



Anna-Karin Mevik

Notater

Estimering av månedlig omsetning innenfor bergverksdrift og industri

1. Innledning	2
2. Omsetningsstatistikken for industrien.....	2
3. Registerfeil og målefeil.....	6
3.1. Registerfeil.....	6
3.2. Målefeil.....	8
4. To alternative estimatorer for den månedlige omsetningen	11
4.1. Modell.....	11
4.2. Estimator A.....	13
4.2.1. Modellbasert analyse av estimator A.....	14
4.2.2. Talleksempel	16
4.3. Estimator B.....	19
4.3.1. Modellbasert analyse av estimator B.....	20
4.3.2. Talleksempel	22
5. Sammendrag	25
Referanser	26
Vedlegg A	27
Vedlegg B.....	28
Vedlegg C	30
Vedlegg D	33

1. Innledning

Månedsomsetningen innen olje- og gassutvinning, bergverksdrift, industri og kraftforsyning blir estimert i Omsetningsstatistikken for industrien. I dette notatet skal vi se på alternative estimatorer for estimeringen av omsetningen innen bergverksdrift og industri. (Grunnen til at vi ikke ser på olje- og gassutvinning og kraftforsyning er at datagrunnlaget for disse næringene er litt annerledes slik at de må behandles spesielt).

Til estimeringen av månedsomsetningen brukes blant annet registerbasert statistikk fra Skattedirektoratet. Det er ønskelig å gå vekk fra dette, og i stedet bare bruke data fra BoF (Bedrifts- og foretaksregisteret) i tillegg til de innsamlede utvalgsdataene. Dette blant annet fordi dataene fra Skattedirektoratet er registrert terminvis (januar/februar, mars/april, osv.) og på foretaksnivå. Dataene må derfor først fordeles ned på månedsbasis og bedriftsnivå før de kan brukes i estimeringen, og dette arbeidet er resurskrevende. (Men det er ønskelig å bruke dataene fra Skattedirektoratet som referanse/kontroll for de estimerte omsetningstallene).

I dette notatet skal vi derfor se på alternative estimatorer som ikke bruker tall fra Skattedirektoratet. Som registergrunnlag for estimeringen har vi valgt å bruke IS (Strukturstatistikk for industri), dvs. vi bruker ikke BoF.¹ BoF er det mest oppdaterte registeret med tanke på hvilke bedrifter som er med i populasjonen, men mangler viktige registervariabler for en del av bedriftene. IS gir et dekkende bilde over populasjonen slik den var to år tilbake i tid. Dvs. IS er ikke like oppdatert som BoF med hensyn til hvilke bedrifter som er med i den aktuelle populasjon, men til gjengjeld mangler ikke IS registervariabler for noen av bedriftene. Dermed unngår vi problemet med manglende verdier når vi bruker IS.²

Vi har altså brukt IS og innsamlede opplysninger fra utvalget som datagrunnlag for estimeringen. Testing og valg av estimatorer er gjort på bakgrunn av data for 2005, og vi har endt opp med to estimatorer som ser ut til å estimere den månedlige omsetningen minst like bra som når data fra Skattedirektoratet brukes.

Vi starter i kapittel 2 med en kort presentasjon av Omsetningsstatistikken, der vi blant annet ser på populasjonen, utvalget og noen av aggregeringsnivåene det skal estimeres omsetningstall for. I kapittel 3 ser vi litt på registerfeil og målefeil. I kapittel 4 presenterer vi de to alternative estimatorene for estimering av månedlig omsetning. Til slutt avslutter vi med en kort oppsummering i kapittel 5.

2. Omsetningsstatistikken for industrien

I Omsetningsstatistikken for industrien estimeres den månedlige omsetningen til bedrifter som driver innen olje- og gassutvinning, bergverksdrift, industri og kraftforsyning. Populasjonen er definert ved BoF og omfatter alle bedrifter, unntatt enmannsbedrifter, som driver innen disse næringene. I 2005 var den omtrent 25 000 bedrifter i populasjonen, hvorav drøye 22 000 drev innen bergverksdrift og industri.

Fordi vi bare skal se på alternative estimatorer for omsetningen innen bergverksdrift og industri, begrenser vi oss stort sett til å omtale disse næringene.

Populasjonen er delt inn i mange delpopulasjoner som det skal estimeres omsetningstall for. Vi snakker gjerne om nivåer, eller aggregeringsnivåer. Det laveste nivået det estimeres tall for er

¹ I noen av korttidsstatistikkene for industri er det valgt å bruk BoF som registergrunnlag, i andre er IS valgt som registergrunnlag.

² Opplysningene i BoF, for industribedriftene, kommer fra IS. Dermed mangler vi opplysninger for de nye bedriftene i BoF som ikke fins i IS.

sektornivå, også kalt bearbeidingsnivå, men disse tallene publiseres ikke. Sektornivå tilsvarer i all hovedsak en inndeling etter næringsgruppe, dvs. 4-sifret NACE (se Standard for næringsgruppering). Bergverksdrift og industri er delt opp i 120 sektorer.

Alle høyere aggregeringsnivå er sammensatt av en eller flere sektorer, og estimert omsetning for disse nivåene framkommer ved å summere over estimert omsetning til sektorene. Eks. på et høyere aggregeringsnivå er det som kalles publiseringsnivå. Dette nivået deler inn bergverksdrift og industri i 20 næringer som vist i tabell 1.

Tabell 1 Populasjonstall for bergverksdrift og industri i 2005, fordelt på publiseringsnivå*

Næring	NACE	Antall sektorer	Antall bedrifter	Årsomsetning	
				i millioner kr	% av tot. omsetning
Bergverksdrift	10, 12-14	8	726	9 433	1.5
Næringsmiddelindustri	151-158	10	2 287	127 320	20.3
Nytelsesmiddelindustri	159-160	4	114	21 057	3.4
Tekstil- og bekledningsindustri	17-19	13	1 371	6 570	1.0
Trelast- og trevareindustri	20	5	2 111	23 811	3.8
Treforedling	21	3	130	19 277	3.1
Forlag og grafisk industri	22	4	3 398	40 230	6.4
Oljeraffinering	23	8	229	67 712	10.8
Kjemisk industri, unntatt kjemiske råvarer	242-247				
Kjemiske råvarer	241	3	118	31 362	5.0
Gummivare- og plastindustri	25	2	468	9 042	1.4
Mineralproduktindustri	26	7	1 002	20 479	3.3
Metallindustri, unntatt ikke-jernholdige metaller	271-273, 275	4	152	13 503	2.2
Ikke-jernholdige metaller	274	2	48	40 054	6.4
Metallvareindustri	28	7	2 449	27 356	4.4
Maskinindustri	29	7	2 696	42 682	6.8
Elektroteknisk og optisk industri	30-33	15	1 351	38 339	6.1
Transportmiddelindustri	34,35111-3,6,7	9	1 246	32 993	5.3
Oljeplattformer	351114,5	1	151	38 443	6.1
Møbelindustri og annen industri	36-37	8	1 996	16 987	2.7
Bergverksdrift og industri i alt		120	22 043	626 650	100.0

*) Tallene for oljeraffinering og kjemisk industri, unntatt kjemiske råvarer, er slått sammen pga. konfidensialitet.

Målt ut fra antall bedrifter er det forlag og grafisk industri som er den største næringen på publiseringsnivå. Måler vi derimot ut fra omsetning er det næringsmiddelindustrien som er desidert størst. I 2005 sto denne næringen for hele 20,3% av totalen for årsomsetningen innen bergverksdrift og industri. Nest størst er oljeraffinering og kjemisk industri (unntatt kjemiske råvarer) med 10,8% av den totale årsomsetningen, mens resten av næringene har mellom 1,0 og 6,8% av årsomsetningen i 2005.

Bruttoutvalget, inklusive olje- og gassutvinning og kraftforsyning, er på om lag 1800 bedrifter, og frafallet er på bare ca. 2%. Utvalget er trukket stratifisert med en trekksannsynlighet proporsjonal med størrelsen på bedriften målt i antall sysselsatte, der strataene er bestemt ut fra omsetning, sysselsetting og næring. Stratifiseringen og allokeringen av utvalgsenheter er gjort slik at alle bedrifter med minst 100 sysselsatte, og alle med minst 10% av årsomsetningen på sektornivå, er tatt med i utvalget. Videre trekkes det ikke bedrifter blant de som har færre enn 10 sysselsatte. Fordelingen av utvalgsenheter

mellom de resterende strataene er bestemt ved en optimal allokering basert på bedriftenes omsetning. Denne utvalgsplanen gjør at dekningsgraden for utvalget, med hensyn til omsetning, er veldig stor.

Bruttoutvalget holdes relativt konstant gjennom året, men pga. frafall blir det litt forskjell i nettoutvalget fra måned til måned. Dekningsgraden for nettoutvalget er likevel nokså stabil gjennom året. For bergverksdrift og industri har vi f.eks. i 2005 at nettoutvalget varierer fra 1 574 til 1 687 bedrifter, mens dekningsgraden varierer fra 65,9 til 67,4%. I tabell 2 ser vi hvordan nettoutvalget for november 2005 fordeler seg på publiseringsnivå.

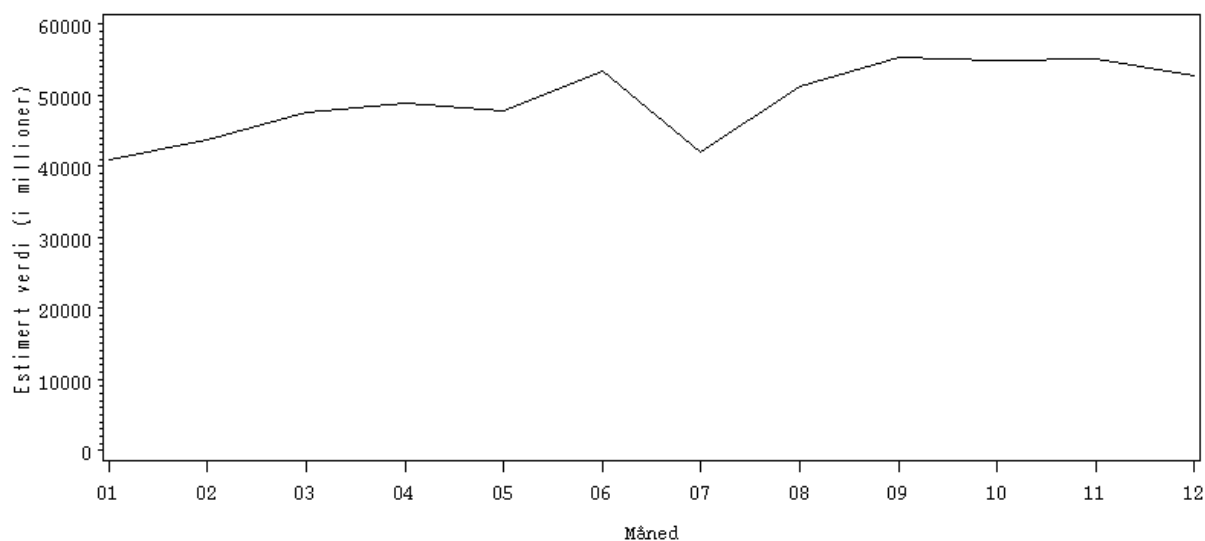
Tabell 2 Utvalget for bergverksdrift og industri i 2005, fordelt på publiseringsnivå*

Næring	NACE	Antall bedrifter	Dekningsgrad målt ved årsomsetning (%)
Bergverksdrift	10, 12-14	47	37.5
Næringsmiddelindustri	151-158	245	56.1
Nytelsesmiddelindustri	159-160	18	84.5
Tekstil- og bekledningsindustri	17-19	67	51.3
Trelast- og trevareindustri	20	127	56.7
Treforedling	21	31	76.1
Forlag og grafisk industri	22	153	60.5
Oljeraffinerings og kjemisk industri, unntatt kjemiske råvarer	23 og 242-247	33	89.5
Kjemiske råvarer	241	45	84.8
Gummivare- og plastindustri	25	51	56.4
Mineralproduktindustri	26	100	55.1
Metallindustri, unntatt ikke-jernholdige metaller	271-273, 275	33	92.1
Ikke-jernholdige metaller	274	19	97.6
Metallvareindustri	28	178	52.7
Maskinindustri	29	136	57.4
Elektroteknisk og optisk industri	30-33	78	52.5
Transportmiddelindustri	34,35111-3,6,7	97	70.1
Oljeplattformer	351114,5	50	72.7
Møbelindustri og annen industri	36-37	88	48.1
Bergverksdrift og industri i alt		1 596	67.3

*) Tallene gjelder for nettoutvalget i november 2005

Gjeldende estimeringsopplegg i Omsetningsstatistikken er en ratebasert estimator, der datagrunnlaget er registerdata fra Skattedirektoratet og BoF, samt de innsamlede opplysningene fra utvalget. I figur 1 ser vi hvordan den estimerte månedsomsetningen (med gjeldende estimator) for bergverksdrift og industri i alt utvikler seg fra måned til måned i 2005. Som vi ser er det stort sett en svak økning i den estimerte omsetningen utover året, med unntak av juli hvor omsetningen faller kraftig fra foregående måned, og desember hvor det er en svak nedgang fra foregående måned.

Estimert månedsomsetning i 2005, bergverksdrift og industri



Figur 1 Estimert månedsomsetning innen bergverksdrift og industri i 2005, med gjeldende estimeringsopplegg i Omsetningsstatistikken. De estimerte verdiene er i millioner kr.

Ved å summere de estimerte månedstallene får vi estimert verdi for årsumsetningen. Den estimerte årsumsetningen kan vi så i ettertid sammenligne mot omsetningstallene fra BoF. I BoF har vi nemlig årsumsetning for alle bedrifter i populasjonen, men pga. etterslep er disse tallene ikke tilgjengelig før nærmere to år etter gjeldende år. Det betyr f.eks. at 2005-omsetningen ikke er på plass i BoF før i 2007. Vi skal se på tallene i BoF som fasit for årsumsetningen.

Hvis vi sammenligner den estimerte årsumsetningen for 2005 (beregnet med gjeldende estimator) med fasiten fra BoF, finner vi at vi treffer relativt bra. Den estimerte årsumsetningen for bergverksdrift og industri i alt er på nesten 594 milliarder kr, mens den faktiske årsumsetningen er i underkant av 627 milliarder. Dvs. vi underestimerer årsumsetningen i 2005 med ca. 33 milliarder. Regnet i % av faktisk årsumsetning tilsvarer dette et avvik på 5,2%, som vi må kunne si er et nokså lite avvik.

Gjør vi tilsvarende sammenligning på publiseringnivå, får vi varierende resultater (se tabell 3). For en god del næringer treffer vi veldig bra med den estimerte omsetningen, men det fins også en del næringer der vi bommer veldig mye. F.eks. bommer vi med hele 23,0% for næringsmiddelindustrien, mens vi for maskinindustri kun har et avvik på -0,7%. (Hvor store avvik vi har varierer fra år til år, se tabell A1 i vedlegg A).

