekstrem observasjon er en observasjon der sysselsettingen i et enkelt år avviker uforholdsmessig mye fra normen over tid. Ekstreme observasjoner kan skyldes feil ved koblingen mellom individ og bedrift over tid. Det er alltid et problem å trekke et skille mellom *reell variasjon* og *ekstrem variasjon*, her implisitt forstått som feilaktig variasjon. En kan imidlertid trekke opp visse tentative kriterier for hva som kan regnes som ekstremt avvikende og deretter korrigere de tilfellene som blir identifisert ved å erstatte den manglende eller ekstreme observasjonen med et glidende gjennomsnitt av observasjonen før og etter. Deretter kan den korrigerte tidsserien tjene som en mal og som en indikator på omfanget av uregelmessighetene i koblingen mellom individ og bedrift over tid.

Utgangspunktet for glattings-algoritmen er sysselsettingen i en bedrift over tid totalt sett og innen de tre utdanningsgruppene vi nytter i vår analyse. Dersom sysselsettingen mangler eller er, utfra gitte kriterier, ekstremt avvikende i en mellomliggende periode (for ekstreme observasjoner er det bare snakk om et enkelt år), blir den erstattet med gjennomsnittet av sysselsettingen i året rett før og rett etter denne perioden, forutsatt at «bnr» eller «anr» er det samme før og etterpå. Observasjonen blir vurdert som ekstrem dersom følgende tre kriterier *samlet* er oppfylt:

- i. Observasjonen utgjør mindre enn 5 prosent av gjennomsnittet av observasjonene før og etter
- ii. Summen av antall sysselsatte over 3-årsperioden skal være større enn 10
- iii. Sysselsettingen før og etter skal være større en 20 prosent av et glidende gjennomsnitt

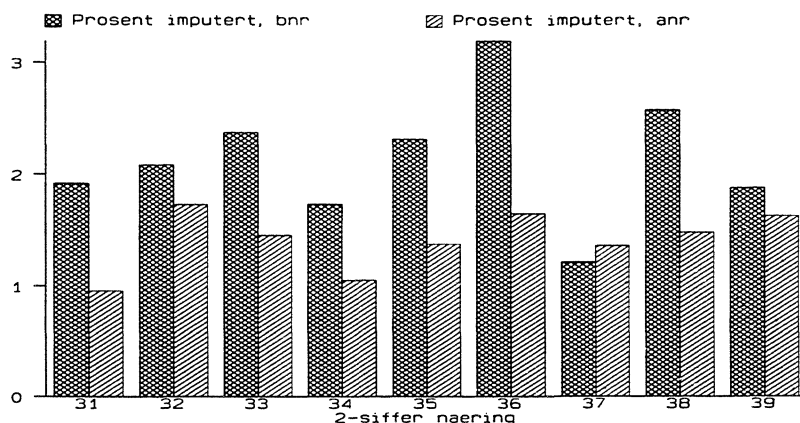
I vår analyse grupperte vi også sysselsettingen etter utdanningsnivå, (ii) er ment å begrense korreksjon innen bestemte utdanningsgruppe. En bedrift med 1 person med høy utdanning vil dersom han slutter gi en endring fra 1 til 0 sysselsatte i gruppen av høyt utdannede som ved (i) vil gjelde som en ekstrem endring. (iii) tar hensyn til gradvis vekst eller nedleggelse. Dersom observasjonen blir identifisert

³ En definisjon av de ulike jobbkreasjonsratene er gitt i vedlegg A

som ekstrem utfra de ovenfor gitte kriterier, blir den erstattet med gjennomsnittet av observasjonene før og etter.

