



Anna-Karin Mevik

Notater

Revisjon av Strukturstatistikk for industrien
Et forslag til selektiv revisjon

1. Innledning	2
2. Data	2
3. Seleksjonsmetoden	6
4. Utprøving av seleksjonsmetoden.....	8
4.1. Valg av α og β	8
4.1.1. Den ideelle situasjonen hvor alle store feil blir rettet.....	9
4.2. Anslått verdi	14
4.3. Revisjon av bedriftene som plukkes ut vha metoden	15
5. Automatisk korreksjon av tusenfeil.....	20
6. Oppsummering.....	22
Referanser	23
Vedlegg A	24
A.1 Variabel v220	25
A.2 Variabel v230	26
A.3 Variabel v240	27
A.4 Variabel v250	28
A.5 Variabel v291	30
A.6 Variabel v260	31
A.7 Variabel v290	32
A.8 Variabel v310	33
A.9 Variabel v314	34
A.10 Variabel v316	35
A.11 Variabel v320	36
A.12 Variabel v330	37
A.13 Variabel v340	38
A.14 Variabel v350	39
A.15 Variabel v360	40
A.16 Variabel v370	41
A.17 Variabel v380	42
A.18 Variabel v390	43
Sist utkommet i serien Notater.....	44

1. Innledning

Strukturstatistikk for industri er en årlig statistikk som blant annet skal gi tall for sysselsetting, produksjonsinntekter, produksjonskostnader og investeringer for bedriftene innen næringene bergverksdrift og industri. Til beregning av disse tallene benyttes innsamlede data fra ca. 5000 utvalgte bedrifter.

Det er manuell revisjon av hele utvalget. Men for mange av bedriftene medfører revisjonen bare små endringer. For å ikke bruke like mye tid på retting av bedrifter med bare småfeil, som på bedrifter med store feil, er det ønskelig å kunne skille ut bedriftene som har store feil helt i starten av revisjonsarbeidet. Dermed kan disse kontrolleres manuelt, mens bedriftene med bare småfeil eventuelt kan korrigeres ved hjelp av en automatisk opprettingsrutine. I dette notatet skal vi se på en ny metode for å skille ut bedriftene som skal kontrolleres manuelt. Dvs. vi skal se på en selektiv metode som har til hensikt å plukke ut de av bedriftene som har store feil.

Vi starter i kapittel 2 med å gi en kort presentasjon av dataene som vi har testet seleksjonsmetoden på. I kapittel 3 presenterer vi selve metoden, og i kapittel 4 viser vi resultatene fra testingen. (Fordi vi ser på mange variable, blir det veldig mange figurer som skal presenteres. For å gjøre notatet mer lesbart, har vi valgt å samle de fleste av figurene i et vedlegg). I tillegg til å se på den selektive metoden, har vi sett på muligheten for automatisk korreksjon av såkalte tusenfeil ("1000-feil"). Resultatene av dette gir vi i kapittel 5. Helt til slutt gir vi en oppsummering i kapittel 6.

2. Data

Til utprøving av seleksjonsmetoden bruker vi data fra 2001-utvalget til strukturstatistikken. Vi har begrenset oss til å se på bedriftene som tilhører enbedriftsforetak. Grunnen til at vi ikke ser på bedriftene som tilhører flerbedriftsforetak, er at det ikke skal gjøres noen seleksjon blant disse (alle skal tas ut til manuell revisjon). Videre tar vi ikke med nyetablerte bedrifter. For å bruke seleksjonsmetoden må vi ha relevant tilleggsinformasjon om bedriftene, som f.eks. forrige års resultater eller registervariable. Fordi vi ofte ikke har slik informasjon for nyetablerte bedrifter, kan metoden ikke brukes på disse.

Når vi har fjernet de ovennevnte bedriftene, sitter vi igjen med 2298 bedrifter. For disse bedriftene har vi valgt å se på de 19 variablene som er listet opp i tabell 1. Dette er numeriske variable, og omhandler produksjonsinntekter og produksjonskostnader til den enkelte bedrift.

Tabell 1 Variabeloversikt

Variabel	Forklarende tekst
	Produksjonsinntekter fordelt på:
v210	Salg av egenproduserte varer
v220	Salg av handelsvarer
v230	Reparasjonsarbeid
v240	Leiearbeid
v250	Andre salgsinntekter
v291	(Sum av variablene v210, v220, v230, v240 og v250)
v260	Øvrige driftsinntekter
v290	Produksjonsinntekter i alt (= v291 + v260)
	Produksjonskostnader fordelt på:
v310	Råvarer, halvfabrikata, hjelpstoffer og emballasje
v314	Solgte handelsvarers kostnad

v316	Leiearbeid
v320	Energiforbruk
v330	Frakt og spedisjon vedrørende salget
v340	Lønn, feriepenger, honorarer mv.
v350	Arbeidsgiveravgift
v360	Reparasjon og vedlikehold
v370	Leiekostnader
v380	Forbruk av andre varer og tjenester
v390	Produksjonskostnader i alt (= v310 + + v380)

For de 2298 bedriftene vi ser på, har vi både originale og reviderte data. De originale dataene kalles også rådata eller ureviderte data. Dette er dataene slik de er før revisjonen tar til.¹

De reviderte dataene kalles også statistikkdata eller endelige data. Dette er de dataene vi ender opp med når revisjonen er ferdig. Vi omtaler alle verdier i dette datasettet som revidert verdi (selv om det eventuelt ikke er gjort noen endring i forhold til original verdi).

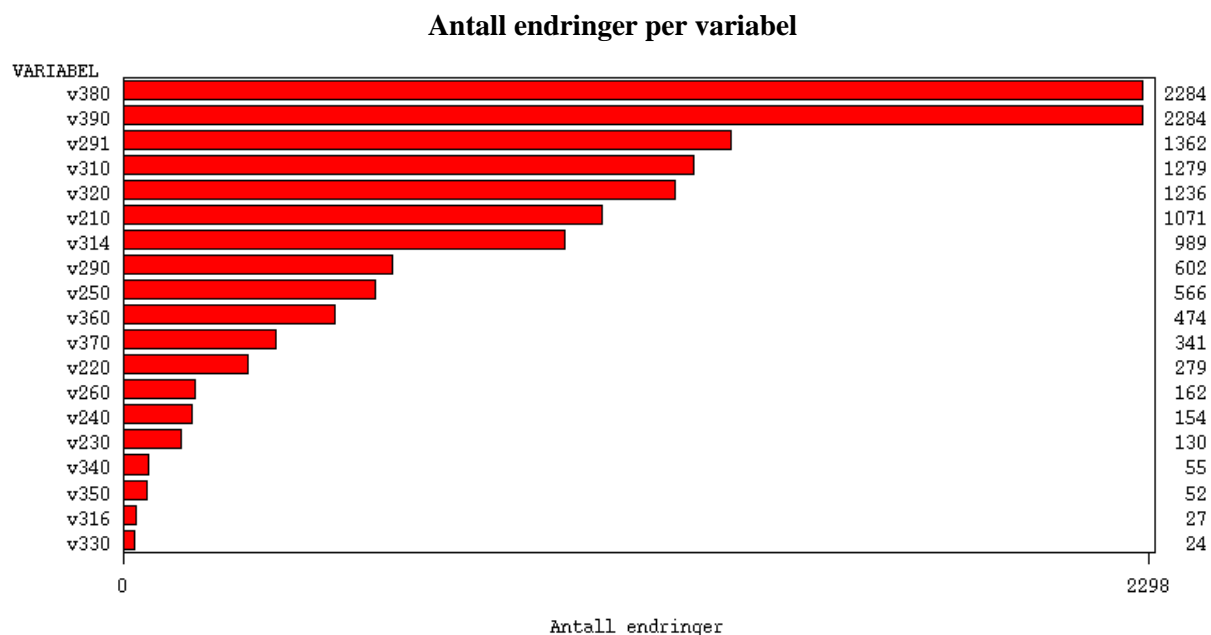
I dette notatet vil vi behandle de reviderte verdiene som om de er de korrekte, eller riktige, verdiene. Dette er selvsagt ikke alltid tilfelle. Selv om alle bedriftene tas ut til manuell kontroll, vet revisoren sjelden hva de korrekte verdiene skal være. Men som oftest vil de reviderte verdiene være nærmere de riktige verdiene enn de originale.