Merk at denne sammenligningen gjelder årstall. Vi får ikke gjort tilsvarende sammenligning for månedstallene fordi det kun er for årstallene at vi i ettertid får fasit fra BoF.

Tabell 3 Estimert årsomsetning innen bergverksdrift og industri i 2005 (med gjeldende estimeringsopplegg i Omsetningsstatistikken), fordelt på publiseringsnivå. Avviket er gitt ved faktisk årsomsetning (BoF) minus estimert verdi

Næring	NACE	Estimert årsomsetning (i millioner kr)	Avvik mot faktisk årsomsetning (%)
Bergverksdrift	10, 12-14	11 193	-18.7
Næringsmiddelindustri	151-158	98 090	23.0
Nytelsesmiddelindustri	159-160	19 997	5.0
Tekstil- og bekledningsindustri	17-19	6 885	-4.8
Trelast- og trevareindustri	20	22 958	3.6
Treforedling	21	16 538	14.2
Forlag og grafisk industri	22	37 186	7.6
Oljeraffinering og kjemisk industri, unntatt kjemiske råvarer	23 og 242-247	60 950	10.0
Kjemiske råvarer	241	33 427	-6.6
Gummivare- og plastindustri	25	8 900	1.6
Mineralproduktindustri	26	21 197	-3.5
Metallindustri, unntatt ikke-jernholdige metaller	271-273, 275	14 308	-6.0
Ikke-jernholdige metaller	274	42 756	-6.7
Metallvareindustri	28	28 068	-2.6
Maskinindustri	29	42 967	-0.7
Elektroteknisk og optisk industri	30-33	44 265	-15.5
Transportmiddelindustri	34,351111-3,6,7	27 758	15.9
Oljeplattformer	351114,5	38 899	-1.2
Møbelindustri og annen industri	36-37	17 629	-3.8
Bergverksdrift og industri i alt		593 971	5.2

For mer informasjon om Omsetningsstatistikken for industrien, se Omsetningsstatistikk for olje- og gassutvinning, bergverksdrift, industri og kraftforsyning (NOS D 251) eller Om statistikken.

3. Registerfeil og målefeil

All statistikk er heftet med registerfeil og målefeil i større eller mindre grad. Her skal vi se på registerfeil som vi får når vi bruker IS som registergrunnlag ved estimeringen.³ Vi skal også se litt på målefeilene vi har i de innsamlede omsetningstallene fra utvalget, ut fra antagelsen om at det er omsetningstallene i BoF som er korrekte.

3.1. Registerfeil

Vi kan se på IS som et register over populasjonen slik den var to år tilbake i tid. Det betyr at når vi bruker IS som registergrunnlag ved estimeringen, så blåser vi opp utvalget til feil populasjon.

For estimeringen av månedsomsetningen i 2005 f.eks. betyr det at vi blåser opp utvalget til en populasjon som består av de samme bedriftene som i 2003. I 2003 var det 22 557 bedrifter innen bergverksdrift og industri, mot 22 043 bedrifter i 2005. Sammenligner vi bedrift for bedrift finner vi at 18 830 bedrifter var med i både 2003- og 2005-populasjonen. 3 727 av bedriftene i 2003 var altså ikke

³ Vi ville også fått registerfeil om vi hadde brukt BoF som registergrunnlag, men da i form av manglende registervariable.

med i 2005-populasjonen (avganger) og 3 213 av bedriftene i 2005 var ikke med i 2003-populasjonen (tilganger).

Både avgangene og tilgangene er stort sett små bedrifter målt ved omsetning. Men tilgangene er noe større enn avgangene. F.eks. er gjennomsnittlig 2003-omsetning for alle bedriftene i 2003 og for de 3 727 avgangene henholdsvis 23,1 og 4,9 millioner kr, mens gjennomsnittlig 2005-omsetning for alle bedriftene i 2005 og for de 3 213 tilgangene henholdsvis er 28,4 og 8,3 millioner kr.

For å få et inntrykk av effekten av registerfeilen har vi gjort et eksperiment der vi har estimert årsomsetningen for 2005 to ganger: Først en gang med registerfeil, så i en konstruert situasjon uten registerfeil. Årsomsetningen er estimert direkte, uten å gå veien om estimerte månedsomsetninger. Som estimator har vi brukt stratifisert rateestimator, med 2003-omsetningen som x -variabel og stratuminddeling tilsvarende næringsinndelingen på publiseringsnivå.

I situasjonen med registerfeil er IS brukt som registergrunnlag for estimeringen. Estimatoren for stratum h kan dermed skrives som

$$\hat{Y}_h = \frac{\sum_{i \in \text{utv. stratum } h} y_i}{\sum_{i \in \text{utv. stratum } h} x_i} \cdot X_h,$$

der $y_i = 2005$ -omsetningen til bedrift i , $x_i = 2003$ -omsetningen til bedrift i , og $X_h = \sum_{i \in \text{IS stratum } h} x_i$ er summen av 2003-omsetningen til bedriftene som ifølge IS fins i stratum h .

For tilfellet uten registerfeil trenger vi en situasjon hvor registergrunnlaget og den faktiske populasjonen består av de samme bedriftene, og hvor registervariabelen x_i ikke mangler for noen bedrifter. For å få til dette har vi redusert IS og 2005-populasjonen til å kun bestå av de 18 830 bedriftene som fins i både 2003-populasjonen (IS) og 2005-populasjonen. Dermed har vi fått en konstruert situasjon uten registerfeil, og estimatoren for stratum h kan skrives som

$$\tilde{Y}_h = \frac{\sum_{i \in \text{utv. stratum } h} y_i}{\sum_{i \in \text{utv. stratum } h} x_i} \cdot X_h^-,$$

der $X_h^- = \sum_{i \in \text{redusert pop., stratum } h} x_i$ er summen av 2003-omsetningen til bedriftene som er med i stratum h i den reduserte IS (dette er de samme bedriftene som fins i stratum h i den reduserte 2005-populasjonen).

Merk at y_i er den faktiske 2005-omsetningen som vi i ettertid kjenner fra BoF. Dermed unngår vi å bruke den innrapporterte omsetningen som vi kan beregne via de innsamlede månedsomsetningene, og som er heftet med målefeil. På den måten blander vi ikke inn effekten av målefeil i denne estimeringen.

Resultatet av estimeringen er gitt i tabell 4. Ikke uventet får vi best estimat når vi ikke har registerfeil for de fleste av næringene. Men for noen næringer har faktisk registerfeilen ført til bedre estimat. Generelt må vi likevel si at registerfeil fører til dårligere estimering. For bergverksdrift og industri i alt og for en del av næringene er estimatet med registerfeil ikke vesentlig dårligere enn estimatet uten registerfeil, mens for andre næringer har registerfeilen gitt et estimat som er en del dårligere. (Hvor

mye dårligere estimeringen blir pga. registerfeil avhenger blant annet av hvor stor forskjell det er på registergrunnlaget og den aktuelle populasjonen, og dette vil variere fra år til år).

Tabell 4 Avvik mellom estimert årsomsetning og faktisk årsomsetning (BoF) i 2005. Her er årsomsetningen estimert direkte uten å gå veien om månedsomsetningene, og med data uten målefeil. Avviket er gitt ved faktisk årsomsetning minus estimert verdi

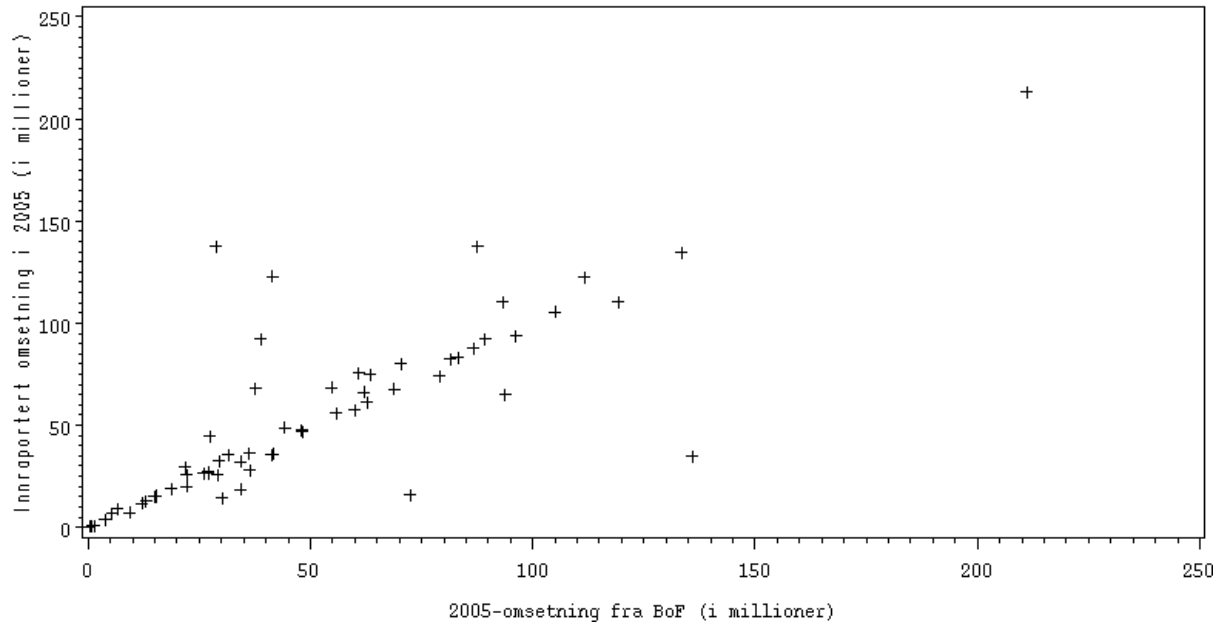
Næring	NACE	Avvik (%)	
		for den konstruerte situasjon uten registerfeil	når vi har registerfeil
Bergverksdrift	10, 12-14	2.6	3.7
Næringsmiddelindustri	151-158	1.3	1.2
Nytelsesmiddelindustri	159-160	0.1	0.4
Tekstil- og bekledningsindustri	17-19	1.9	6.0
Trelast- og trevareindustri	20	3.2	3.1
Treforedling	21	8.7	5.7
Forlag og grafisk industri	22	-1.2	0.7
Oljeraffinering og kjemisk industri, unntatt kjemiske råvarer	23 og 242-247	0.1	6.6
Kjemiske råvarer	241	-3.2	0.7
Gummivare- og plastindustri	25	0.4	-5.9
Mineralproduktindustri	26	2.6	3.0
Metallindustri, unntatt ikke-jernholdige metaller	271-273, 275	-1.8	-5.4
Ikke-jernholdige metaller	274	-5.2	-7.2
Metallvareindustri	28	-1.6	0.0
Maskinindustri	29	1.7	7.2
Elektroteknisk og optisk industri	30-33	-1.9	-3.7
Transportmiddelindustri	34,35111-3,6,7	-2.0	-1.3
Oljeplattformer	351114,5	-3.0	-7.3
Møbelindustri og annen industri	36-37	2.4	6.5
Bergverksdrift og industri i alt		-0.1	0.8

3.2. Målefeil

Her skal vi se litt på hvilke målefeil vi har i de innsamlede omsetningstallene fra utvalget, ut fra antagelsen om at det er omsetningstallene i BoF som er korrekte. (De innsamlede omsetningstallene er revidert, så det er snakk om gjenværende målefeil). Fordi vi bare har årsomsetningen i BoF, får vi bare sett på målefeilene for årsomsetningen, ikke for månedsomsetningene.

De fleste bedriftene i utvalget er med alle 12 månedene gjennom året. For disse kan vi beregne innrapportert årsomsetning som vi i ettertid kan sammenligne mot omsetningen i BoF. Vi har gjort denne sammenligningen med 2005-tallene, og det viste at det er målefeil i alle dataene, men at de fleste av feilene kun er småavvik som oppveier hverandre. (Se figur 2). Disse feilene vil derfor ikke medføre noen estimeringsskjevhet, og vi velger å kalle dem små tilfeldige målefeil. Resten av feilene kaller vi store systematiske målefeil. Det er ikke så veldig mange av disse målefeilene, men noen av dem er veldig store, og da ofte ved at den innrapporterte verdien er alt for stor. Vi kan derfor ikke regne med at de store målefeilene vil oppveie hverandre. Tvert imot må vi regne med at de medfører en estimeringsskjevhet. (Se vedlegg B for flere figurer).

Målefeil



Figur 2 Figuren viser sammenhengen mellom innrapportert 2005-omsetning og 2005-omsetningen fra BoF, for utvalgsbedriftene innen tekstil- og bekledningsindustri. Den innrapportert omsetningen er kun beregnet for bedriftene som har vært med i nettoutvalget alle 12 månedene i 2005.

Merk at selv om vi i ettertid kan se at det er en stor målefeil i innrapportert årsomsetning for en bedrift, kan vi ikke vite om det er store målefeil i alle innrapporterte månedstallene for bedriften, eller om det bare er for noen av månedene at der er store målefeil. Tilsvarende kan vi ikke være sikre på at det ikke fins store målefeil i noen av månedstallene selv om det kun er en liten målefeil i årsomsetningen.

Denne beskrivelsen vi har gjort av målefeilene passer ganske bra for stort sett alle næringene. Det eneste er at det for noen næringer ikke er så tydelig hva som bør regnes som små tilfeldige målefeil og hva som bør regnes som store systematiske målefeil. (Se figur for næringsmiddelindustrien i vedlegg B).

For å få ett inntrykk av effekten av målefeilene har vi gjort et eksperiment der vi har estimert årsomsetningen for 2005 to ganger. En gang der vi har brukt data med målefeil, og en gang der vi har brukt data uten målefeil. Som i delkapittel 3.1 har vi estimert årsomsetningen direkte uten å gå veien om månedsomsetningene, og vi har brukt stratifisert rateestimator med 2003-omsetningen som x -variabel og stratuminndeling tilsvarende næringsinndelingen på publiseringsnivå.

For å ikke blande inn effekten av registerfeil har vi gjort estimeringen for den konstruerte situasjonen der vi ikke har registerfeil. Vi har redusert IS og 2005-populasjonen til kun å bestå av de bedriftene som fins i både 2003-populasjonen (IS) og 2005-populasjonen. Så bruker vi den reduserte IS som registergrunnlag for estimering av omsetningen til den reduserte 2005-populasjonen. Dermed har vi ikke registerfeil, verken i form av manglende x -variabel eller i form av feil bedrifter i registergrunnlaget.

I situasjonen der vi bruker data med målefeil i estimeringen er det de innrapporterte omsetningstallene fra utvalget som brukes. Lar vi w_i = innrapportert 2005-omsetningen for bedrift i , kan estimatoren for stratum h skrives som

$$\hat{Y}_h = \frac{\sum_{i \in \text{utv. stratum } h} w_i}{\sum_{i \in \text{utv. stratum } h} x_i} \cdot X_h^-$$

der $x_i = 2003$ -omsetningen til bedrift i , og $X_h^- = \sum_{i \in \text{redusert pop., stratum } h} x_i$ er summen av 2003-omsetningen til

bedriftene som er med i stratum h i den reduserte IS (dette er de samme bedriftene som fins i stratum h i den reduserte 2005-populasjonen).