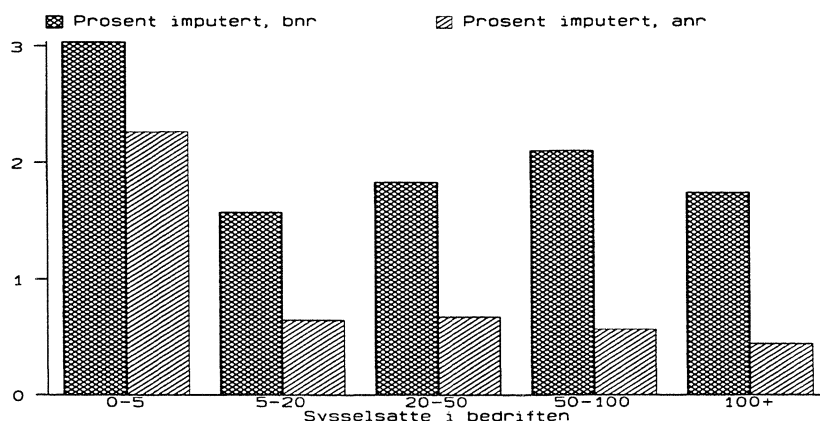
Figur 10 viser den prosentvise andelen manglende eller ekstreme observasjoner som er blitt korrigeret i ulike næringer i industrien (2-siffer, ISIC):

Figur 10. Korreksjon etter næring



Som vi ser av figuren blir gjennomgående færre observasjoner korrigeret dersom vi nyter «anr» (1-2 prosent korrigererte) fremfor «bnr» (2-3 prosent korrigererte) en refleksjon at algoritmen korrigerer for «99-bedrifter». Omfanget av korreksjonene mellom de ulike næringene med opp mot 1 prosent ved bruk av «anr» og 2 prosent ved bruk av «bnr». Figur 11 viser at antallet korrigererte observasjoner er klart høyere for bedrifter med færre enn 5 sysselsatte. Dersom vi aggregerer på basis av «bnr» korrigerer vi i 3 prosent av tilfellene i gruppen av bedrifter med mindre enn 5 ansatte mens den er mellom 1.5 og 2 prosent for større bedrifter.

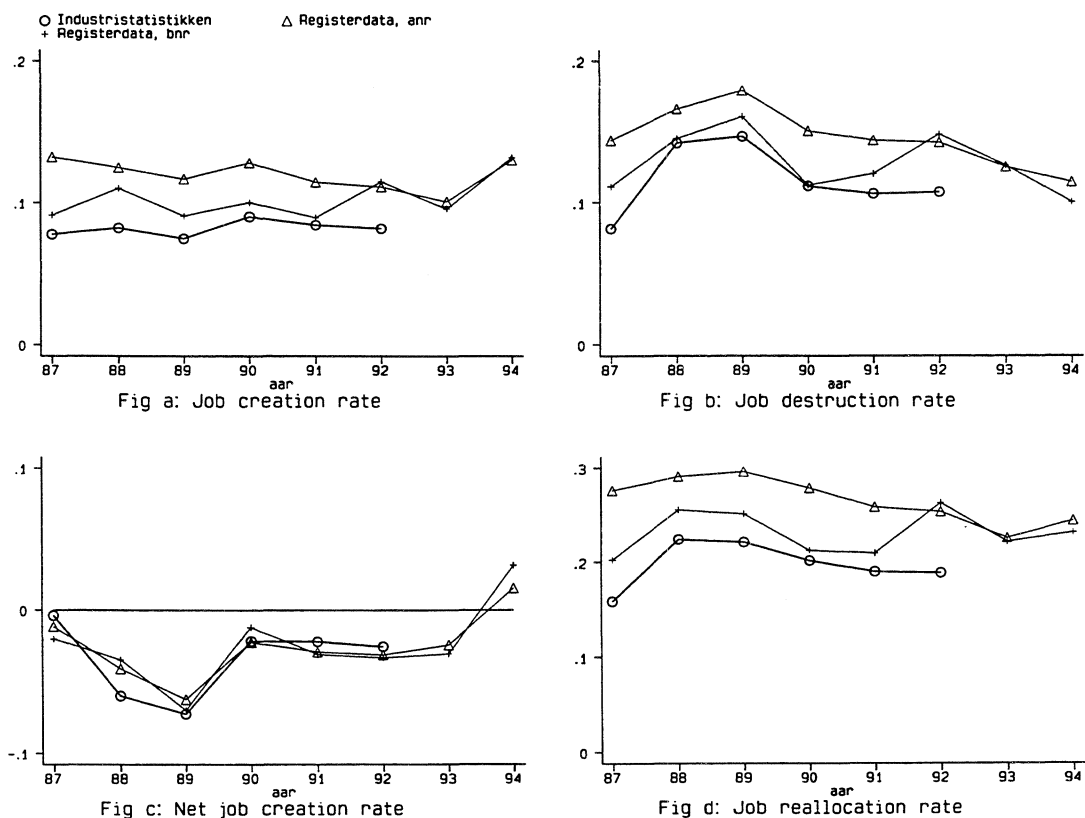
Figur 11. Korreksjon etter bedriftstørrelse