Fordi revidert verdi blir betraktet som riktig verdi, vil en original verdi bli betraktet som feil hvis den ikke er lik den reviderte verdien. En endring er dermed ensbetydende med at original verdi er feil. (Med endring mener vi at original verdi ikke er lik revidert verdi). Dette trenger selvsagt ikke være tilfelle. Det kan tenkes at en original verdi som er korrekt, blir endret i løpet av revisjonen. Men forhåpentligvis skjer dette ikke så ofte.

Ved å sammenligne originale med reviderte data, finner vi at det bare er en bedrift som ikke har fått endret noen av sine variable. De andre bedriftene har fått endret alt fra 1 til 15 av sine variable. I gjennomsnitt er 5.8 variable endret for en bedrift. Dvs. gjennomsnittlig er litt under 1/3 av variablene til en bedrift endret.

Teller vi opp antall endringer per variabel, finner vi at det varierer fra 24 til 2284 (se figur 1). Det er variabel v330 som har færrest endringer, og det er variablene v380 og v390 som har flest endringer. I gjennomsnitt er en variabel endret 704 ganger. Dvs. i gjennomsnitt er en variabel endret for drøye 30% av bedriftene.

¹ Det er ingen manglende verdier (missing) i disse dataene.



Figur 1 Stolpediagram som viser antall endringer som er gjort per variabel.

Hvor store endringene er, varierer veldig. Variabel v390 har f.eks. bare veldig små endringer, mens variabel v210 har noen veldig store endringer der original verdi er mye større enn den reviderte verdien (se figur 2). Situasjonen vi har for v210 gjelder også for variablene v220, v230, v240, v250 og v291. Mange av disse store endringene skyldes det vi kaller tusenfeil, dvs. at bedriften har oppgitt variabelverdien i hele kroner mens den egentlig skal gis i tusen kroner. Dette kommer vi tilbake til i kapittel 5.

For å få et inntrykk av hvor mye som endres for en variabel, har vi for hver variabel beregnet størrelsene

$$a1 = \frac{\sum \text{original verdi}}{\sum \text{revidert verdi}}$$

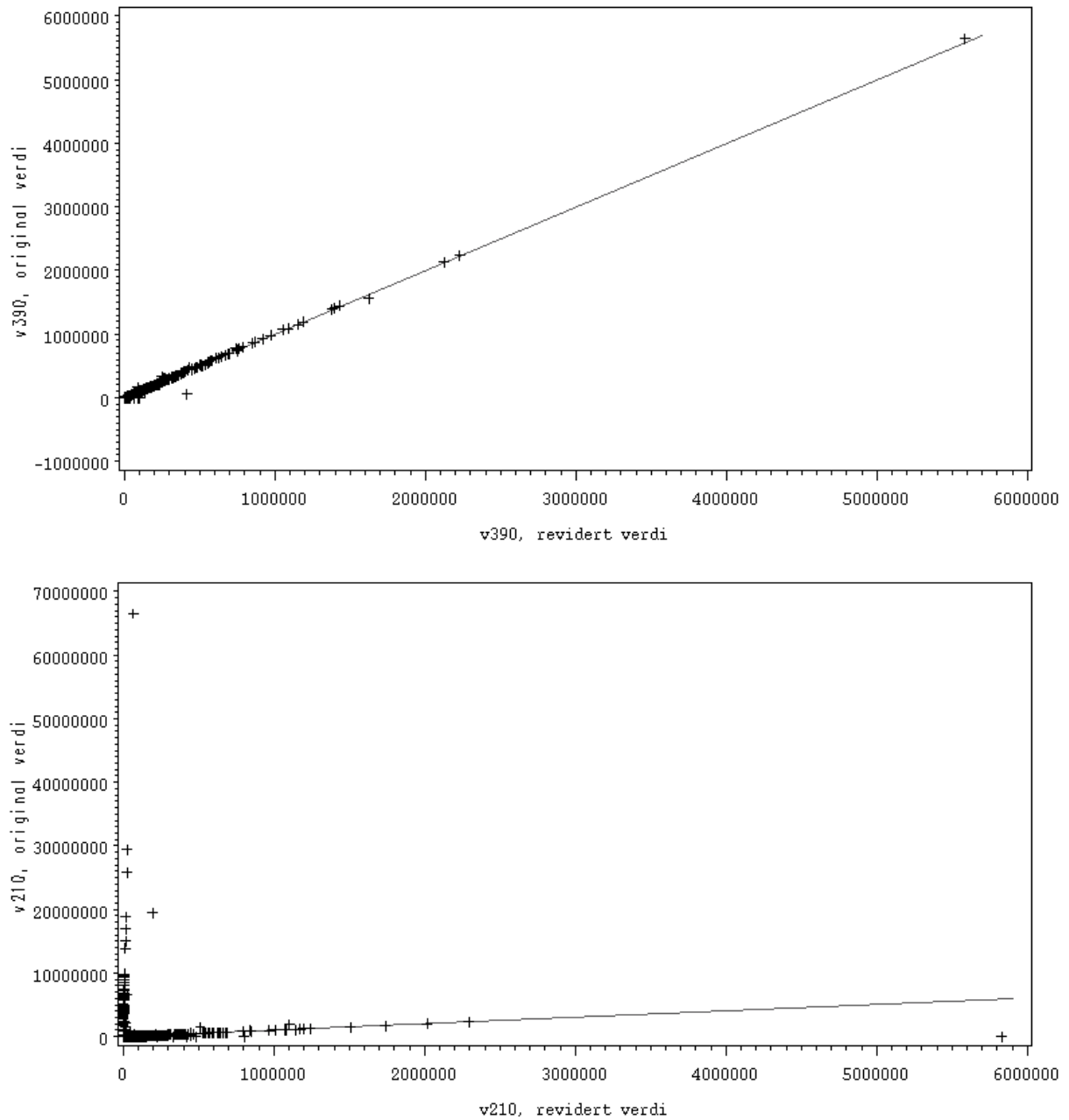
og

$$b1 = \frac{\sum |\text{revidert verdi} - \text{original verdi}|}{\sum \text{revidert verdi}} \cdot 100.$$

Her går summen er over alle 2298 bedriftene, og tegnet $||$ betyr absoluttverdi (dvs. at $|3| = 3$ mens $|-3| = 3$).