For estimeringen uten målefeil bytter vi ut de innrapporterte omsetningstallene med omsetningstallene fra BoF. Dvs. estimatoren for stratum h er gitt ved

$$\tilde{Y}_h = \frac{\sum_{i \in \text{utv. stratum } h} y_i}{\sum_{i \in \text{utv. stratum } h} x_i} \cdot X_h^-$$

der $y_i = 2005$ -omsetningen fra BoF.

Resultatet av estimeringen er gitt i tabell 5. Som vi ser har vi fått dårligere estimat når vi bruker data med målefeilene enn når vi bruker data uten målefeil, unntatt for tre næringer. For en del av næringene er forverringen veldig stor, som f.eks. nytelsesmiddelindustrien. Vi kan også se at målefeilene har ført til flere overestimeringer. Overestimering betyr at den estimerte omsetningen er større enn den faktiske omsetningen (BoF), og da blir avviket mellom faktisk og estimert verdi negativt (avviket er gitt ved faktisk omsetning minus estimert omsetning).

Tabell 5 Avvik mellom estimert årsumsetning og faktisk årsumsetning (BoF) i 2005. Her er årsums. estimert direkte uten å gå veien om månedsoms., og for en konstruert situasjon uten registerfeil. Avviket er gitt ved faktisk årsums. minus estimert verdi

Næring	NACE	Avvik (%) når vi bruker data	
		uten målefeil	med målefeil
Bergverksdrift	10, 12-14	2.6	-37.8
Næringsmiddelindustri	151-158	1.3	-5.0
Nytelsesmiddelindustri	159-160	0.1	12.2
Tekstil- og bekledningsindustri	17-19	1.9	-3.8
Trelast- og trevareindustri	20	3.2	2.3
Treforedling	21	8.7	7.4
Forlag og grafisk industri	22	-1.2	-1.2
Oljeraff. og kjemisk indu., unntatt kjemiske råvarer	23 og 242-247	0.1	5.2
Kjemiske råvarer	241	-3.2	-6.2
Gummivare- og plastindustri	25	0.4	-1.3
Mineralproduktindustri	26	2.6	-6.3
Metallindustri, unntatt ikke-jernholdige metaller	271-273, 275	-1.8	-4.7
Ikke-jernholdige metaller	274	-5.2	20.6
Metallvareindustri	28	-1.6	-2.1
Maskinindustri	29	1.7	-4.9
Elektroteknisk og optisk industri	30-33	-1.9	-9.0
Transportmiddelindustri	34,35111-3,6,7	-2.0	-4.5
Oljeplattformer	351114,5	-3.0	-18.4
Møbelindustri og annen industri	36-37	2.4	-13.9
Bergverksdrift og industri i alt		-0.1	-2.4

4. To alternative estimatorer for den månedlige omsetningen

I dette kapitlet skal vi se på to alternative estimatorer for den månedlige omsetningen. Som mange andre estimatorer innen økonomisk statistikk er disse estimatorene modellbaserte, og de bygger på en stratifisert ratemodell. Den ene estimatoren tilsvarer den vanlige stratifiserte rateestimator, mens den andre er en stratifisert ratebasert estimator.

Som x -variabel i ratemodellen bruker vi bedriftenes årsumsetning to år tilbake i tid. Som strata bruker vi næringene på publiseringnivå.

Vi gjør med andre ord ikke noen ytterligere stratifisering etter sysselsetting og omsetning, som er vanlig innen industristatistikk. Modellmessig ser det ikke ut til å være nødvendig med en slik stratifisering, og testing viser at det heller ikke gir bedre estimat. Men som vi skal komme tilbake til fins det avvikende bedrifter som vi skal skille ut og behandle spesielt.

Selv om vi bruker næring som stratum så kan vi beregne estimert omsetning på lavere nivå enn næring. Når vi bruker modellbasert estimator så predikerer vi månedsomsetningen til hver enkelt bedrift utenfor utvalget. Dermed kan vi lage en populasjonsfil med innrapportert månedsomsetning for bedriftene som er med i utvalget og predikert månedsomsetning for resten av bedriftene. Ut fra denne populasjonsfilen kan vi så beregne estimert omsetning for et hvilket som helst aggregeringsnivå helt enkelt ved å summere predikert/innrapportert verdi over de tilhørende bedriftene.

Når vi bruker modellbaserte estimatorer fins det alltid noen observasjoner som avviker fra den antatte modellen. Disse kalles gjerne avvikere eller ekstremverdier, og de kan ha veldig stor innflytelse på den estimerte verdien. For å redusere effekten av avvikerne er det vanlige å skille dem ut i et eget stratum, og så behandle dette stratimet som et fulltellingsstratum.

Også her har vi avvikende observasjoner som vi skal skille ut. Men i stedet for å plassere dem i et eget stratum skal vi ikke bruke den innrapporterte månedsomsetningen for disse bedriftene i det hele tatt i estimeringen. Vi skal helt enkelt behandle avvikerne som om de ikke var med i utvalget, dvs. vi predikerer nye omsetningstall for dem.

Grunnen til at vi gjør dette er at mange av de avvikende observasjonene viser seg å være store målefeil som generelt ikke oppveier hverandre slik de små tilfeldige målefeilene gjør. Det er derfor bedre å ikke bruke de innrapporterte omsetningstallene til disse bedriftene i det hele tatt, framfor å plassere dem i et eget stratum. Testing med 2005-tall viser at dette gjelder til tross for at ikke alle avvikende observasjoner som vi fjerner nødvendigvis er store målefeil.⁴

4.1. Modell

Det litt spesielle med modellen som presenteres her er at vi ikke ser bort fra at det er målefeil. Vi skal skille mellom den innrapporterte månedsomsetningen som har målefeil, og den faktiske månedsomsetningen som vi ønsker å estimere totalen til. (De innrapporterte månedsomsetningene er revidert, så det er snakk om gjenværende målefeil). Som notasjon for disse to størrelsen bruker vi

$w_{i,m}$ = innrapportert månedsomsetning for bedrift i , måned m

$y_{i,m}$ = faktisk månedsomsetning for bedrift i , måned m .

⁴ Det vi har funnet er at mange av bedriftene som har en avvikende månedsomsetning, for en eller flere måneder, har en stor målefeil i den innrapporterte årsumsetningen. Men vi kan aldri være sikre på om det faktisk er en stor målefeil i den aktuelle måneden vi ser på, selv om den er veldig avvikende.

Det er altså $w_{i,m}$ vi får fra bedriftene i utvalget, mens det er $y_{i,m}$ vi ønsker å estimere totalen til.

Vi antar at den faktiske månedsomsetningen $y_{i,m}$ kan modelleres med den stratifiserte ratemodellen. Dvs.

$$E[y_{i,m}] = \beta_{h,m} x_i \quad \text{og} \quad V(y_{i,m}) = \sigma_{h,m}^2 x_i \quad \text{for } i \in U_h,$$

der x_i = årsomsetning for bedrift i to år tidligere (fra BoF), og U_h er bedriftene i populasjonen som tilhører stratum h .

Sammenhengen mellom faktisk og innrapportert månedsomsetning har vi valgt å beskrive med

$$w_{i,m} = \lambda_{i,m} y_{i,m} + \varepsilon_{i,m},$$

der $\varepsilon_{i,m}$ er en stokastisk variabel med forventning 0 og varians $\tau_{h,m}^2 x_i$, og uavhengig av $y_{i,m}$, mens $\lambda_{i,m} > 0$ er en ukjent parameter.

Når $\lambda_{i,m} = 1$ har vi kun en tilfeldig målefeil i den innrapporterte månedsomsetningen. Dette er ment å beskrive tilfellene som vi i delkapittel 3.2 kalte små tilfeldige målefeil. Når $\lambda_{i,m} \neq 1$ har vi i tillegg til den tilfeldige målefeilen en systematisk målefeil. Dette er ment å beskrive tilfellene som vi kalte store målefeil, og som vi ikke har så mange av men som vi ikke kan regne med vil oppveie hverandre.

Fordi vi ikke kjenner den faktiske månedsomsetningen $y_{i,m}$ har vi ikke mulighet til å teste denne modellen. Valg av denne modellen har vi gjort ved først å tilpasse en modell for årsomsetningen (som vi kan gjøre fordi vi kjenner den faktiske årsomsetning fra BoF). Så har vi overført den modellen til månedsomsetningen, og håper at den passer også her. (Vår modellering av målefeilene kan nok diskuteres, men den ser ut til å være bra nok til å gi en rimelig estimator).

Hvis vi i tillegg antar at bedriftene er uavhengige av hverandre får vi at

$$E[w_{i,m}] = \lambda_{i,m} \beta_{h,m} x_i \quad \text{og} \quad V(w_{i,m}) = (\lambda_{i,m}^2 \sigma_{h,m}^2 + \tau_{h,m}^2) x_i \quad \text{for } i \in U_h.$$

Spesielt for bedriftene som bare har små tilfeldige målefeil ($\lambda_{i,m} = 1$) får vi at

$$(1) \quad E[w_{i,m}] = \beta_{h,m} x_i \quad \text{og} \quad V(w_{i,m}) = (\sigma_{h,m}^2 + \tau_{h,m}^2) x_i.$$

Dvs. at innrapportert månedsomsetning for disse bedriftene følger en stratifisert ratemodell.

Vi kan sjekke om denne modelleringen av innrapportert månedsomsetning virker rimelig (fordi vi kjenner $w_{i,m}$ for utvalget). Dvs. vi kan sjekke om de innrapporterte månedsomsetningene ser ut til å følge den stratifiserte ratemodellen (1), og om bedriftene som avviker fra modellen har store målefeil i innrapportert årsomsetning. For 2005-dataene stemmer dette nokså bra. Ratemodellen ser ut til å passe bra for de fleste næringene, og de fleste av avvikene har store målefeil i innrapportert årsomsetning. Men det fins også noen avvikere som ikke har store målefeil i innrapportert årsomsetning.

Som definisjon på en avvikere har vi brukt regresjonsdiagnostikk. For aktuell måned beregnes bedriftenes studentiserte residualer (R) og innflytelse (DFFITs) i forhold til ratemodellen

$E[w_{i,m}] = \beta_{h,m} x_i$ og $V(w_{i,m}) = (\sigma_{h,m}^2 + \tau_{h,m}^2) x_i$. Dette gjøres innen hver næring (stratum). For transportmiddelindustrien defineres så en bedrift som avviker hvis $|R| \notin (-1, 2.5)$ og $|DFFITS| > \frac{2}{\sqrt{n_h}}$, der n_h er antall bedrifter i næringen (stratumet). For resten av næringene defineres en bedrift som avviker hvis $|R| \notin (-2, 2.5)$ og $|DFFITS| > \frac{2}{\sqrt{n_h}}$.

Grensene for når en bedrift skal telle som avviker er valgt ut fra ønsket om å få med flest mulig av de store målefeilene, uten å få med for mange av de små målefeilene. (Analyse med 2005-dataene viste at det er hensiktsmessig å velge litt andre grenser for transportmiddelindustrien enn for de andre næringene).

Merk at definisjonen gjelder per måned. Antall avvikere, og hvilke bedrifter det er, vil altså variere fra måned til måned. I 2005 var det mellom 44 og 58 avvikere per måned. De fleste av disse hadde stor målefeil i innrapportert årsomsetning, men ikke alle.

Som et alternativ til å skille ut de store målefeilene har vi gjort forsøk på å rette opp feilene ved å estimere $\lambda_{i,m}$. Vi estimerte $\lambda_{i,m}$ med den samme målefeilen som bedriften hadde for to år siden (forholdet mellom innrapportert årsomsetning for to år siden og x_i). Men resultatet av dette var ikke bra.

4.2. Estimator A

Estimator A er den stratifiserte rateestimatoren der vi kun bruker de innrapporterte månedsomsetningene som ikke er avvikende. Bedriftene som ikke er med i utvalget, og bedriftene som er avvikende, får predikert månedsomsetningen med

$$\hat{\beta}_{h,m} x_i,$$

der

$$\hat{\beta}_{h,m} = \frac{\sum_{i \in s_{h,m}^-} w_{i,m}}{\sum_{i \in s_{h,m}^-} x_i}$$

og $s_{h,m}^-$ er bedriftene i utvalget til stratum h som ikke er identifisert som avviker. (Hvilke bedrifter som regnes som avvikere kan variere fra måned til måned, derfor markerer vi utvalget med m).

I stratum h estimeres dermed den faktiske månedsomsetningen

$$Y_{h,m} = \sum_{i \in U_h} y_{i,m}$$

med estimatoren

$$\begin{aligned}
\hat{Y}_{h,m} &= \sum_{i \in s_{h,m}^-} w_{i,m} + \sum_{i \in U_h \setminus s_{h,m}^-} \hat{\beta}_{h,m} x_i \\
&= \frac{\sum_{i \in s_{h,m}^-} w_{i,m}}{\sum_{i \in s_{h,m}^-} x_i} X_h \\
&= \hat{\beta}_{h,m} X_h,
\end{aligned}$$

der $X_h = \sum_{i \in U_h} x_i$ er årsomsetingen i stratum h to år tilbake i tid.

Denne estimatoren kan vi beregne med SAS-applikasjonen STRUKTUR. Men variasjonskoeffisienten og konfidensintervallet vi får med STRUKTUR tar ikke hensyn til at det er målefeil i dataene. Vi skal derfor gjøre en modellbasert analyse av estimator A for å finne en variasjonskoeffisient og et konfidensintervall som inkluderer usikkerheten som følger av målefeilene.

4.2.1. Modellbasert analyse av estimator A

Analysen gjøres under antagelsen at vi vet hvilke bedrifter som kun har tilfeldig målefeil ($\lambda_{i,m} = 1$) og hvilke som har systematiske målefeil ($\lambda_{i,m} \neq 1$), slik at $s_{h,m}^-$ er de bedriftene i utvalget som kun har en tilfeldig målefeil.⁵ Dermed er $s_{h,m}^-$ en ustokastisk mengde, og $\lambda_{i,m} = 1$ for $i \in s_{h,m}^-$. Videre skal vi se bort fra frafall og registerfeil (men vi ser altså ikke bort fra målefeil).

Under disse forutsetningene får vi at

$$\begin{aligned}
E[\hat{Y}_{h,m} - Y_{h,m}] &= E[\hat{\beta}_{h,m} X_h] - E\left[\sum_{i \in U_h} y_{i,m}\right] \\
&= \beta_{h,m} X_h - \beta_{h,m} X_h \\
&= 0.
\end{aligned}$$

Dvs. estimator A er forventningsrett for den faktiske månedsomsetningen i stratum h . (Det kan vises at dette også gjelder for andre aggregeringsnivå en stratumnivå).