6.2. En sammenligning med Industristatistikken

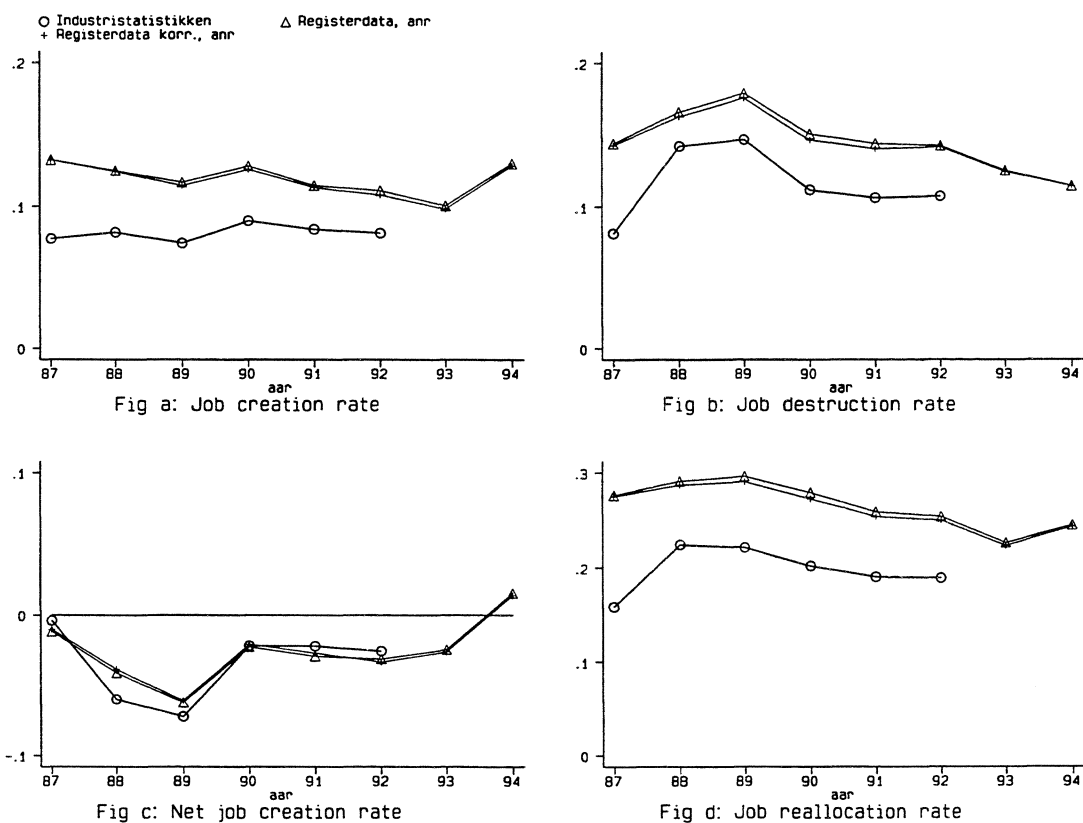
Rater for totale endringer i arbeidsstokken på bedriftsnivå kan også genereres på basis av Industristatistikken. Vi vil i det følgende sammenligne jobbskapsningsrater generert på basis av Registerdatene og Industristatistikken. Figur 12 består av fire delfigurer som viser utviklingen i jobbkreasjon, jobbdestruksjon, netto jobbskaping og jobb reallokering i Industrien for perioden 1987 til 1994. Målsetningen er å se på samsvaret mellom ratene generert på basis av Industristatistikken og ratene generert på basis av Registerdataene. Vi ser av Figur 11a og Figur 11b at Registerdataene gir høyere jobbskaping- og jobbdestruksjons rater enn Industristatistikken. Dersom vi antar at Industristatistikken gir det korrekte bildet av situasjonen kan en slutte følgende: Ratene er høyere i Registerdataene som et resultat av en for høy variasjon på bedriftsnivå. Denne variasjonen skyldes uregelmessigheter i koblingen mellom individ og bedrift over tid. En analyse basert på «bnr»et i Registerdataene vil influeres av «99-bedriftene» som i kraft av å fremstå som manglende i enkelte bedrifters tidsserie vil resultere i et for høyt anslag på nedleggelser og etableringer. I tillegg vil feil ved tilordning av bedriftsnummeret resultere i høyere rater. En analyse basert på «anr» i Registerdataene blir høyere siden skifte i eierforhold blir regnet som nedleggelse og nyetablering. Det er imidlertid verdt å merke seg at selv om ratene blir høyere er konjunkturforløpet er det samme. Dette kommer særlig klart frem ved de sammenfallende netto jobbskapsningsratene.