Størrelsen $a1$ gir forholdet mellom summert verdi før og etter revisjon. At $a1$ er nær 1 kan bety at det er få og små endringer, eller at positive og negative endringer oppveier hverandre. (Med positiv endring mener vi at revidert verdi er større enn original verdi, og med negativ endring mener vi at revidert verdi er mindre enn original verdi).

Plott av original verdi mot revidert verdi



Figur 2 *Plott av original mot revidert verdi for variablene v390 og v210 (verdiene er gitt i tusen kroner). Den heltrukne linjen markerer hvor original verdi er lik revidert verdi. Dvs. langs denne linjen markeres bedrifter som ikke har endret den aktuelle variabelen.*

Størrelse b_1 gir den totale endringen til variabelen, i prosent av sum revidert verdi. Denne størrelsen vil være liten hvis det er få og små endringer, og stor hvis det er mange og store endringer. Fordi det er absoluttverdien til endringene som benyttes i b_1 , vil ikke positive og negative endringer kunne oppveie hverandre. Dermed kan b_1 bli stor selv om a_1 er nær 1.

Tallene vi har fått for a_1 og b_1 , samt antall endringer per variabel, er gitt i tabell 2. Som vi ser er $a_1 \approx 1$ for mange av variablene, også for v390 som er endret for nesten alle bedriftene. For de fleste av variablene hvor $a_1 \approx 1$, er b_1 liten. For variablene v210, v220, v230, v240, v250 og v291, som

domineres av noen originale verdier som er mye større enn de reviderte, er naturlig nok *a1* og *b1* blitt veldig store. (De ekstreme verdiene til variabel v250 skyldes en original verdi som er eksepsjonelt stor i forhold til den reviderte verdien, se figur A7 i vedlegget).

Tabell 2 Sammenligning av originale og reviderte data
(Kolonne 3 viser størrelsen *a1*, og kolonne 4 viser størrelsen *b1*)

Variabel	Antall endringer	Forh. mellom summene basert på originale og reviderte verdier	Total endring i % av sum revidert verdi
v210	1071	4.73	394.10
v220	279	5.97	555.10
v230	130	6.16	554.98
v240	154	7.20	670.59
v250	566	96.40	9658.14
v291	1362	4.26	343.45
v260	162	0.97	13.71
v290	602	1.00	2.30
v310	1279	0.75	31.72
v314	989	3.38	298.58
v316	27	0.98	2.58
v320	1236	0.96	21.69
v330	24	1.00	0.24
v340	55	0.99	1.67
v350	52	0.99	1.23
v360	474	0.99	1.51
v370	341	1.00	0.85
v380	2284	1.12	17.11
v390	2284	1.00	2.81

For en sammenligning av originale og reviderte data fra 1996, se Abrahamsen, Andersen og Ragnarsøn, 2001.

3. Seleksjonsmetoden

Selektiv revisjon betyr at bare noen av bedriftene tas ut til (manuell) revisjon. Hvilke bedrifter som tas ut, avhenger av metoden. Metoden vi skal se på her, har til hensikt å plukke ut de av bedriftene som har stor feil for en gitt variabel (variabelen må være numerisk). Utplukkingen vil være automatisk så fort metoden er tilpasset og implementert i revisjonssystemet.

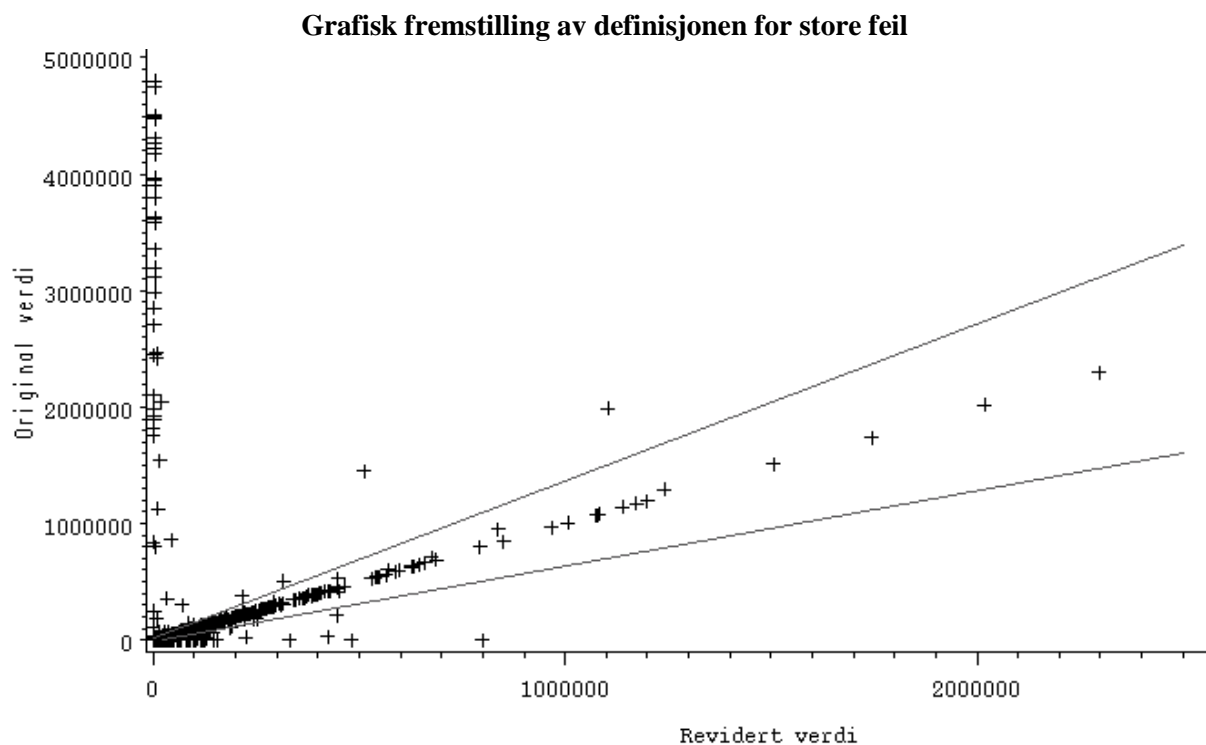
Med stor feil mener vi at avviket, eller avstanden, mellom original og revidert verdi er stor. Mer presist definerer vi en stor feil ved at

$$|\text{revidert verdi} - \text{original verdi}| > \alpha + \beta \cdot |\text{revidert verdi}|,$$

der $\alpha \geq 0$ og $0 \leq \beta < 1$. Hensikten med metoden er med andre ord å få plukket ut de bedriftene som tilfredsstiller denne ulikheten.