Med en del regning kan det vises at variansen til prediksjonsfeilen $\hat{Y}_{h,m} - Y_{h,m}$ kan skrives som

$$\begin{aligned}
V(\hat{Y}_{h,m} - Y_{h,m}) &= V\left(\sum_{i \in s_{h,m}^-} w_{i,m} + \frac{\sum_{i \in s_{h,m}^-} w_{i,m}}{\sum_{i \in s_{h,m}^-} x_i} \sum_{i \in U_h \setminus s_{h,m}^-} x_i - \sum_{i \in U_h} y_{i,m}\right) \\
&= (\sigma_{h,m}^2 + \tau_{h,m}^2) \left\{ \frac{\left(\sum_{U_h \setminus s_{h,m}^-} x_i\right)^2}{\sum_{s_{h,m}^-} x_i} + \sum_{U_h \setminus s_{h,m}^-} x_i \right\} + \tau_{h,m}^2 X_h.
\end{aligned}$$

⁵ Denne antagelsen er urealistisk. Men også i SAS-applikasjonen STRUKTUR gjøres tilsvarende antagelser når det fins avvikende observasjoner som behandles spesielt.

Denne variansen er større enn den vi får fra SAS-applikasjonen STRUKTUR. Med STRUKTUR får vi beregnet variansen

$$\begin{aligned} V(\hat{Y}_{h,m} - W_{h,m}) &= V\left(\sum_{i \in s_{h,m}^-} w_{i,m} + \frac{\sum_{i \in s_{h,m}^-} w_{i,m}}{\sum_{i \in s_{h,m}^-} x_i} \sum_{i \in U_h \setminus s_{h,m}^-} x_i - \sum_{i \in U_h} w_{i,m}\right) \\ &= (\sigma_{h,m}^2 + \tau_{h,m}^2) \left\{ \frac{\left(\sum_{U_h \setminus s_{h,m}^-} x_i\right)^2}{\sum_{s_{h,m}^-} x_i} + \sum_{U_h \setminus s_{h,m}^-} x_i \right\} \end{aligned}$$

som ikke tar hensyn til at det er målefeil i dataene.

Vi har at $E[w_{i,m}] = \beta_{h,m} x_i$ og $V(w_{i,m}) = (\sigma_{h,m}^2 + \tau_{h,m}^2) x_i$ for $i \in s_{h,m}^-$, så vi estimerer $\sigma_{h,m}^2 + \tau_{h,m}^2$ med

$$\widehat{\sigma_{h,m}^2 + \tau_{h,m}^2} = \frac{1}{n_{h,m} - 1} \sum_{i \in s_{h,m}^-} \frac{(w_{i,m} - \hat{\beta}_{h,m} x_i)^2}{x_i},$$

der $n_{h,m}$ er antall bedrifter i $s_{h,m}^-$.

Vi får ikke estimert $\tau_{h,m}^2$ på tilsvarende vanlig måte, så vi bruker $\widehat{\sigma_{h,m}^2 + \tau_{h,m}^2}$ som en konservativ estimator for $\tau_{h,m}^2$. Den estimerte prediksjonsvariansen blir dermed

$$\begin{aligned} \hat{V}(\hat{Y}_{h,m} - Y_{h,m}) &= \left(\widehat{\sigma_{h,m}^2 + \tau_{h,m}^2}\right) \left\{ \frac{\left(\sum_{U_h \setminus s_{h,m}^-} x_i\right)^2}{\sum_{s_{h,m}^-} x_i} + \sum_{U_h \setminus s_{h,m}^-} x_i \right\} + \left(\widehat{\sigma_{h,m}^2 + \tau_{h,m}^2}\right) X_h \\ &= \left(\widehat{\sigma_{h,m}^2 + \tau_{h,m}^2}\right) \left\{ \frac{\left(\sum_{U_h \setminus s_{h,m}^-} x_i\right)^2}{\sum_{s_{h,m}^-} x_i} + \sum_{U_h \setminus s_{h,m}^-} x_i + X_h \right\}, \end{aligned}$$

og den estimerte variasjonskoeffisienten blir

$$\widehat{cv} = \frac{\sqrt{\hat{V}(\hat{Y}_{h,m} - Y_{h,m})}}{\hat{Y}_{h,m}} \cdot 100.$$

For å kunne bruke $\hat{Y}_{h,m} \pm 1.96 \sqrt{\hat{V}(\hat{Y}_{h,m} - Y_{h,m})}$ som et tilnærmet 95% konfidensintervall for $Y_{h,m}$ antar vi at $Y_{i,m}$ og $\varepsilon_{i,m}$ er normalfordelte. Da er nemlig

$$\hat{Y}_{h,m} - Y_{h,m} = \sum_{i \in \bar{s}_{h,m}} \left(\frac{X_h}{\sum_{i \in \bar{s}_{h,m}} x_i} - 1 \right) y_{i,m} + \frac{X_h}{\sum_{i \in \bar{s}_{h,m}} x_i} \sum_{i \in \bar{s}_{h,m}} \varepsilon_{i,m} - \sum_{i \in U_h \setminus \bar{s}_{h,m}} y_{i,m}$$

også normalfordelt (med forventning lik 0) slik at $\hat{Y}_{h,m} \pm 1.96\sqrt{V(\hat{Y}_{h,m} - Y_{h,m})}$ er et 95%

konfidensintervall for $Y_{h,m}$. Vi putter så inn estimert prediksjonsvarians for $V(\hat{Y}_{h,m} - Y_{h,m})$, og bruker

$$\left(\hat{Y}_{h,m} - 1.96\sqrt{\hat{V}(\hat{Y}_{h,m} - Y_{h,m})} \quad , \quad \hat{Y}_{h,m} + 1.96\sqrt{\hat{V}(\hat{Y}_{h,m} - Y_{h,m})} \right)$$

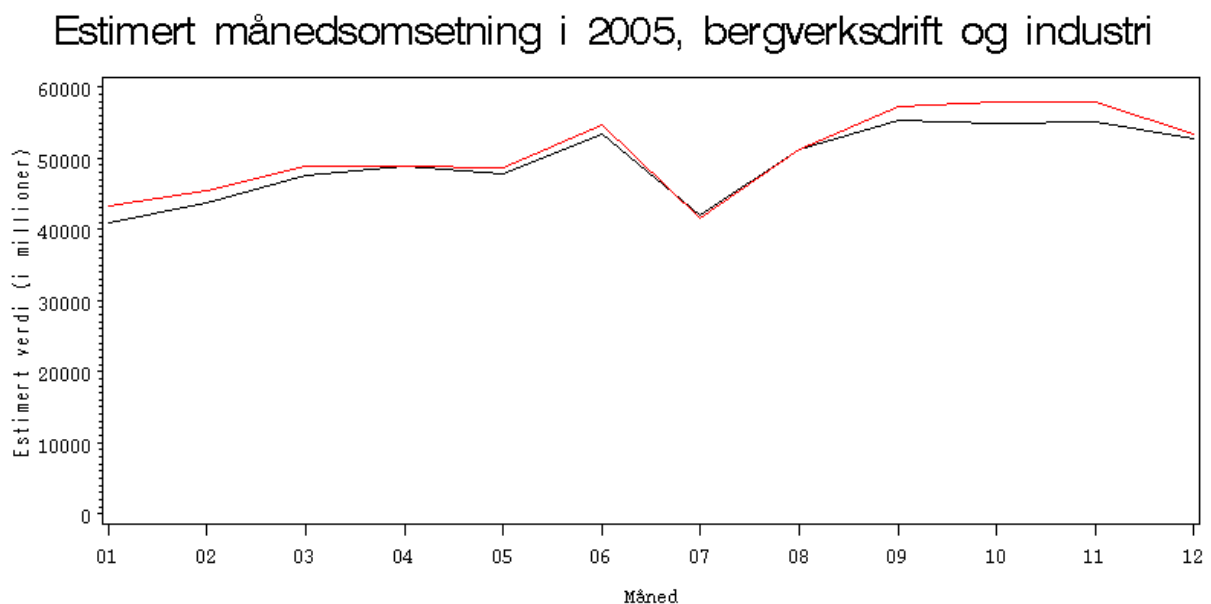
som et tilnærmet 95% konfidensintervall for den månedlige omsetningen i stratum h .

4.2.2. Talleksempel

Her skal vi se hvordan estimert verdi blir, og hvor stor usikkerhet vi har, når vi estimerer månedsomsetningen i 2005 med estimator A.

Vi starter med å sammenligne estimatene vi får fra estimator A med estimatene vi får fra gjeldende estimator i Omsetningsstatistikken. Ser vi på bergverksdrift og industri i alt, er estimatene fra de to estimatorene ikke så veldig forskjellige. Utviklingen gjennom året er så å si den samme for begge estimatorene, men estimator A gir stort sett noe større estimat (se figur 3).

På publiseringsnivå har vi større forskjeller. Men for mange av næringene gir de to estimatorene grovt sett samme utvikling gjennom året, selv om det nivåmessig kan være en viss forskjell. For en god del av næringene ligger månedsestimatene med estimator A, for alle 12 månedene, konsekvent over eller under estimatene med gjeldende estimator, og i noen tilfeller er det veldig stor forskjell på estimatene. (Se vedlegg C for flere figurer).



Figur 3 Estimert månedsomsetning innen bergverksdrift og industri i 2005. Den røde linjen er estimert verdi med estimator A, den sorte linjen er estimert verdi med gjeldende estimeringsopplegg i Omsetningsstatistikken.

Beregner vi den estimerte årsomsetningen (sum av estimerte månedsomsetninger) og sammenligner mot faktisk årsomsetning (fra BoF), finner vi at vi treffer veldig bra for bergverksdrift og industri i alt. Den estimerte årsomsetningen er på 609 milliarder kr mens den faktiske årsomsetningen er i underkant av 627 milliarder. Dvs. vi underestimerer årsomsetningen i 2005 med ca. 18 milliarder, og det tilsvarer et avvik på 2,8% av faktisk årsomsetning. Til sammenligning underestimerer vi med 5,2% når vi bruker gjeldende estimator.

På publiseringsnivå varierer resultatet fra næring til næring (se tabell 6). For en del næringer treffer vi veldig bra med estimator A, og for over halvparten av næringene bommer vi med mindre enn 6%, noe vi må kunne si er bra. For de resterende næringene bommer vi mer. Men det er bare for to næringer at vi bommer med mer enn 10%, og det er for oljeraffinering og kjemisk industri unntatt kjemiske råvarer (næring 23, 242-247) med et avvik på 10,6%, og ikke-jernholdige metaller (næring 274) med et avvik på 15,0%. Til sammenligning bommer vi med mer enn 10% for 6 næringer når vi bruker gjeldende estimator, og da har vi på det verste et avvik på 23,0%.

Basert på disse resultatene ser det ut til at estimator A totalt sett er noe bedre enn gjeldende estimator.

Tabell 6 Estimert årsomsetning innen bergverksdrift og industri i 2005, fordelt på publiseringsnivå. Avviket er gitt ved faktisk årsomsetning (BoF) minus estimert verdi

Næring	Estimert årsomsetning med estimator A (i millioner kr)	Avvik mot faktisk årsomsetning (%)	
		med estimator A	med gjeldende estimator
10, 12-14	9 872	-4.7	-18.7
151-158	125 049	1.8	23.0
159-160	19 211	8.8	5.0
17-19	6 193	5.7	-4.8
20	22 924	3.7	3.6
21	18 452	4.3	14.2
22	40 092	0.3	7.6
23 og 242-247	60 563	10.6	10.0
241	31 358	0.0	-6.6
25	9 598	-6.2	1.6
26	21 461	-4.8	-3.5
271-273, 275	14 457	-7.1	-6.0
274	34 063	15.0	-6.7
28	25 854	5.5	-2.6
29	40 186	5.8	-0.7
30-33	39 694	-3.5	-15.5
34,35111-3,6,7	30 623	7.2	15.9
351114,5	42 071	-9.4	-1.2
36-37	17 301	-1.8	-3.8
Bergverksdrift og industri i alt	609 023	2.8	5.2

Usikkerheten til estimator A, målt ved den estimerte variasjonskoeffisienten, varierer fra næring til næring og fra måned til måned (se tabell 7). Noe av variasjonen vi ser fra måned til måned skyldes nok at den estimerte månedsomsetningen varierer, og ikke bare at variansen har endret seg.

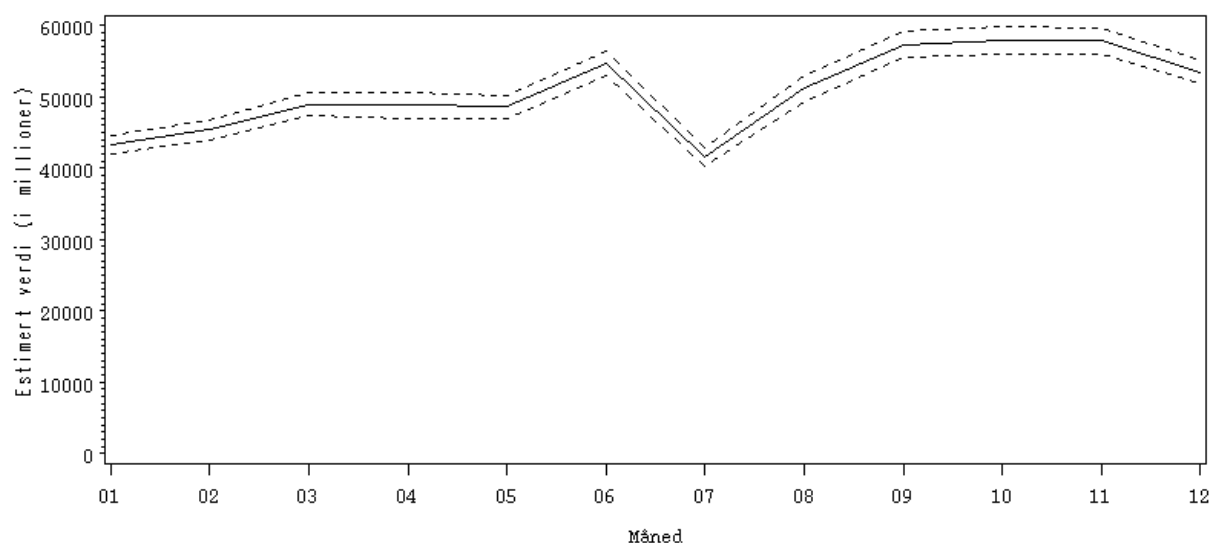
For bergverksdrift og industri i alt ligger de estimerte variasjonskoeffisientene mellom 1,5% og 1,9%, noe som tyder på veldig liten usikkerhet. Også for forlag og grafisk industri (næring 22) har vi veldig liten usikkerhet. Som en følge av liten usikkerhet får vi smale konfidensintervall for disse nivåene (se figur 4 og vedlegg D).