Figur 11. Jobbkreasjon og destruksjon basert på «anr» og «bnr» i Registerdataene og Industristatistikken

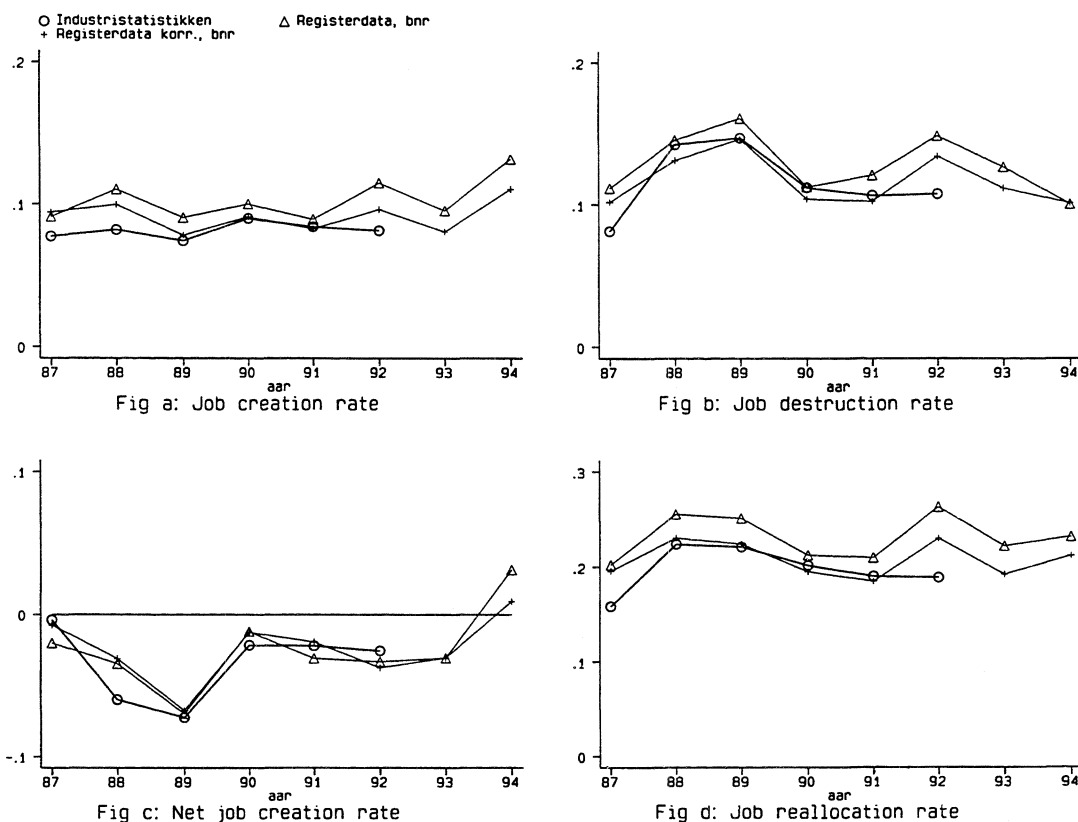


Figur 13 og 14 tilsvarer figur 12 og viser hvordan glattingen av dataene påvirker ratene i Registerdataene. Av figur 12 ser vi at glattingen har en knapt merkbar effekt på ratene dersom vi baserer oss på «anr» i Registerdataene. Siden glattingen er ment å korrigerer for store uregelmessigheter i sysselsettingen på bedriftsnivå som følge av mangler ved koblingen mellom individ og virksomhet vil korrigeringen bli mer avgjørende dersom vi nytter «bnr». Dette fremkommer av Figur 14 og er et resultat av «99-bedriftene» og eventuelle feil som er oppstått ved tilordning av «bnr».

Figur 12. Jobkreasjon og destruksjon basert på «anr» i en korrigert og ukorrigert versjon sammenlignet med Industristatistikken



Figur 13. Jobkreasjon og destruksjon basert på «bnr» i en korrigeret og ukorrigeret versjon sammenlignet med Industristatistikken



7. Konklusjon

Vi har i dette notatet fokusert på koblingen mellom arbeidstaker og bedrift i Registerdataene. Ved å nytte Industristatistikken som et sammenligningsgrunnlag kunne vi sammenligne forskjeller i sysselsettingstall og antall bedrifter i de to datasettene. Videre undersøkte i hvor stor grad det var mulig å koble informasjon fra de to datasettene på bedriftsnivå.