Hvis $\beta > 0$ så betyr det at vi aksepterer større avvik mellom original og revidert verdi jo større den revidert verdien er (se figur 3 for en illustrasjon). Hvis $\alpha = 0$ så betyr det at vi ikke aksepterer feil av typen original verdi = 0 mens revidert verdi $\neq 0$, eller motsatt original verdi $\neq 0$ mens revidert verdi = 0.

For å bestemme hvor store α og β skal være, må vi sammenligne originale og reviderte data fra tidligere perioder. Dette vil gi oss et inntrykk av hvor mange og hvilke typer feil som gjøres, og hvor store feilene er. Ved å prøve ulike verdier for α og β , får vi også en pekepinne på hvilke og hvor mange av disse feilene som skal plukkes ut. Valg av α og β blir så en avveining mellom hvor nøyaktig tall vi ønsker og hvor mye resurser som skal brukes til retting.



Figur 3 Plott av original verdi mot revidert verdi for variabelen v210 (i tusen kroner). De to heltrukne linjene markerer grensene for når en feil er stor; bedrifter med store feil er markert enten over den øverste linjen eller under den nederste linjen. (For å gjøre figuren mer lesbar er ikke alle bedriftene markert i figuren). I denne figuren er $\alpha = 15\,000$ (når vi måler i tusen kroner) og $\beta = 0.35$.

Utplukkingen av bedriftene som skal revideres, skal selvsagt skje før vi kjenner de reviderte verdiene. For å få til dette bruker vi anslåtte (estimerte) verdier for de reviderte verdiene. Dvs. vi anslår, eller gjetter på, de reviderte verdiene (før selve opprettingen av eventuelle feil tar til). Så plukker vi ut bedriftene hvor avviket mellom original og anslått verdi er stor. Dvs. vi plukker ut bedrifter hvor

$$|\text{anslått verdi} - \text{original verdi}| > \alpha + \beta \cdot |\text{anslått verdi}|.$$

For å bestemme hvordan vi skal anslå den reviderte verdien, må vi sette oss inn i hvordan variabelen blir revidert. Vi må prøve med ulike anslag, og se hvilket som egner seg best til å plukke ut de ønskede bedriftene. Til denne utprøvingen trenger vi ikke bare originale og reviderte data fra tidligere perioder, men også eventuelle andre data som benyttes ved revisjonen. (Det er viktig å huske at den anslåtte verdien skal kunne programmeres som en automatisk rutine, uten vurderinger fra de som utfører revisjonen).

Det er en stor fordel om de anslåtte verdiene bare gjør bruk av data som er tilgjengelig samtidig som bedriftene svarer. Da kan vi nemlig avgjøre om en bedrift skal tas ut til revisjon med en gang bedriften har svart (i motsetning til en del andre selektive metoder der det kreves at alle bedriftene må svare før denne avgjørelsen kan tas). Dermed kan revisjonsarbeidet starte så fort de første bedriftene har svart.

Hvor treffsikker metoden er, dvs. hvor mange av de store feilene som plukkes ut, avhenger av den anslåtte verdien. Jo bedre vi klarer å anslå den reviderte verdien, jo flere av de store feilene vil bli plukket ut (samtidig som færre småfeil plukkes ut). Dette er med andre ord en metode som egner seg for variable hvor det fins tilgjengelig informasjon som kan brukes til å gi et godt anslag på den reviderte verdien.

Slik vi har definert store feil, kan vi ha tilfeller hvor original verdi < 0 mens revidert verdi > 0 , uten at det regnes som en stor feil. Et eks. på dette kan være at $\alpha = 1000$, original verdi er -500 og revidert verdi er 500 . Ønsker vi at dette skal regnes som en stor feil, kan definisjonen modifieres.

Det er viktig å være klar over at metoden vi har presentert her ikke nødvendigvis plukker ut de feilene som har størst innvirkning på estimatoren (dvs. det oppblåste tallet som skal representere hele populasjonen). Hvis det er disse feilene vi vil plukke ut, må vi ta hensyn til vektene i tillegg til størrelsen på feilene.

4. Utprøving av seleksjonsmetoden

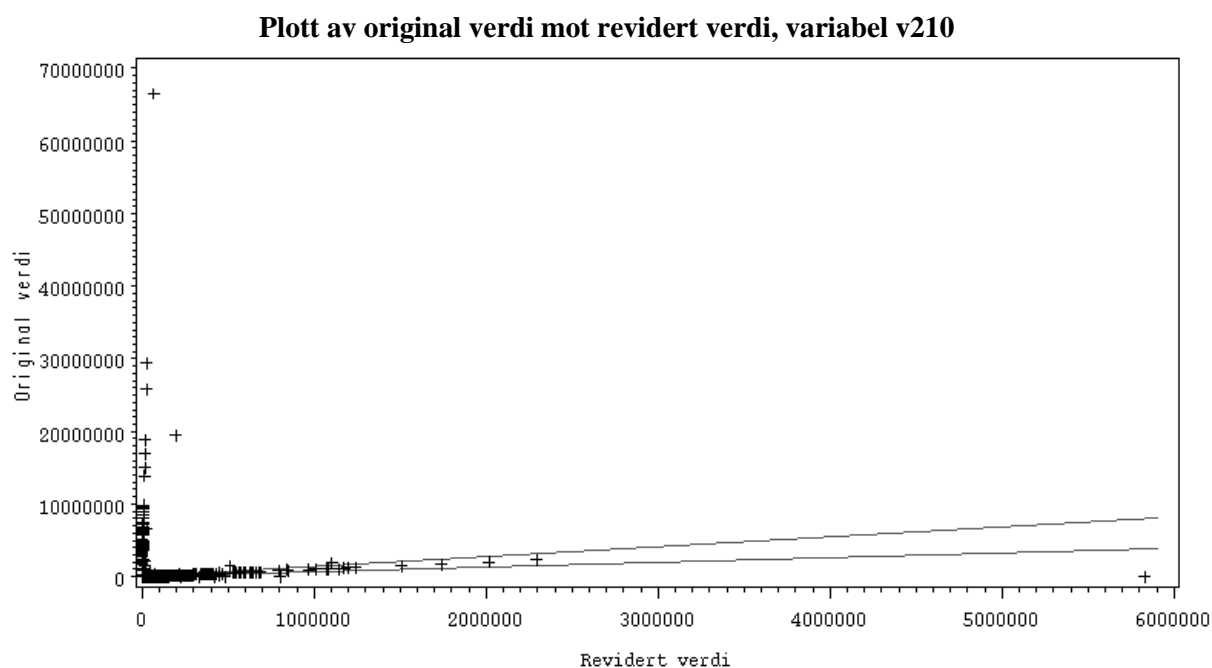
Til utprøving av metoden bruker vi dataene som er beskrevet i kapittel 2. Dvs. vi bruker data fra 2001-utvalget til strukturstatistikken. Strengt tatt skulle vi brukt data fra to år, ett datasett til å tilpasse metoden og ett til å se hvor godt den fungerer. Fordi vi bruker samme datasett til både tilpassing og testing, vil resultatene antyde at metoden (med den tilpassingen vi gjør her) er bedre enn det som er tilfelle. Men med en bedre tilpassing vil nok metoden være minst like bra som resultatene tilsier. (Det avgjørende for hvor bra metoden fungerer, er hvor godt vi klarer å anslå de reviderte verdiene, og dette kan nok gjøres en del bedre enn det vi gjør i dette notatet).