Noen få næringer, som f.eks. næringsmiddelindustrien (næring 151-158), har en estimert variasjonskoeffisient som er mindre enn 5% alle månedene. Dette tyder på at usikkerheten til disse næringene er stabilt liten gjennom hele året. Det fins også noen næringer som har en estimerte variasjonskoeffisient på under 5% for noen av månedene, mens variasjonskoeffisienten for resten av året er en del større. Resten av næringene har stort sett middels til stor usikkerhet. Det fins også noen tilfeller hvor den estimerte variasjonskoeffisienten er større enn 10%, og det er bergverksdrift (næring 10, 12-14) som har de største verdiene. Denne næringen har estimerte variasjonskoeffisienter på drøye 18% i både august, september og oktober, og i juli er den på 16,1%. Dette tyder på ekstremt stor usikkerhet disse månedene.

Tabell 7 Den estimerte variasjonskoeffisienten til estimator A i 2005, på publiseringnivå

Næring	Jan.	Feb.	Mars	April	Mai	Juni	Juli	Aug.	Sep.	Okt.	Nov.	Des.
10, 12-14	7.7	8.2	8.3	7.6	5.5	8.4	16.1	18.3	18.4	18.9	11.3	8.0
151-158	3.9	3.9	4.1	4.4	4.2	4.1	4.5	3.8	4.0	4.6	4.0	3.9
159-160	7.9	10.0	14.5	13.0	7.0	6.6	5.4	16.0	4.9	5.8	3.0	4.1
17-19	6.0	6.6	10.0	10.0	8.7	7.5	9.5	10.2	8.0	7.5	7.4	7.2
20	4.2	4.4	3.7	2.3	4.2	4.1	5.3	3.5	2.7	3.2	4.8	3.5
21	7.0	7.0	3.2	4.2	3.9	4.2	4.7	4.1	5.1	5.4	5.0	5.0
22	1.7	1.7	1.4	1.6	1.7	1.7	2.6	1.7	1.6	1.5	1.9	1.7
23 og 242-247	5.9	5.7	6.7	12.5	8.9	5.7	5.2	6.9	6.1	6.2	6.5	5.7
241	5.8	5.1	9.5	11.7	9.2	3.9	6.0	11.4	9.0	9.3	9.5	7.6
25	4.0	5.0	4.7	4.8	5.1	5.5	6.0	5.5	6.2	6.0	5.2	5.3
26	4.2	3.9	7.1	8.2	7.3	5.4	5.5	10.8	5.5	4.9	4.7	5.3
271-273, 275	5.6	4.9	4.4	5.7	5.1	3.5	5.6	14.4	4.6	5.2	6.0	4.7
274	6.5	7.0	4.4	9.9	6.2	9.6	4.2	5.5	9.4	7.4	7.7	6.7
28	3.1	3.4	3.7	3.6	3.2	3.5	4.3	3.4	3.5	2.9	2.7	3.8
29	5.8	5.5	5.0	5.6	6.6	5.0	8.0	5.1	3.9	4.7	4.5	4.6
30-33	7.1	8.1	8.3	7.4	7.4	6.8	7.2	7.8	6.8	5.3	6.7	5.4
34,35111-3,6,7	4.9	5.6	5.7	6.6	8.1	7.7	8.6	9.2	8.0	10.0	7.9	7.0
351114,5	13.9	11.0	11.1	9.6	9.9	9.4	10.4	11.0	9.3	8.5	8.4	8.7
36-37	3.5	3.3	3.7	4.3	4.2	4.5	4.8	4.0	4.4	4.0	3.7	3.7
Bergverksdrift og industri i alt	1.5	1.5	1.7	1.9	1.7	1.6	1.7	1.9	1.6	1.7	1.6	1.5

95% KI for månedsomsetning i 2005, bergverksdrift og industri



Figur 4 Et tilnærmet 95% konfidensintervall for den månedlige omsetningen innen bergverksdrift og industri i 2005 (stiplede linjer). Den heltrukne linjen er estimert månedsomsetning (estimator A).

Det må huskes at de estimerte variasjonskoeffisientene vi presenterer her, inkluderer usikkerheten vi har på grunn av målefeilene. Velger vi å se bort fra at det er målefeil i dataene, blir den målte usikkerheten mindre. Som eks. skal vi se hvor stor den estimerte variasjonskoeffisienten blir for metallvareindustrien og oljeplattformer når vi ser bort fra at det er målefeil i dataene. For metallvareindustrien varierer den estimerte variasjonskoeffisienten fra 2,7% til 4,3% når vi tar høyde for målefeilene. Ser vi bort fra målefeilene får vi en estimert variasjonskoeffisient som varierer fra 1,9% til 2,9%. For oljeplattformer varierer den estimerte variasjonskoeffisienten fra 8,4% til 13,9% når vi tar høyde for målefeilene, og fra 4,7% til 8,6% når vi ser bort fra målefeilene.

4.3. Estimator B

I estimator B bruker vi bedriftenes innrapporterte årsomsetning fra to år tilbake i tid til estimering av $\beta_{h,m}$.

De fleste av bedriftene i utvalg var også med i utvalget for to år siden. For disse bedriftene har vi

$$z_i = \text{innrapporter årsomsetning for bedrift } i \text{ to år tidligere (sum av innrapportert månedsomsetning)}.$$

Årsomsetningen z_i gjelder altså samme år som x_i (x_i er faktisk årsomsetning fra BoF, mens z_i er innrapportert omsetning som har målefeil).

Hadde vi ikke hatt målefeil i de innrapporterte månedsomsetningene, dvs. hadde vi fått $y_{i,m}$ i stedet for $w_{i,m}$ fra bedriftene i utvalget, ville vi estimert $\beta_{h,m}$ med $\sum y_{i,m} / \sum x_i$. Det vi nå håper på er at forholdet mellom de innrapporterte omsetningene $w_{i,m}$ og z_i er lik forholdet mellom de faktiske omsetningene $y_{i,m}$ og x_i , slik at $\sum w_{i,m} / \sum z_i \approx \sum y_{i,m} / \sum x_i$. I så fall kan vi estimere $\beta_{h,m}$ med $\sum w_{i,m} / \sum z_i$.

Fordi $y_{i,m}$ er ukjent får vi ikke sjekket om tilnærmingen holder. Men vi har sjekket tilsvarende tilnærming for årsomsetningene, og det viste for 2005 at $\sum w_i / \sum z_i$ er veldig lik $\sum y_i / \sum x_i$ (når vi fjerner noen avvikende observasjoner). Det eneste unntaket er for transportmiddelindustrien, der $\sum y_i / \sum x_i$ er nesten dobbel så stor som $\sum w_i / \sum z_i$.

Vi håper at det samme gjelder for månedsomsetningene. Unntatt for transportmiddelindustrien estimerer vi derfor $\beta_{h,m}$ med

$$(2) \quad \tilde{\beta}_{h,m} = \frac{\sum_{i \in s_{h,m}^*} w_{i,m}}{\sum_{i \in s_{h,m}^*} z_i},$$

der $s_{h,m}^*$ er de bedriftene i utvalget til stratum h som også var med i utvalget for to år siden (og som ikke avviker i forhold til modellen $E[w_{i,m}] = \beta_{h,m} z_i$ og $V(w_{i,m}) = \sigma_{h,m}^2 z_i$).

For transportmiddelindustrien estimerer vi $\beta_{h,m}$ på samme måte som i estimator A, dvs. $\tilde{\beta}_{h,m} = \hat{\beta}_{h,m}$. (Vi har prøvd med å bruke (2) også for denne næringen, men da fikk vi et veldig dårlig resultat).

Bedrifter som ikke er med i utvalget, og bedrifter som er avvikende ifølge definisjonen i delkapittel 4.1, får nå predikert månedsomsetningen med

$$\tilde{\beta}_{h,m} x_i,$$

og for stratum h kan estimator B skrives som

$$\begin{aligned} \tilde{Y}_{h,m} &= \sum_{i \in \bar{s}_{h,m}} w_{i,m} + \sum_{i \in U_h \setminus \bar{s}_{h,m}} \tilde{\beta}_{h,m} x_i \\ &= \sum_{i \in \bar{s}_{h,m}} w_{i,m} + \tilde{\beta}_{h,m} \sum_{i \in U_h \setminus \bar{s}_{h,m}} x_i. \end{aligned}$$

For transportmiddelindustrien blir estimator B lik estimator A, dvs. $\tilde{Y}_{h,m} = \hat{Y}_{h,m}$.

4.3.1. Modellbasert analyse av estimator B

Analysen gjelder for alle strata unntatt transportmiddelindustrien. For transportmiddelindustrien har vi at estimator B er lik estimator A, og da er det analysen i delkapittel 4.2.1 som gjelder.

I tillegg til antagelsene vi gjorde under analysen av estimator A, skal vi anta at

$\sum_{s_{h,m}^*} (y_{i,m} + \varepsilon_{i,m}) / \sum_{s_{h,m}^*} x_i$ gir en god tilnærming av $\sum_{s_{h,m}^*} w_{i,m} / \sum_{s_{h,m}^*} z_i$ innen hvert stratum. Dvs. at

$$(3) \quad \frac{\sum_{s_{h,m}^*} w_{i,m}}{\sum_{s_{h,m}^*} z_i} \approx \frac{\sum_{s_{h,m}^*} (y_{i,m} + \varepsilon_{i,m})}{\sum_{s_{h,m}^*} x_i}.$$

Grunnen til at vi gjør denne antagelse er at vi ønsker oss et tilnærmet uttrykk for prediksjonsvariansen som vi kan estimere. Den eksakte prediksjonsvariansen klarer vi ikke å estimere fordi den avhenger av $\lambda_{i,m}$ som er ukjent.

Vi kunne alternativt valgt å bruke tilnærmingen $\sum_{s_{h,m}^*} w_{i,m} / \sum_{s_{h,m}^*} z_i \approx \sum_{s_{h,m}^*} y_{i,m} / \sum_{s_{h,m}^*} x_i$, som vi har sjekket at holder veldig bra for årstallene. Men i tilfelle den ikke holder like bra for månedsomsetningen har vi valgt å bruke tilnærming (3). (Den estimerte prediksjonsvariansen blir noe mindre med den alternative tilnærmingen enn med tilnærming (3)).

Med tilnærming (3) får vi

$$\tilde{Y}_{h,m} \approx \sum_{i \in \bar{s}_{h,m}} w_{i,m} + \frac{\sum_{s_{h,m}^*} (y_{i,m} + \varepsilon_{i,m})}{\sum_{s_{h,m}^*} x_i} \sum_{i \in U_h \setminus \bar{s}_{h,m}} x_i.$$

Dermed er

$$\begin{aligned} E[\tilde{Y}_{h,m} - Y_{h,m}] &\approx E\left[\sum_{i \in \bar{s}_{h,m}} w_{i,m} + \frac{\sum_{s_{h,m}^*} (y_{i,m} + \varepsilon_{i,m})}{\sum_{s_{h,m}^*} x_i} \sum_{i \in U_h \setminus \bar{s}_{h,m}} x_i\right] - E\left[\sum_{i \in U_h} y_{i,m}\right] \\ &= \beta_{h,m} \sum_{i \in \bar{s}_{h,m}} x_i + \frac{\beta_h \sum_{s_{h,m}^*} x_i}{\sum_{s_{h,m}^*} x_i} \sum_{i \in U_h \setminus \bar{s}_{h,m}} x_i - \beta_{h,m} \sum_{i \in U_h} x_i \\ &= \beta_{h,m} \sum_{i \in U_h} x_i - \beta_{h,m} \sum_{i \in U_h} x_i \\ &= 0 \end{aligned}$$

Dvs. estimator B er tilnærmet forventningsrett.

Men en del regning får vi at prediksjonsvariansen til $\tilde{Y}_{h,m}$ kan tilnærmes ved

$$\begin{aligned} (4) \quad V(\tilde{Y}_{h,m} - Y_{h,m}) &\approx V\left(\sum_{i \in \bar{s}_{h,m}} w_{i,m} + \frac{\sum_{s_{h,m}^*} (y_{i,m} + \varepsilon_{i,m})}{\sum_{s_{h,m}^*} x_i} \sum_{i \in U_h \setminus \bar{s}_{h,m}} x_i - \sum_{i \in U_h} y_{i,m}\right) \\ &= (\sigma_{h,m}^2 + \tau_{h,m}^2) \left\{ \frac{\left(\sum_{U_h \setminus \bar{s}_{h,m}} x_i\right)^2}{\sum_{s_{h,m}^*} x_i} + \sum_{i \in U_h \setminus \bar{s}_{h,m}} x_i - 2 \frac{\sum_{U_h \setminus \bar{s}_{h,m}} x_i}{\sum_{s_{h,m}^*} x_i} \sum_{i \in \bar{s}_{h,m} \setminus \bar{s}_{h,m}} x_i \right\} + \tau_{h,m}^2 X_h. \end{aligned}$$

Estimerer $\sigma_h^2 + \tau_h^2$ med

$$\widehat{\sigma_{h,m}^2 + \tau_{h,m}^2} = \frac{1}{n_{h,m} - 1} \sum_{i \in \bar{s}_{h,m}} \frac{(w_{i,m} - \hat{\beta}_{h,m} x_i)^2}{x_i},$$

og bruker $\widehat{\sigma_{h,m}^2 + \tau_{h,m}^2}$ som en konservativ estimator for $\tau_{h,m}^2$. (Dette er samme estimatorene som vi brukte i prediksjonsvariansen til estimator A).

Den estimerte prediksjonsvariansen blir dermed

$$\begin{aligned} \hat{V}(\tilde{Y}_{h,m} - Y_{h,m}) &= \left(\widehat{\sigma_{h,m}^2 + \tau_{h,m}^2} \right) \left\{ \frac{\left(\sum_{U_h \setminus s_{h,m}^-} x_i \right)^2}{\sum_{s_{h,m}^*} x_i} + \sum_{i \in U_h \setminus s_{h,m}^-} x_i - 2 \frac{\sum_{U_h \setminus s_{h,m}^-} x_i}{\sum_{s_{h,m}^*} x_i} \sum_{i \in s_{h,m}^* \setminus s_{h,m}^-} x_i \right\} + \left(\widehat{\sigma_{h,m}^2 + \tau_{h,m}^2} \right) X_h \\ &= \left(\widehat{\sigma_{h,m}^2 + \tau_{h,m}^2} \right) \left\{ \frac{\left(\sum_{U_h \setminus s_{h,m}^-} x_i \right)^2}{\sum_{s_{h,m}^*} x_i} + \sum_{i \in U_h \setminus s_{h,m}^-} x_i - 2 \frac{\sum_{U_h \setminus s_{h,m}^-} x_i}{\sum_{s_{h,m}^*} x_i} \sum_{i \in s_{h,m}^* \setminus s_{h,m}^-} x_i + X_h \right\}, \end{aligned}$$

og den estimerte variasjonskoeffisienten blir

$$\widehat{cv} = \frac{\sqrt{\hat{V}(\tilde{Y}_{h,m} - Y_{h,m})}}{\tilde{Y}_{h,m}} \cdot 100.$$

For å lage et tilnærmet 95% konfidensintervall for $Y_{h,m}$ antar vi at $Y_{i,m}$ og $\varepsilon_{i,m}$ er normalfordelte.