Avviket i sysselsettingstall mellom Registerdataene og Industristatistikken for hele industrien var aldri mer enn i underkant av 3 prosent. Avviket viste imidlertid stor variasjon avhengig av størrelsen på bedriftene. Avviket var størst dersom vi ønsket å identifisere bedriften ved «bnr» som samsvarer med bedriftenes identifikasjonsnummer i Industristatistikken. Sysselsettingen var i dette tilfellet lavere i Registerdataene for alle størrelsesgrupper med færre enn 200 ansatte. Størst var avviket for bedrifter med mindre enn 5 ansatte (15 prosent lavere). Avviket i sysselsetting viste også stor variasjon innenfor ulike 2-siffer næringer (ISIC). Ulikheter i klassifisering av bedrifters næringstilhørighet er en viktig forklaringsfaktor her.

En sammenligning av bedrifter i de to datasettene viste at antall bedrifter i Registerdataene med et tilordnet «bnr» var lavere i Industristatistikken. Dette avviket var størst for de minste bedriftene, en refleksjon av de mange «99-bedriften» i denne gruppen.

Kobling av bedrifter i Registerdataene med bedrifter i Industristatistikken på basis av «bnr» viste at 60-70 prosent fant en makker. Koblingsgraden var lavest for de minste bedriftene. For bedrifter med færre enn 5 sysselsatte, fant bare i overkant av 40 prosent en makker mens koblingsgraden var rundt 80 prosent for større bedrifter.

Det viste seg å være tildels store forskjeller i sysselsetting på bedriftsnivå mellom de to datasettene. Med utgangspunkt i de bedriftene som fant en «makker», ble korrelasjonen mellom bedriftenes sysselsetting i henhold til Registerdataene og i henhold til Industristatistikken estimert til 0.92.