4.1. Valg av α og β

Første steg i tilpassingen av metoden er valg av α og β . En mulighet er å velge ulike verdier for ulike variable. Dette kan være aktuelt hvis f.eks. variablene har veldig ulik skala, eller hvis noen av variablene er viktigere enn andre slik at vi ønsker nøyaktigere tall for disse. For enkelthets skyld har vi valgt å bruke samme α og β for alle variablene, og vi har valgt verdiene ut fra at det ikke skal bli for

mange store feil. Som mål har vi at ikke flere enn ca. $\frac{1}{4}$ av bedriftene skal tas ut til revisjon. For å få til dette må vi ha $\alpha > 0$. Hvis $\alpha = 0$ regnes det som stor feil når original verdi = 0 mens revidert verdi $\neq 0$, eller motsatt original verdi $\neq 0$ mens revidert verdi = 0. Fordi variabel v314 har 701 slike feil, blir det for mange bedrifter som skal tas ut til revisjon om $\alpha = 0$.²

Ved å prøve oss frem har vi endt opp med $\alpha = 15\ 000$ (når vi måler verdiene i tusen kroner) og $\beta = 0.35$. Dette betyr f.eks. at 153 av variabel v210 sine feil er definert som store, og disse inkluderer alle de dominerende endringene der original verdi er mye større en revidert verdi (se figur 4 for en illustrasjon av hvilke feil som er store). De andre variablene har alt fra 0 til 282 store feil (se tabell 3). Spesielt kan vi nevne v380 og v390, som begge har 2284 feil, men hvor bare henholdsvis 8 og 10 er store (se vedlegg for figurer tilsvarende figur 4).



Figur 4 Plott av original verdi mot revidert verdi for variabelen v210 (i tusen kroner). De to heltrukne linjene markerer grensene for når en feil er stor.

4.1.1. Den ideelle situasjonen hvor alle store feil blir rettet

I en ideell situasjon blir alle de store feilene rettet. For å få et inntrykk av hvor mye de gjenværende feilene, dvs. småfeilene, utgjør i forhold til de fullstendig reviderte dataene, har vi beregnet størrelsene

$$a2 = \frac{\sum_{s_2} \text{revidert verdi} + \sum_{s_2} \text{original verdi}}{\sum_{\text{alle bedrifter}} \text{revidert verdi}}$$

og

² Fra og med 2003 er rutinene for innhenting av variabelverdiene til v314 endret. Dette er gjort i den hensikt å bedre kvaliteten på de originale dataene til variabelen, og dermed redusere antall feil.

$$b2 = \frac{\sum_{s_2} |\text{revidert verdi} - \text{original verdi}|}{\sum_{\text{alle bedrifter}} \text{revidert verdi}} \cdot 100,$$

der $s_2 =$ bedrifter med stor feil (for den aktuelle variabelen)

$\bar{s}_2 =$ resten av bedriftene, dvs. de som ikke har stor feil (for den aktuelle variabelen).

Merk at s_2 og \bar{s}_2 vil være forskjellig fra variabel til variabel. For variabel v210 består f.eks. s_2 av de 153 bedriftene som har stor feil for denne variabelen, mens for v316 består s_2 av de 2 bedriftene som har stor feil for v316.

Størrelse $a2$ gir forholdet mellom summert verdi basert på delvis reviderte data og fullstendig reviderte data. Med delvis reviderte data mener vi her at bare de store feilene er rettet opp, mens fullstendig reviderte data betyr at alle feil er rettet opp. Størrelse $b2$ gir total endring til de gjenværende feilene, i prosent av sum reviderte verdier.

Størrelsene $a2$ og $b2$ tilsvarer størrelsen $a1$ og $b1$ ved å bytte ut de originale dataene med de delvis reviderte dataene. For variable som ikke har store feil (som f.eks. v330) er derfor $a2 = a1$ og $b2 = b1$.

Det er viktig å være klar over at $a2$ ikke kan tolkes som forholdet mellom estimatene som er basert på delvis reviderte data og på fullstendig reviderte data. (Med estimat mener vi det oppblåste tallet som skal representere hele populasjonen). Selv om $a2 \approx 1$, trenger det ikke bety at estimatet basert på delvis reviderte data er tilnærmet lik estimatet basert på fullstendig reviderte data. Det samme gjelder for tilsvarende forhold vi skal se på senere.

Tallene vi har fått for $a2$ og $b2$ er vist i tabell 3. Vi ser at $a2 \approx 1$ for nesten alle variablene, og at $b2$ er liten for mange av variablene. Sammenligner vi $b2$ med $b1$, ser vi at $b2$ er mindre eller lik $b1$ for alle variablene (slik det skal være). For variablene som har veldig stor $b1$, er $b2$ atskillig mindre enn $b1$. Dette betyr at de store feilene, selv om det ikke er så mange, utgjør mesteparten av den totale feilen til disse variablene.

Det er to variable som skiller seg ut med stor $b2$, og det er v250 og v314. Småfeilene til disse variablene utgjør så mye som 35.49% og 42.79% av sum reviderte verdier, respektivt. Ønsker vi mindre tall, må vi velge mindre verdier for α og β .

Tabell 3 Sammenligning av delvis reviderte og fullstendig reviderte data, når delvis reviderte data betyr at de store feilene er rettet. (Kolonne 3 viser størrelsen $a2$, og kolonne 4 viser størrelsen $b2$)

Variabel	Antall store feil	Forh. mellom summene basert på delvis og fullstendig rev. data	Total endring til gjenværende feil, i % av sum rev. verdi
v210	153	0.99	2.70
v220	62	1.01	7.67
v230	22	0.94	13.87
v240	25	0.96	19.54
v250	70	1.23	35.49
v291	153	1.00	1.92
v260	3	1.00	3.57
v290	16	1.00	0.26
v310	215	0.96	6.97
v314	282	1.24	42.79
v316	2	1.00	0.34
v320	4	1.00	14.45
v330	0	1.00	0.24
v340	3	0.99	0.90
v350	0	0.99	1.23
v360	0	0.99	1.51
v370	0	1.00	0.85
v380	8	1.11	14.12
v390	10	1.01	2.15

Fordi det er sammenhenger mellom mange av variablene, f.eks. $v290 = v291 + v260$, kan ikke variablene revideres uavhengig av hverandre. Tanken er derfor at alle bedrifter som har stor feil for minst en av sine variable skal tas ut til revisjon, og så rettes alle variablene til disse bedriftene. (Hvis variablene kan deles inn i grupper som kan revideres uavhengig av hverandre, kan det tas ut bedrifter for hver enkelt gruppe, og da skal kun variablene til den aktuelle gruppen rettes).