Da er $\tilde{Y}_{h,m} - Y_{h,m}$ tilnærmet normalfordelt, og $\tilde{Y}_{h,m} \pm 1.96\sqrt{\hat{V}}$ er et tilnærmet 95% konfidensintervall for $Y_{h,m}$. Her er \hat{V} den tilnærmede prediksjonsvariansen til $\tilde{Y}_{h,m}$ gitt ved (4). Vi putter så inn den estimerte prediksjonsvariansen, og bruker

$$\left(\tilde{Y}_{h,m} - 1.96\sqrt{\hat{V}(\tilde{Y}_{h,m} - Y_{h,m})}, \tilde{Y}_{h,m} + 1.96\sqrt{\hat{V}(\tilde{Y}_{h,m} - Y_{h,m})} \right)$$

som et tilnærmet 95% konfidensintervall for den månedlige omsetningen i stratum h .

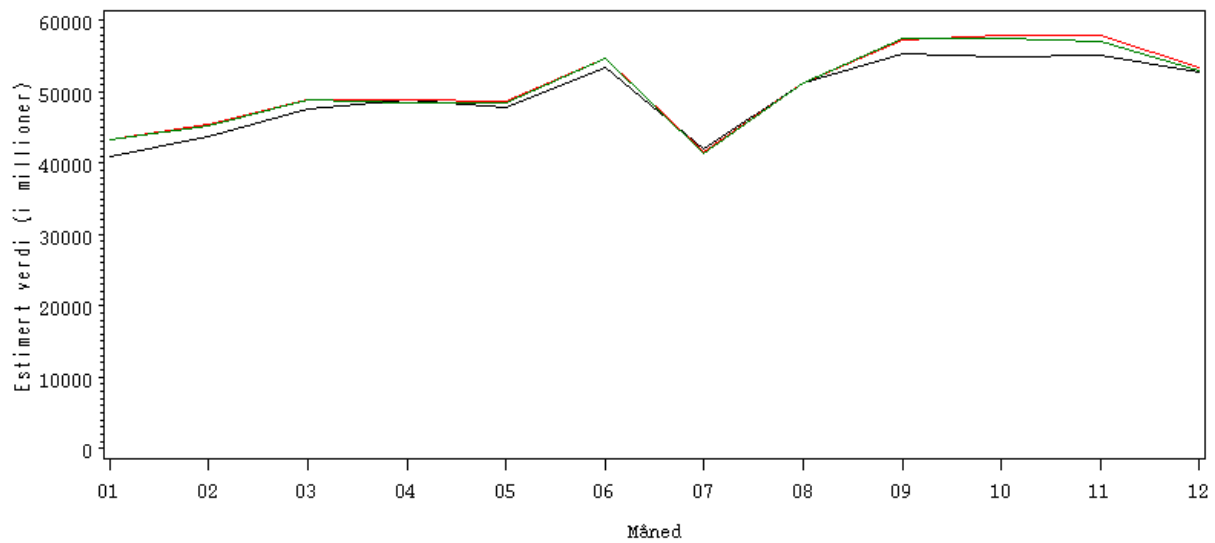
Merk at variasjonskoeffisienten og konfidensintervallet inkluderer usikkerheten vi har på grunn av målefeilene.

4.3.2. Talleksempel

Nå skal vi se hvordan de estimerte verdiene for 2005 blir med estimator B, og hvor stor usikkerhet vi får med denne estimatoren.

Sammenligning av de estimerte månedsomsetningene vi får fra estimator B, med de estimerte verdiene vi får fra estimator A, viser at de to estimatorene stort sett gir veldig like estimat. Dette gjelder både for bergverksdrift og industri i alt og for de fleste av næringene (se figur 5 og vedlegg C). Dermed får vi stort sett samme resultat ved sammenligning av estimatene fra estimator B og gjeldende estimator, som vi fikk ved sammenligningen av estimatene fra estimator A og gjeldende estimator.

Estimert månedsomsetning i 2005, bergverksdrift og industri



Figur 5 Estimert månedsomsetning innen bergverksdrift og industri i 2005. Den grønne linjen er estimert verdi med estimator B, den røde linjen er estimert verdi med estimator A, og den sorte linjen er estimert verdi med gjeldende estimeringsopplegg i Omsetningsstatistikken.

Fordi de estimerte månedsomsetningene blir nokså like med estimator A og estimator B, blir også de estimerte årstallene nokså like. Dette igjen betyr at vi bommer omtrent like mye på den faktiske årsumsetningen (BoF) med begge estimatorene. Men som vi kan se av tabell 8 er det litt forskjell. For bergverksdrift og industri i alt f.eks. bommer vi med 3.2% med estimator B mot 2.8% med estimator A. Det er ikke store forskjellen, men vi treffer altså litt bedre med estimator A.

Det er større forskjell for en del av næringene på publiseringsnivå. Eksempelvis har vi et avvik på 4,8% med estimator B, mot 8,8% med estimator A for nytelsesmiddelindustrien (næring 159-160), og for møbelindustri og annen industri (næring 36-37) er avviket på 4,8% med estimator B mot -1,8% med estimator A. Totalt sett ser det ut til at vi treffer litt bedre med estimator B enn med estimator A på publiseringsnivå.

Sammenligner vi hvor godt vi treffer den faktiske årsumsetningen med estimator B, med hvor godt vi treffer med gjeldende estimator, ender vi opp med samme konklusjon som for estimator A. Nemlig at estimator B totalt sett ser ut til å være noe bedre enn gjeldende estimator.

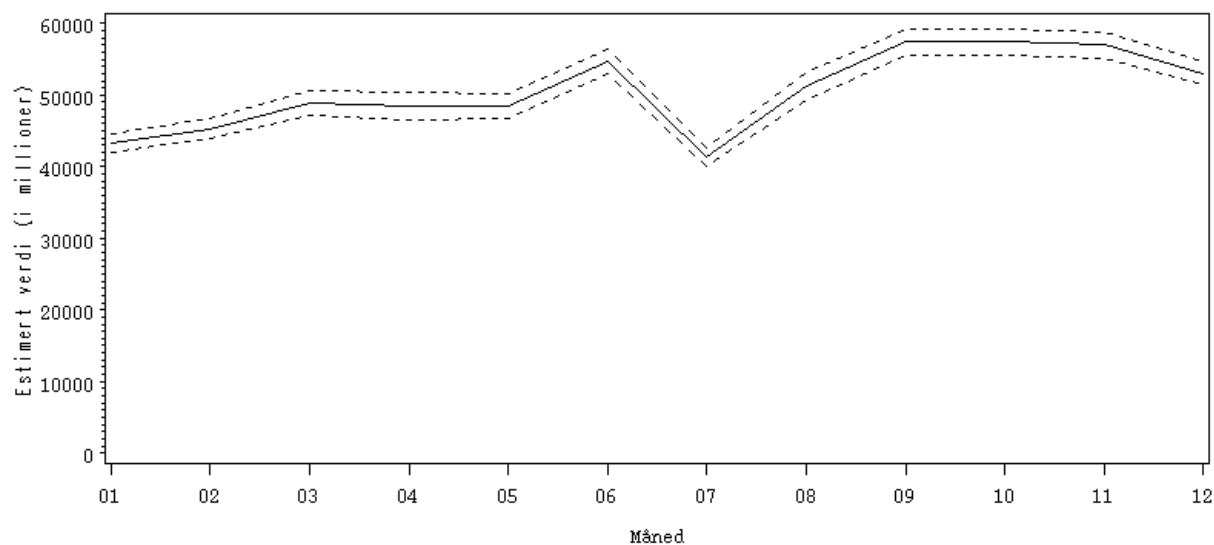
Usikkerheten til estimator B, målt ved den estimerte variasjonskoeffisienten, er generelt nokså lik usikkerheten til estimator A. Størst forskjell har vi for bergverksdrift (næring 10, 12-14). Den estimerte variasjonskoeffisienten for denne næringen ligger mellom 6,1% og 21,0% med estimator B, og mellom 5,5% og 18,9% med estimator A.

Testingen med 2005-dataene viser altså at estimator A og estimator B er nokså like. Det kan se ut til at vi muligens treffer litt bedre med estimator B, men ikke nok til å oppveie følgende to fordeler med estimator A. For det første: Estimator A er den velkjente rateestimatoren, mens estimator B er litt mer komplisert. For det andre: For å kunne bruke estimator B kreves det at de aller fleste av utvalgsbedriftene også var med i utvalget for to år siden, mens estimator A ikke krever dette. På bakgrunn av dette anser vi derfor estimator A for å være bedre egnet enn estimator B.

Tabell 8 Estimert årsomsetning innen bergverksdrift og industri i 2005, fordelt på publiseringsnivå. Avviket er gitt ved faktisk årsomsetning (BoF) minus estimert verdi

Næring	Estimert årsomsetning med estimator B (i millioner kr)	Avvik mot faktisk årsomsetning (%)		
		med estimator B	med estimator A	med gjeldende estimator
10, 12-14	9 275	1.7	-4.7	-18.7
151-158	122 993	3.4	1.8	23.0
159-160	20 042	4.8	8.8	5.0
17-19	6 122	6.8	5.7	-4.8
20	23 373	1.8	3.7	3.6
21	18 389	4.6	4.3	14.2
22	40 112	0.3	0.3	7.6
23 og 242-247	62 440	7.8	10.6	10.0
241	30 927	1.4	0.0	-6.6
25	9 592	-6.1	-6.2	1.6
26	20 617	-0.7	-4.8	-3.5
271-273, 275	14 463	-7.1	-7.1	-6.0
274	34 243	14.5	15.0	-6.7
28	26 183	4.3	5.5	-2.6
29	40 476	5.2	5.8	-0.7
30-33	39 173	-2.2	-3.5	-15.5
34,351111-3,6,7	30 623	7.2	7.2	15.9
351114,5	41 474	-7.9	-9.4	-1.2
36-37	16 170	4.8	-1.8	-3.8
Bergverksdrift og industri i alt	606 687	3.2	2.8	5.2

95% KI for månedsomsetning i 2005, bergverksdrift og industri



Figur 6 Et tilnærmet 95% konfidensintervall for den månedlige omsetningen innen bergverksdrift og industri i 2005 (stiplede linjer). Den heltrukne linjen er estimert månedsomsetning (estimator B).

Tabell 9 Den estimerte variasjonskoeffisienten til estimator B i 2005, på publiseringnivå

Næring	Jan.	Feb.	Mars	April	Mai	Juni	Juli	Aug.	Sep.	Okt.	Nov.	Des.
10, 12-14	8.1	8.6	9.3	8.0	6.1	8.9	17.7	20.3	20.6	21.0	12.4	7.8
151-158	4.0	4.1	4.2	4.6	4.2	4.0	4.5	3.8	4.0	4.7	4.2	4.0
159-160	7.2	6.3	14.1	12.9	6.5	6.2	5.2	16.1	4.7	3.7	2.9	2.8
17-19	6.1	6.8	10.3	10.4	8.9	7.4	9.6	10.6	8.1	7.8	8.0	7.6
20	4.1	4.4	3.7	2.3	4.1	4.0	5.3	3.4	2.6	3.1	4.7	3.4
21	7.1	7.0	3.2	4.2	3.9	4.2	4.7	4.1	5.1	5.4	5.0	4.9
22	1.7	1.7	1.4	1.6	1.7	1.8	2.8	1.7	1.6	1.6	1.9	1.7
23 og 242-247	5.8	5.6	6.6	12.2	8.6	5.6	4.9	6.5	5.9	6.2	6.4	5.3
241	6.0	5.2	9.7	11.9	9.5	3.9	6.1	11.6	9.3	9.4	9.7	7.7
25	4.1	5.2	4.8	4.9	5.1	5.6	6.0	5.4	6.2	6.0	5.2	5.1
26	4.1	3.9	7.1	8.4	7.7	5.8	5.9	11.5	5.8	5.2	4.9	5.5
271-273, 275	5.6	4.9	4.4	5.6	5.1	3.5	5.7	14.5	4.6	5.2	6.0	4.7
274	6.5	6.9	4.4	9.8	6.2	9.5	4.2	5.5	9.7	7.4	7.6	6.6
28	3.1	3.4	3.7	3.6	3.3	3.5	4.2	3.4	3.4	2.9	2.7	3.8
29	5.7	5.6	4.9	5.5	6.5	4.9	8.2	5.0	4.1	5.0	4.6	4.8
30-33	6.9	8.5	8.6	7.6	7.5	7.3	8.1	8.0	6.5	5.5	6.9	6.0
34,35111-3,6,7	4.9	5.6	5.7	6.6	8.1	7.7	8.6	9.2	8.0	10.0	7.9	7.0
351114,5	14.9	11.8	11.6	10.1	10.5	9.8	10.9	10.8	9.7	9.0	8.6	9.3
36-37	3.8	3.3	4.0	4.8	4.5	5.0	5.2	4.5	4.8	4.4	4.0	4.1
Bergverksdrift og industri i alt	1.5	1.6	1.8	2.0	1.7	1.6	1.7	1.9	1.6	1.7	1.6	1.6

5. Sammendrag

I dette notatet har vi presentert to alternative estimatorene for den månedlige omsetningen innen bergverksdrift og industri. Den ene estimatoren er en stratifisert rateestimator (estimator A), den andre er en stratifisert ratebasert estimator (estimator B). I motsetning til gjeldende estimator i Omsetningsstatistikken gjør disse estimatorene ikke bruk av data fra Skattedirektoratet. (Hensikten med å finne alternative estimatorene var ønsket om ikke å bruke dataene fra Skattedirektoratet).

Testing med 2005-data tyder på at estimator A og estimator B er minst like bra som gjeldende estimator. Faktisk ser det ut som om vi treffer litt bedre med de alternative estimatorene. (For å avgjøre hvor bra estimatorene treffer har vi beregnet estimert årsomsetning ved å summere de estimerte månedsomsetningene, og så har vi sammenlignet mot årsomsetningen fra BoF som vi kjenner i ettertid).

Estimator A og estimator B gir stort sett veldig like estimat, og de har stort sett nokså lik usikkerhet. (Vi har ikke noe usikkerhetsmål på den gjeldende estimator så vi får ikke sammenlignet mot usikkerheten til gjeldende estimator). Men estimator A har to fordeler framfor estimator B. Den ene er at estimator A er en velkjent standard estimator som er enklere enn estimator B. Det andre er at estimator B er avhengig av at det ikke er for stor utskifting av bedrifter i utvalget fra år til år, mens estimator A ikke er det. Av disse grunner anser vi estimator A som den beste av disse to estimatorene.

Referanser

Statistisk sentralbyrå (2003): *Omsetningsstatistikk for olje- og gassutvinning, bergverksdrift, industri og kraftforsyning*, NOS D 251.