Ved å sammenligne jobbskappingsrater generert på basis av Industristatistikken med tilsvarende rater generert på basis av Registerdataene fremkom det samme konjunkturforløpet. Jobbreallokeringsratene var imidlertid høyere i Registerdataene. Høyest var jobbreallokeringsratene dersom vi baserte oss på arbeidsgivernummeret og det bedriftsspesifikke undernummeret. Dette var en følge av at endringer i eierforhold endrer arbeidsgivernummeret slik at det fremstår som en nedleggelse og nyetablering. Dersom vi baserer oss på «bnr» i Registerdataene ble jobbreallokeringsraten funnet å være 1-3 prosent høyere i Registerdataene enn i Industristatistikken. Denne høyere jobbreallokeringsraten er et resultat av feil og mangler både ved koblingen mellom individ og bedrift over tid og i tilordningen «bnr». Likevel må man kunne konkludere med at jobbskapingratene generert på basis av Registerdataene samsvarer overraskende bra med de tilsvarende rater generert på basis av Industristatistikken.

Referanser

Grorud, A. K. (1993): Supplering til brukerdokumentasjon for det sentrale bedrifts- og foretaksregisteret, Seksjon for Bedriftsregister.

Tor Jakob Klette og Svein Erik Førre(1995): Innovation and Job Creation in a Small Open Economy. Til publisering i *Economics of Innovation and New Technology*

Olsen, G. (1993): Rutiner for ajorhold av det sentrale bedrifts- og foretaksregisteret. Seksjon for metoder og standarder. Personstatistikk.

Salvenes, K. G. og S. E. Førre (1997): Job creation, heterogenous workers and technical change. *Rapport, SNF*.

Utne, H. og E. Vassnes (1995): Kobling av A/A- og LTO-registeret. *Notat, SSB*

Vassenden, E.(1995): The register of the level of education: A product of the intergration of other registers, census and surveys. *Notat, SSB*

Vedlegg A

La N_{it} være sysselsettingen i bedrift i ved tidspunkt t . La

$$\Delta N_{it}^+ \equiv \begin{cases} 0 & : N_{it} \leq N_{it-1} \\ N_{it} - N_{it-1} & : \textit{ellers} \end{cases}$$

Der Ω_t er mengden av alle bedrifter i industrien i år t . La videre

$$N_{it} = \left(\sum_{i \in \Omega_t} N_{it} + \sum_{j \in \Omega_{t-1}} N_{jt-1} \right) / 2$$

Jobb kreasjonsraten er definert som:

$$\theta_t^{JC} \equiv \frac{\sum_{i \in \Omega_t} \Delta N_{it}^+}{N_t}$$

Jobb destruksjonsraten er tilsvarende definert som:

$$\theta_t^{JD} \equiv \frac{\sum_{i \in \Omega_t} \Delta N_{it}^-}{N_t}$$

Jobb reallokeringsraten er definert som:

$$\theta_t^{JR} \equiv \theta_t^{JC} + \theta_t^{JD}$$

Netto jobb kreasjonsraten er definert som:

$$\theta_t^{JR} \equiv \theta_t^{JC} - \theta_t^{JD}$$

Vedlegg B

Næringskoder (ISIC) :

- 31: Produksjon av næringsmidler, drikkevarer og tobakk
- 32: Produksjon av tekstilvarer, bekledningsvarer, lær og lærvarer
- 33: Produksjon av trevarer
- 34: Treforedling, grafisk produksjon og forlagsvirksomhet
- 35: Produksjon av kjemiske produkter, mineral-, olje-, kull-, gummi- og plastprodukter
- 36: Produksjon av mineralske produkter
- 37: Produksjon av metaller
- 38: Produksjon av verkstedsprodukter
- 39: Industriproduksjon ellers