Det er 524 bedrifter som har stor feil for minst en av sine variable, og det er disse vi ønsker å ta ut til revisjon. (I gjennomsnitt har disse bedriftene 1.96 store feil). De andre 1774 bedriftene har kun småfeil. For å se hvor mye småfeilene til disse bedriftene utgjør i forhold til de fullstendig reviderte dataene, har vi beregnet størrelsene

$$a3 = \frac{\sum_{s_3} \text{revidert verdi} + \sum_{\bar{s}_3} \text{original verdi}}{\sum_{\text{alle bedrifter}} \text{revidert verdi}}$$

og

$$b3 = \frac{\sum_{s_3} |\text{revidert verdi} - \text{original verdi}|}{\sum_{\text{alle bedrifter}} \text{revidert verdi}} \cdot 100,$$

der s_3 = de 524 bedriftene som har stor feil for minst en av sine variable

\bar{s}_3 = de resterende bedriftene, dvs. de 1774 bedriftene som kun har småfeil.

Forskjellen på $a3$ og $a2$, og på $b3$ og $b2$, er at flere feil er rettet i $a3$ og $b3$ enn i $a2$ og $b2$. I $a2$ og $b2$ er kun de store feilene rettet, mens i $a3$ og $b3$ er en god del småfeil rettet i tillegg til de store. Som en følge av dette er $b3$ mindre enn $b2$.

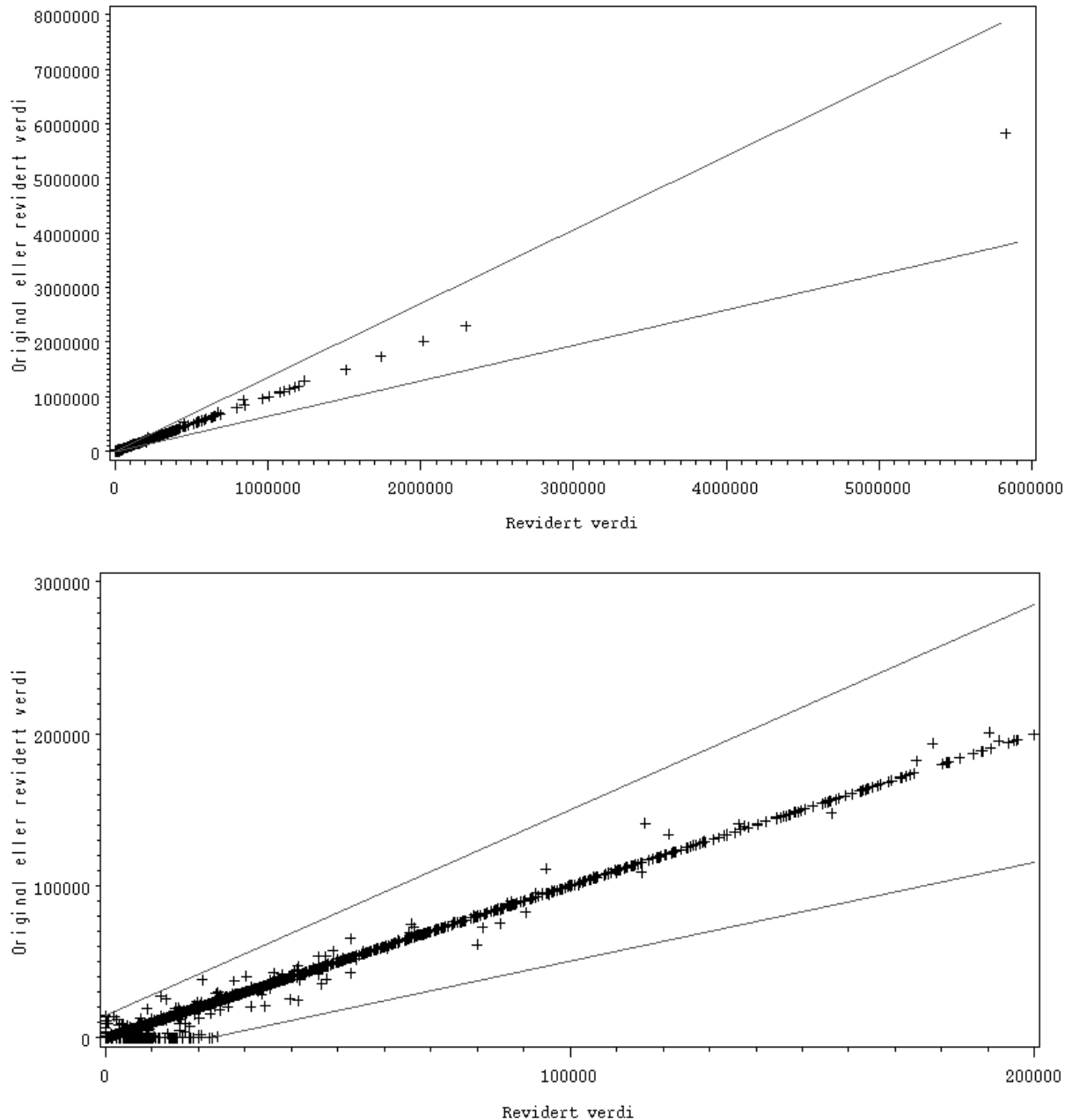
Tallene vi har fått for $a3$ og $b3$ er vist i tabell 4. Det er bare tre variable som har $b3 > 10$, mens syv av variablene har $b3 < 1$. Nesten alle variabelen har $a3 \approx 1$.

Tabell 4 Sammenligning av delvis reviderte og fullstendig reviderte data, når delvis reviderte data betyr revisjon av de 524 bedriftene som har store feil. (Kolonne 2 viser størrelsen $a3$, og kolonne 3 viser størrelsen $b3$)

Variabel	Forh. mellom summene basert på delvis og fullstendig rev. data	Total endring til gjenværende feil, i % av sum rev. verdi
v210	1.00	1.61
v220	1.01	5.21
v230	0.96	9.93
v240	0.99	13.18
v250	1.19	26.92
v291	1.00	1.49
v260	1.00	1.27
v290	1.00	0.23
v310	0.98	4.47
v314	1.20	36.82
v316	1.00	0.21
v320	1.01	8.04
v330	1.00	0.09
v340	1.00	0.38
v350	1.00	0.35
v360	1.00	0.36
v370	1.00	0.47
v380	1.07	8.52
v390	1.00	1.17

For å se hvilke feil som gjenstår, i den ideelle situasjon hvor de 524 bedriftene med store feil blir revidert, har vi plottet delvis reviderte verdier mot reviderte verdier for variablene v210 (se figur 5). Med delvis revidert verdi mener vi revidert verdi hvis bedriften tas ut til revisjon, og original verdi ellers. Dvs. her er delvis revidert verdi = revidert verdi for de 524 bedriftene med store feil, og delvis revidert verdi = original verdi for de 1774 bedriftene som ikke har store feil.