Statistisk sentralbyrå (1994): *Standard for næringsgruppering*, NOS C 182.

Vedlegg A

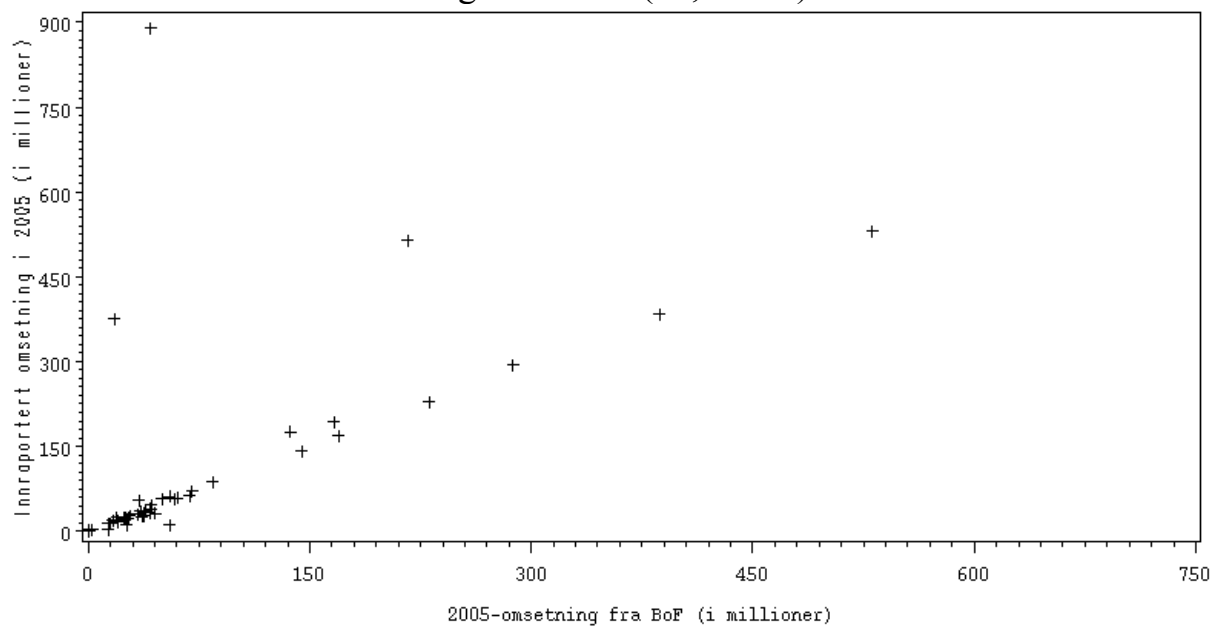
Tabell A1 Sammenligning av estimert årsumsetning (med gjeldende estimeringsopplegg i Omsetningsstatistikken) og faktisk årsumsetning (BoF). Avviket er gitt ved faktisk årsumsetning minus estimert verdi

Næring	NACE	Avvik (%)		
		2003	2004	2005
Bergverksdrift	10, 12-14	-32.7	-40.4	-18.7
Næringsmiddelindustri	151-158	9.0	10.3	23.0
Nytelsesmiddelindustri	159-160	1.6	5.2	5.0
Tekstil- og bekledningsindustri	17-19	-0.4	-3.3	-4.8
Trelast- og trevareindustri	20	1.3	8.3	3.6
Treforedling	21	9.1	17.3	14.2
Forlag og grafisk industri	22	-2.0	3.1	7.6
Oljeraffinering og kjemisk industri, unntatt kjemiske råvarer	23 og 242-247	3.3	3.6	10.0
Kjemiske råvarer	241	2.2	-5.3	-6.6
Gummivare- og plastindustri	25	4.2	5.0	1.6
Mineralproduktindustri	26	-2.7	0.4	-3.5
Metallindustri, unntatt ikke-jernholdige metaller	271-273, 275	10.1	8.4	-6.0
Ikke-jernholdige metaller	274	-7.5	-11.0	-6.7
Metallvareindustri	28	-6.5	5.1	-2.6
Maskinindustri	29	-15.0	-7.7	-0.7
Elektroteknisk og optisk industri	30-33	-6.2	-17.9	-15.5
Transportmiddelindustri	34,35111-3,6,7	-9.4	6.8	15.9
Oljeplattformer	351114,5	-0.9	4.8	-1.2
Møbelindustri og annen industri	36-37	-15.2	-10.3	-3.8
Bergverksdrift og industri i alt		-0.8	1.5	5.2

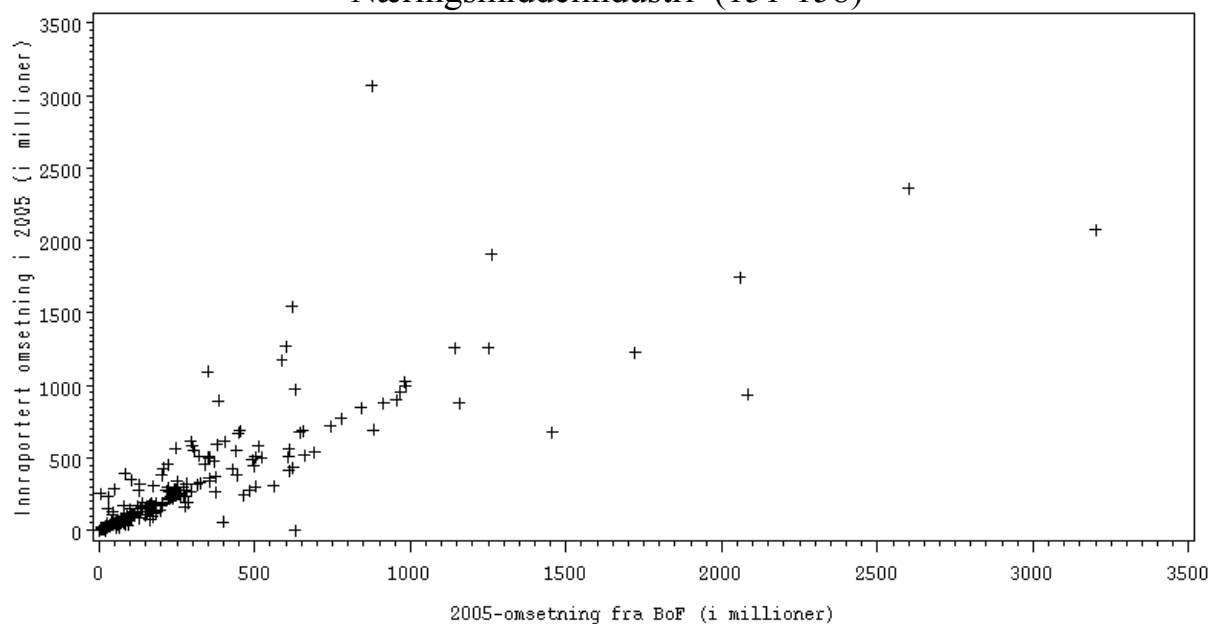
Vedlegg B

Figurene i dette vedlegget viser sammenhengen mellom 2005-omsetningen fra BoF og den innrapporterte omsetningen for utvalgsbedriftene innen noen næringer. Vi har kun tatt med bedrifter som har vært med i nettoutvalget alle 12 månedene i 2005, slik at vi har kunnet beregne innrapportert årsomsetning på bakgrunn av de innrapporterte månedsomsetningene.

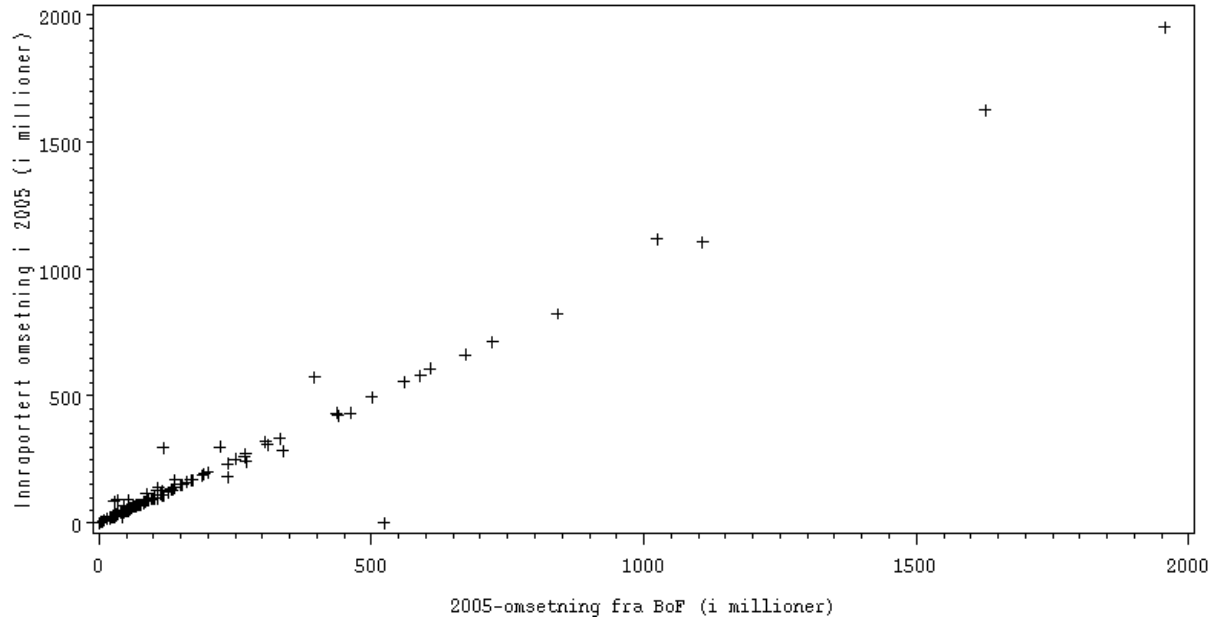
Bergverksdrift (10, 12-14)



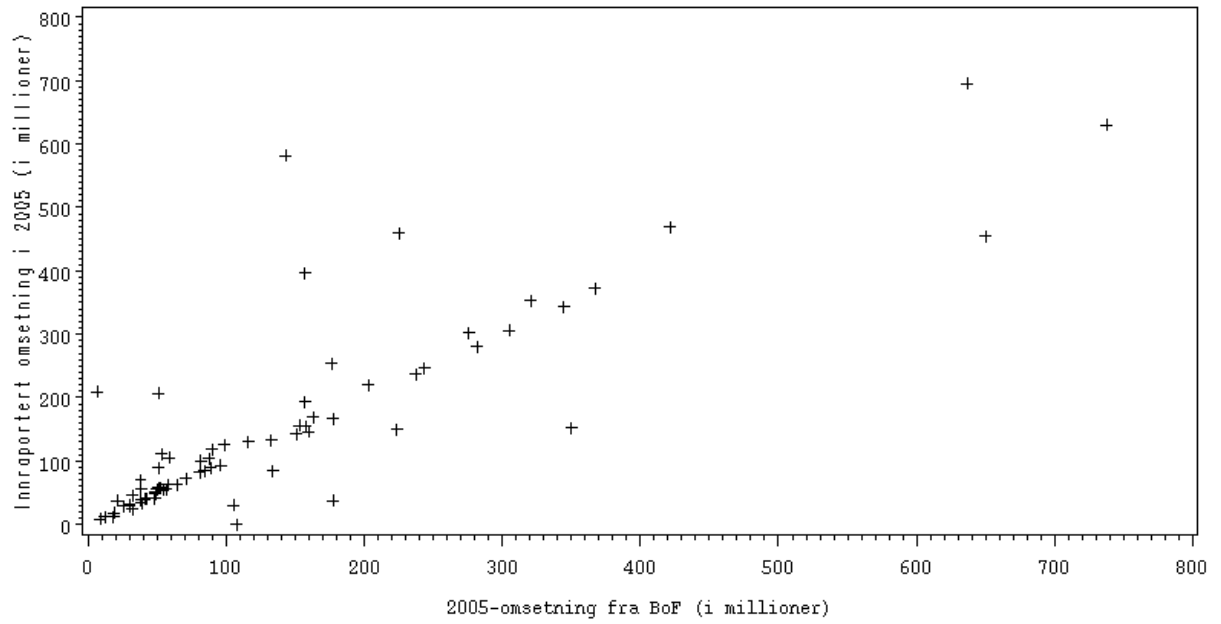
Næringsmiddelindustri (151-158)



Forlag og grafisk industri (22)



Mineralproduktindustri (26)

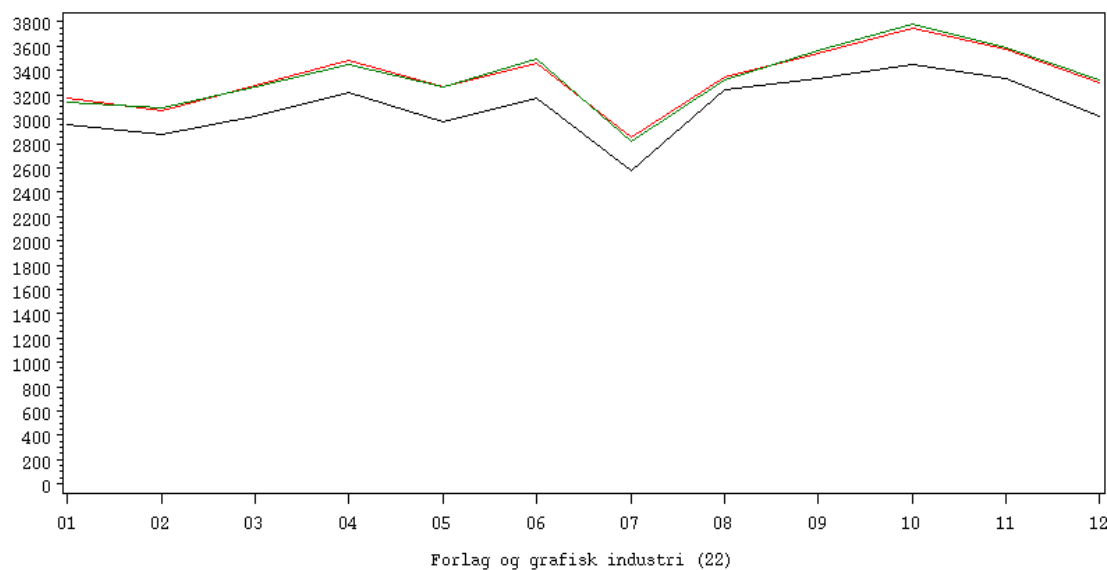
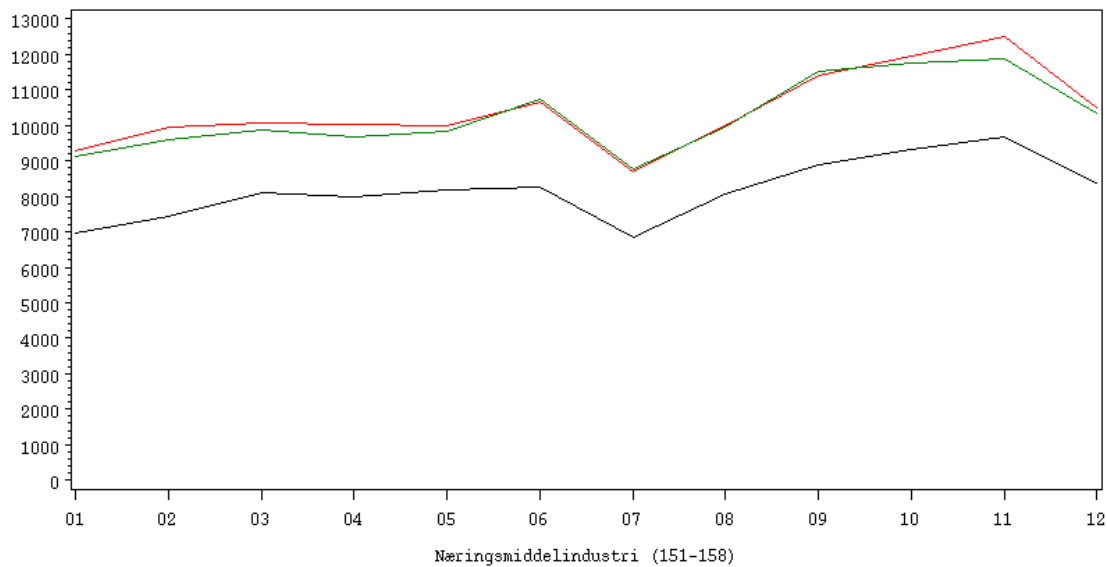