De sist utgitte publikasjonene i serien Notater fra Forskningsavdelingen

- 94/18 Y. Vogt: Innføring i FAME
- 94/22 M.W. Arneberg: LOTTE-TRYGD. Teknisk dokumentasjon
- 95/5 D. Fredriksen: MOSART Teknisk dokumentasjon
- 95/7 K. Olsen: Nytte- og kostnadsvirkninger av en norsk oppfyllelse av nasjonale utslippsmålsettinger
- 95/15 T. Karlsen: Opptimal karbonbeskatning og virkningen på norsk petroleumsformue
- 95/17 Å. Cappelen, T. Skjerpen og J. Aasness: Konsumetterspørsel, tjenesteproduksjon og sysselsetting. En mikro til makroanalyse
- 95/24 H.T. Mysen: Nordisk energimarkedsmodell. Dokumentasjon av delmodell for energi- etterspørsel i industrien
- 95/26 I. Aslaksen, T. Fagerli og H.A. Gravningsmyhr: Produksjon og konsum i husholdningene
- 95/29 B.E. Naug: Eksport- og importlikninger i KVARTS
- 95/31 B.E. Naug: Etterspørsel etter arbeidskraft - en litteraturoversikt
- 95/35 T.J. Klette: Vekst og produktivitet i norsk industri. Hovedrapport fra et NFR-prosjekt
- 95/40 L. Lerskau: Oversikt over konjunkturindikatorer i databasen NORMAP og FAME
- 95/46 B.E. Naug: Estimering av eksportrelasjoner på disaggregerte kvartalsdata
- 95/47 K. Moum: Beregning av bruttoproduksjon og eierinntekt i boligsektoren i nasjonalregnskapet - noen metodiske synspunkter
- 95/52 T. Kornstad: Simulering av konsum og arbeidstilbud i et livsløpsperspektiv
- 95/56 A. Langørgen: Faktorer bak kommunale variasjoner i utgifter til sosialhjelp og barnevern
- 95/58 T. W. Karlsen: Energimarkedet fra 1973 og fram mot 2010
- 96/3 I. M. Smestad: Valg under usikkerhet: En analyse av eksperimentdata basert på kvalitative valgbehandlingsmodeller
- 96/8 B. Lian og K. O. Aarbu: Dokumentasjon av LOTTE-AS
- 96/9 D. Fredriksen: Datagrunnlaget for modellen MOSART, 1993
- 96/10 S. Grepperud og A. C. Bøeng: Konsekvensene av økte oljeavgifter for råoljepris og etterspørsel etter olje. Analyser i PETRO og WOM
- 96/16 K. Gerdrup: Inntektsfordeling og økonomisk vekst i norske fylker: En empirisk studie basert på data for perioden 1967-93
- 96/31 A. Bruvoll og H. Wiig: Konsekvenser av ulike håndteringsmåter for avfall
- 96/33 M. Rolland: Militærutgifter i Norges prioriterte samarbeidsland
- 96/35 A.C. Hansen: Analyse av individers preferanser over lotterier basert på en stokastisk modell for usikre utfall
- 96/36 B.H. Vatne: En dynamisk spillmodell: Dokumentasjon av dataprogrammer
- 96/44 K.-G.Lindquist og B.E.Naug: Makro-økonometriske modeller og konkurranseevne.
- 96/45 R. Golombek og S. Kverndokk (red): Modeller for elektrisitets- og gassmarkedene i Norge, Norden og Europa.
- 96/53 F.R. Aune: Konsekvenser av en nordisk avgiftsharmonisering på elektrisitetsområdet.
- 97/2 E. Berg og K. Rypdal: Historisk utvikling og fremskrivning av forbruket av noen miljøskadelige produkter
- 97/5 Å. Cappelen: SSBs arbeid med investeringsrelasjoner: erfaringer og planer
- 97/30 K.-G. Lindquist: Database for energiintensive næringer. Tall fra industristatistikken
- 97/35 A. Langørgen: Faktorer bak variasjoner i kommunal ressursbruk til pleie og omsorg
- 97/36 S. E. Førre: Registerdataene i lys av industristatistikken

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