Plott av delvis reviderte verdier mot reviderte verdier, variabel v210



Figur 5 Plott av delvis reviderte verdier mot reviderte verdier for variabelen v210 (i tusen kroner), når det er de 524 bedriftene med store feil som blir revidert. Dvs. at delvis revidert verdi = revidert verdi for de 524 bedriftene med store feil, og delvis revidert verdi = original verdi for de 1774 bedriftene som ikke har store feil. De to heltrukne linjene markerer grensene for når en feil er stor. Det nederste plottet er et utsnitt av det over.

Som vi ser av figur 5 er det kun småfeil som gjenstår for v210, akkurat som det skal. Vi ser også at en del av disse feilene er av typen original verdi = 0 mens revidert verdi > 0, eller motsatt original verdi > 0 mens revidert verdi = 0. Teller vi opp finner vi at det er 85 slike feil. Andre variable har også gjenværende feil av denne typen. Vi kan nevne v250 hvor hele 337 av de 399 gjenværende feilene er av denne type, og v314 som har 455 slike gjenværende feil.

4.2. Anslått verdi

Neste steg i tilpassingen av metoden er å bestemme hvordan vi skal anslå de reviderte verdiene. Til hjelp ved revisjonen benyttes flere datakilder, som f.eks. forrige års reviderte verdier, årsregnskap for foretaket og næringsoppgaver. I dette notatet har vi bare benyttet oss av forrige års reviderte verdier, dvs. reviderte verdier for 2000.

Av de 2298 bedriftene vi ser på, er det 512 bedrifter som ikke var med i 2000-utvalget. Dermed har vi ikke forrige års reviderte verdier for disse. Men vi har estimerte verdier. I strukturstatistikken får nemlig alle bedriftene i populasjonen, unntatt de i utvalget, estimert sine variabelverdier. Oppblåsing til populasjonstall skjer så ved summering av estimerte og reviderte verdier. (For mer om estimeringen, se Abrahamsen, Andersen og Ragnarsøn, 2001). Dermed har vi estimerte verdier for de 512 bedriftene som ikke var med i 2000-utvalget. Fordi vi ikke skal gjøre noen forskjell på om forrige års verdi er revidert eller estimert, vil vi for enkelhets skyld omtale alle som reviderte verdier.

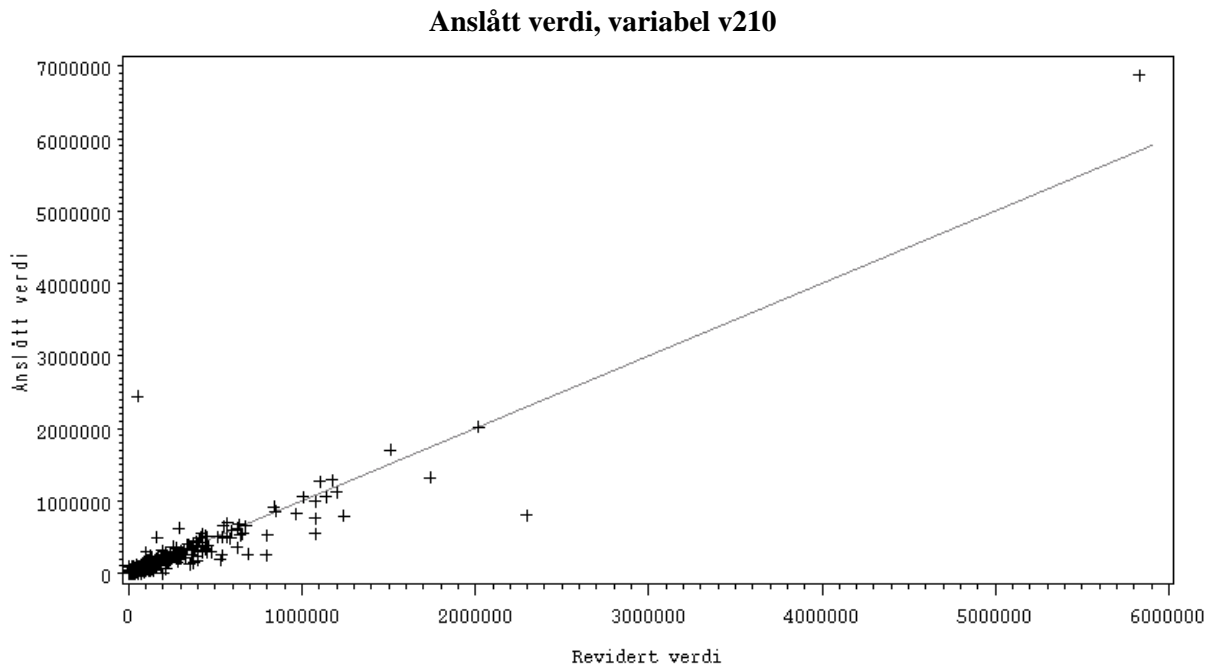
Etter å ha prøvd oss litt frem, har vi for de fleste av variablene valgt å anslå revidert verdi med forrige års reviderte verdi, eller med en enkel kombinasjon av original verdi og forrige års reviderte verdi. For noen av variablene har vi valgt å lage anslåtte verdier som avhenger av anslåtte verdier til andre variable.

Som et eks. på hvordan revidert verdi blir anslått skal vi se på variable v210. For denne variabelen har vi valgt å anslå den reviderte verdien med forrige års revidert verdi, dvs.

$$\text{anslått verdi} = \text{forrige års reviderte verdi}.$$

(Se vedlegg for hvordan anslått verdi er for de andre variablene). Dette anslaget treffer relativt godt for mange av bedriftene. For litt under $\frac{3}{4}$ av bedriftene er avstanden mellom anslått og revidert verdi mindre enn 5000, og for omtrent halvparten av bedriftene er avstanden mindre enn 2000 (når vi måler verdiene i tusen kroner). Men for noen av bedriftene bommer vi veldig på den reviderte verdien (se figur 6).³

³ Denne sammenligningen av anslått og revidert verdi er gjort med samme data som vi har brukt til å bestemme hvordan den anslåtte verdien skal være. For å få et mer realistisk bilde av hvor godt vi klarer å anslå de reviderte verdiene burde vi egentlig brukt et annet datasett til sammenligningen.



Figur 6 Plott av anslått (estimert) verdi mot revidert verdi for variabelen v210 (i tusen kroner). Den heltukne linjen markerer hvor anslått verdi er lik revidert verdi. Dvs. at bedrifter hvor vi klarer å anslå den reviderte verdien nøyaktig, er markert langs denne linjen.

4.3. Revisjon av bedriftene som plukkes ut vha metoden

Metoden plukker ut bedrifter hvor

$$|\text{anslått verdi} - \text{original verdi}| > 15000 + 0.35 \cdot |\text{anslått verdi}|$$

(når vi måler i tusen kroner).