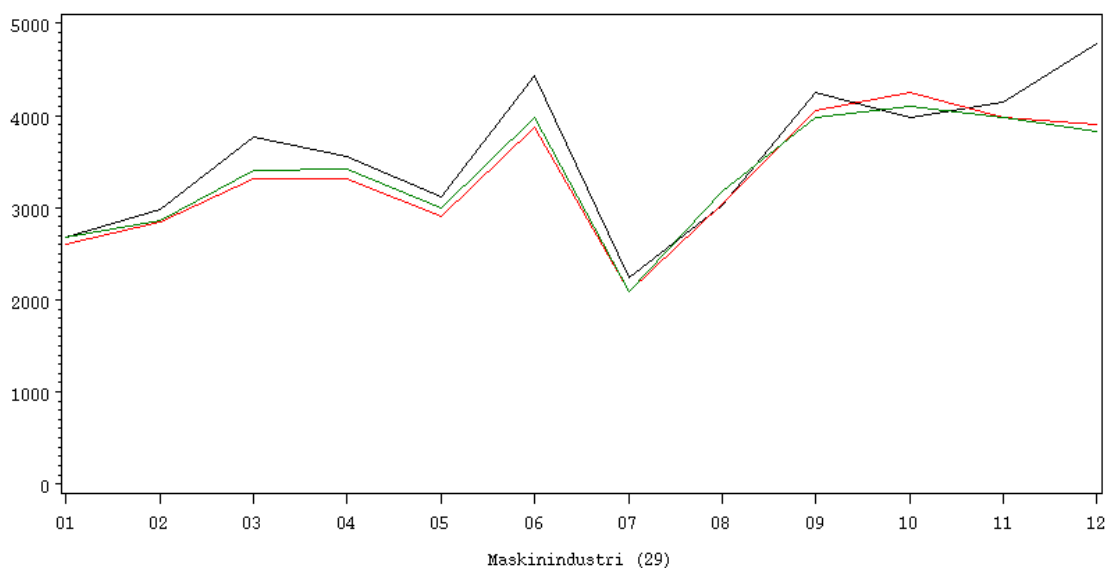
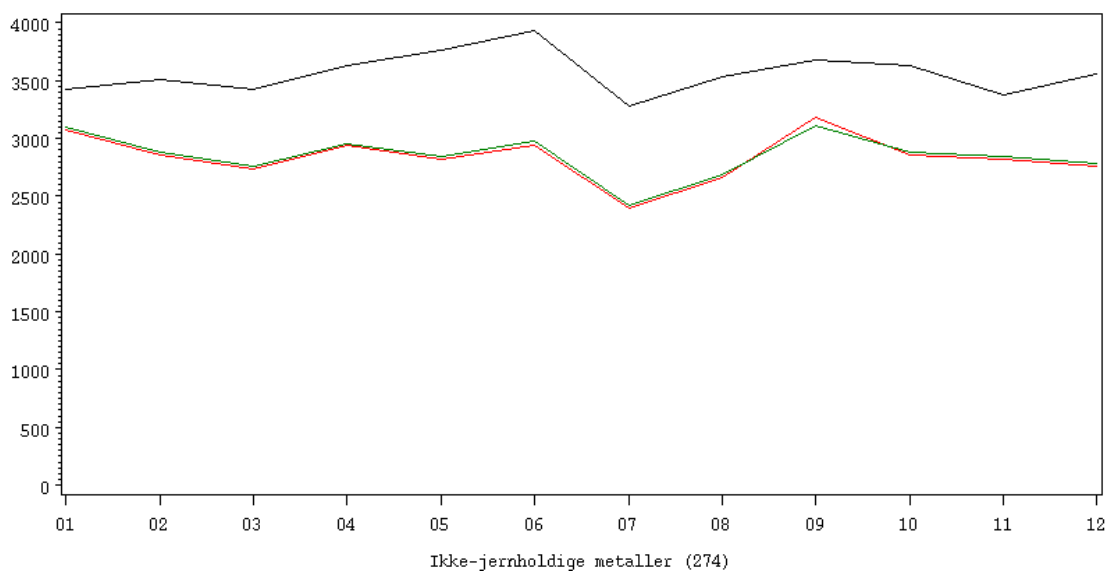
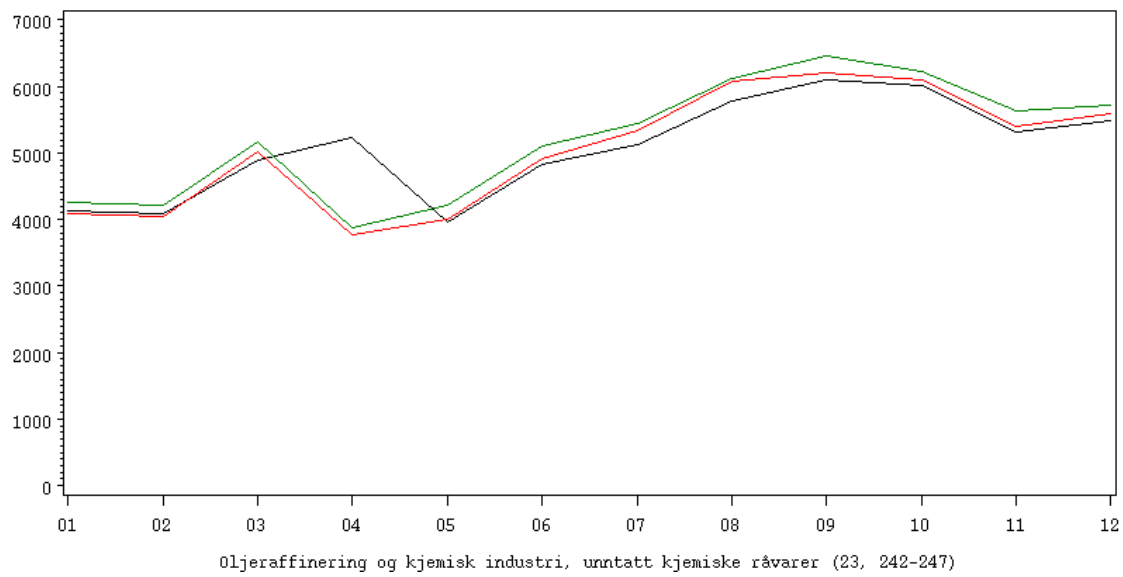
Vedlegg C

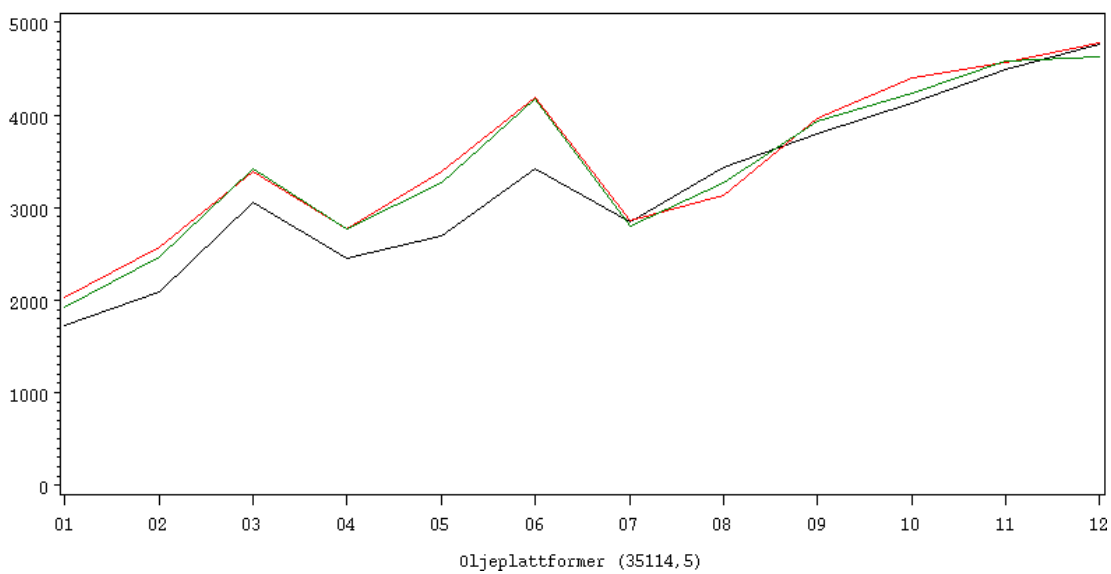
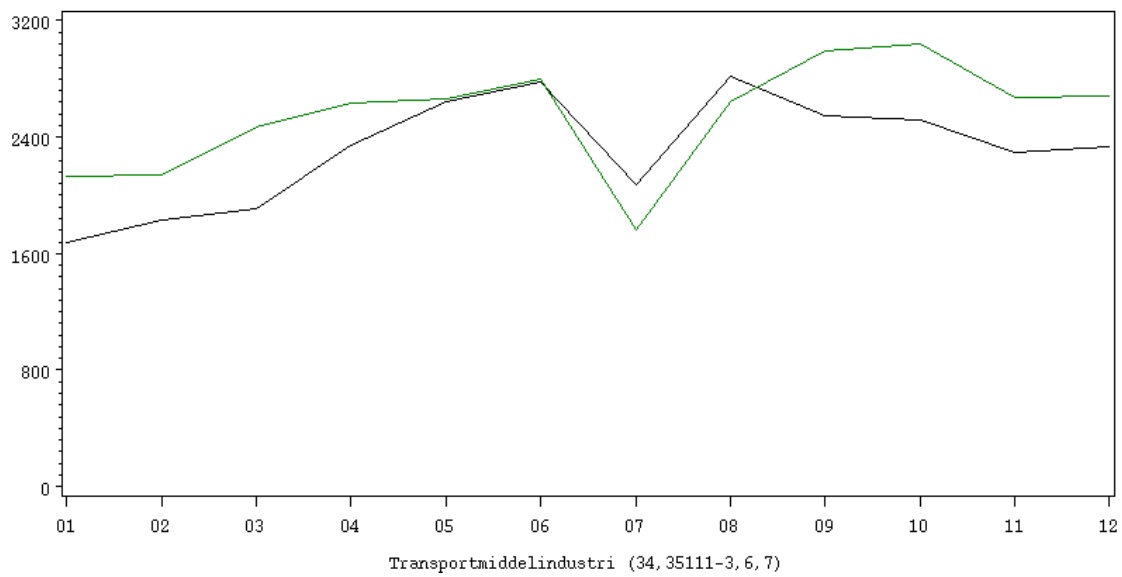
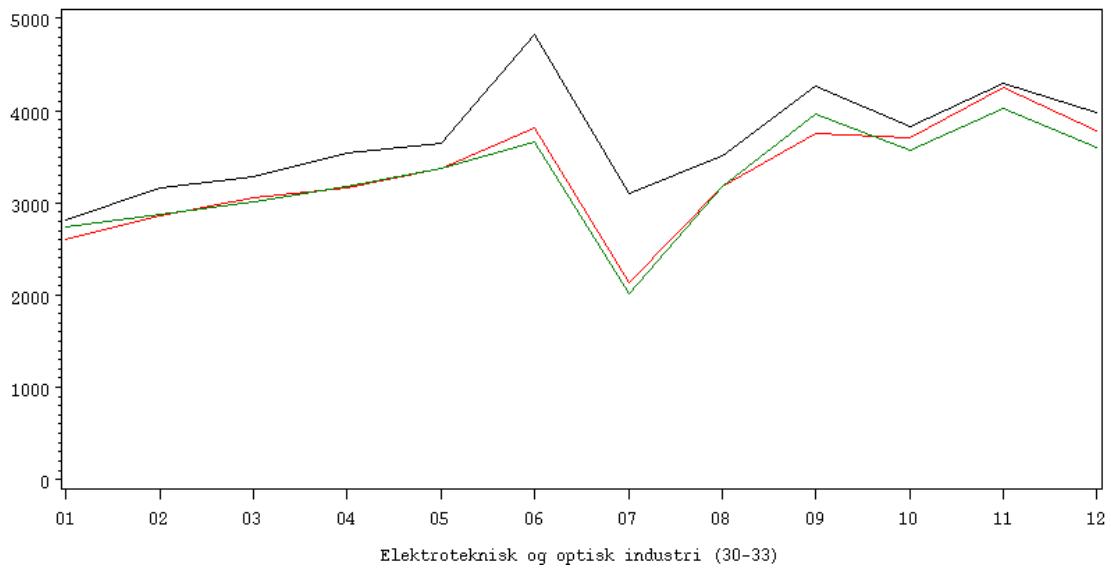
Figurene i dette vedlegget viser estimert månedsomsetning for 2005, for de største næringene på publiseringsnivå. Måned er gitt langs den horisontale akse, mens estimert verdi er gitt langs den vertikale akse (i millioner kr).

- gjeldende estimeringsopplegg
- estimator A
- estimator B

For transportmiddelindustrien (34, 35111-3,6,7) er estimator A = estimator B







Vedlegg D

Figurene i dette vedlegget viser

- estimert månedsomsetning i 2005 med estimator A (heltrukken linje)
- tilnærmet 95% konfidensintervall for den faktiske månedsomsetningen i 2005 (stiplede linjer)

for de største næringene på publiseringsnivå.

Verdiene er i millioner kr.

Konfidensintervallet inkluderer usikkerheten som følger av at det er målefeil i dataene.