Hvor mange bedrifter som plukkes ut per variabel varierer fra 0 til 278. Variable v210 får f.eks. plukket ut 225 bedrifter. Av de 153 bedriftene som har stor feil for variabel v210, er 134 blant disse. Dvs. at 19 av bedriftene med stor feil (for v210) ikke er blant de 225 utplukkede bedriftene, mens 91 bedrifter som ikke har stor feil (for v210) likevel er plukket ut for variabelen. Se tabell 5 for hvor mange bedrifter som plukkes ut per variabel, og hvor mange av de store feilene som ikke er blant disse.

For å se hvor mye de gjenværende feilen utgjør i forhold til fullstendig reviderte verdier, når vi bare retter feilene til de bedriftene som plukkes ut for den aktuelle variabelen, har vi beregnet størrelsene

$$a4 = \frac{\sum_{s_4} \text{revidert verdi} + \sum_{\bar{s}_4} \text{original verdi}}{\sum_{\text{alle bedrifter}} \text{revidert verdi}}$$

og

$$b4 = \frac{\sum_{s_4} |revidert\ verdi - original\ verdi|}{\sum_{alle\ bedrifter} revidert\ verdi} \cdot 100,$$

der s_4 = bedrifter som plukkes ut vha metoden (for den aktuelle variabelen)

\bar{s}_4 = resten av bedriftene, dvs. de som ikke plukkes ut vha metoden (for den aktuelle variabelen).

Tallene vi har fått for $a4$ og $b4$ er vist i tabell 5. Sammenligner vi $a4$ med $a2$ ser vi at de er veldig like hverandre. Det eneste unntaket er for v250, hvor $a4 = 0.72$ og $a2 = 1.23$. Sammenligner vi $b4$ med $b2$, ser vi at de ikke er så ulike for mange av variablene. Men for noen av variablene er $b4$ veldig mye større enn $b2$. Et eks. er variabel v250, hvor $b4 = 89.91$ og $b2 = 35.49$. Denne store forskjellen skyldes at en av de 6 store feilene som ikke plukkes ut for variabelen, er veldig stor i forhold til de reviderte verdiene. Av dette ser vi at det ikke bare er hvor mange av de store feilene som ikke plukkes ut som betyr noe, men også hvor store de er i forhold til reviderte verdier.

Tabell 5 Sammenligning av delvis reviderte og fullstendig reviderte data, når delvis reviderte data betyr at feilene som plukkes ut vha metoden er rettet. (Kolonne 4 viser størrelsen $a4$, og kolonne 5 viser størrelsen $b4$)

Variabel	Antall bed. som plukkes ut	Antall store feil som ikke plukkes ut	Forh. mellom summene basert på delvis og fullstendig rev. data	Total endring til gjenværende feil, i % av sum rev. verdi
v210	225	19	0.99	2.92
v220	80	6	0.91	18.69
v230	35	2	0.93	13.90
v240	37	3	0.96	21.03
v250	67	6	0.72	89.91
v291	161	1	1.00	1.88
v260	0	3	0.97	13.71
v290	16	5	0.99	0.67
v310	203	19	0.95	8.10
v314	278	20	1.24	50.07
v316	0	2	0.98	2.58
v320	4	1	1.02	16.07
v330	0	0	1.00	0.24
v340	2	2	0.99	1.49
v350	0	0	0.99	1.23
v360	0	0	0.99	1.51
v370	0	0	1.00	0.85
v380	21	3	1.11	13.99
v390	8	3	1.01	2.48

Til nå har vi sett hvor mange bedrifter som plukkes ut per variabel. Fordi forskjellige variable kan plukke ut forskjellige bedrifter, blir det totalt tatt ut 606 bedrifter. Dvs. at vi totalt får plukket ut 606 bedrifter vha metoden, og det er disse bedriftene som blir revidert.

Ideelt ønsker vi å reviderte de 524 bedriftene som har store feil. Av de 606 bedriftene som plukkes ut vha metoden, er det 500 som har store feil (de resterende 106 bedrifter har kun småfeil). Dvs. at 24 av de 524 bedriftene vi ønsker å revidere, ikke blir revidert.

Fordi vi retter alle variablene til de 606 bedriftene som tas ut vha metoden, får vi rettet flere feil en det som ligger til grunn for tabell 5. Spesielt får vi rettet flere av de store feilene. F.eks. får vi nå rettet 150 av variabel v210 sine store feil, slik at det bare er tre store feil som ikke blir rettet. For variablene v220, v291, v260, v320, v340 og v380 får vi rettet alle store feil (se tabell 6).

For å se hvor mye feilene til de 1692 bedriftene som ikke plukkes ut vha metoden, utgjør i forhold til fullstendig reviderte data, har vi beregnet størrelsene

$$a5 = \frac{\sum_{s_5} \text{revidert verdi} + \sum_{\bar{s}_5} \text{original verdi}}{\sum_{\text{alle bedrifter}} \text{revidert verdi}}$$

og

$$b5 = \frac{\sum_{s_5} |\text{revidert verdi} - \text{original verdi}|}{\sum_{\text{alle bedrifter}} \text{revidert verdi}} \cdot 100,$$

der s_5 = de 606 bedriftene som plukkes ut vha metoden

\bar{s}_5 = resten av bedriftene, dvs. de 1692 bedriftene som metoden ikke plukker ut.

Forskjellen på $a5$ og $a4$, og på $b5$ og $b4$, er at flere feil er rettet i $a5$ og $b5$ enn i $a4$ og $b4$. I $a4$ og $b4$ er det kun rettet feil til de bedriftene som plukkes ut for den aktuelle variabelen. I $a5$ og $b5$ er det i tillegg rettet feil til bedriftene som plukkes ut for de andre variablene. (Hvilke bedrifter som tas ut til retting i $a4$ og $b4$ varierer fra variabel til variabel, mens i $a5$ og $b5$ er det de samme 606 bedriftene som tas ut til retting for alle variablene).

Forskjellen på $a5$ og $a3$, og på $b5$ og $b3$, er at $a3$ og $b3$ gjelder den ideelle situasjonen hvor vi har revidert de 524 bedriftene med store feil, mens $a5$ og $b5$ gjelder den faktiske situasjonen hvor vi har revidert de 606 bedriftene som plukkes ut vha metoden. Det er derfor interessant å sammenligne $a5$ og $b5$ med $a3$ og $b3$.

Tallene vi har fått for $a5$ og $b5$ er gitt i tabell 6. Vi ser at $a5$ så å si er lik $a3$ for alle variablene. Vi ser også at det ikke er noen nevneverdig forskjell på $b5$ og $b3$. Faktisk er $b5$ en anelse mindre enn $b3$ for de fleste av variablene. F.eks. er $b5 = 1.35$ mens $b3 = 1.61$ for v210. Dette skyldes at flere bedrifter er revidert i $b5$ enn i $b3$, så selv om ikke alle store feil blir rettet, kan $b5$ bli mindre enn $b3$.

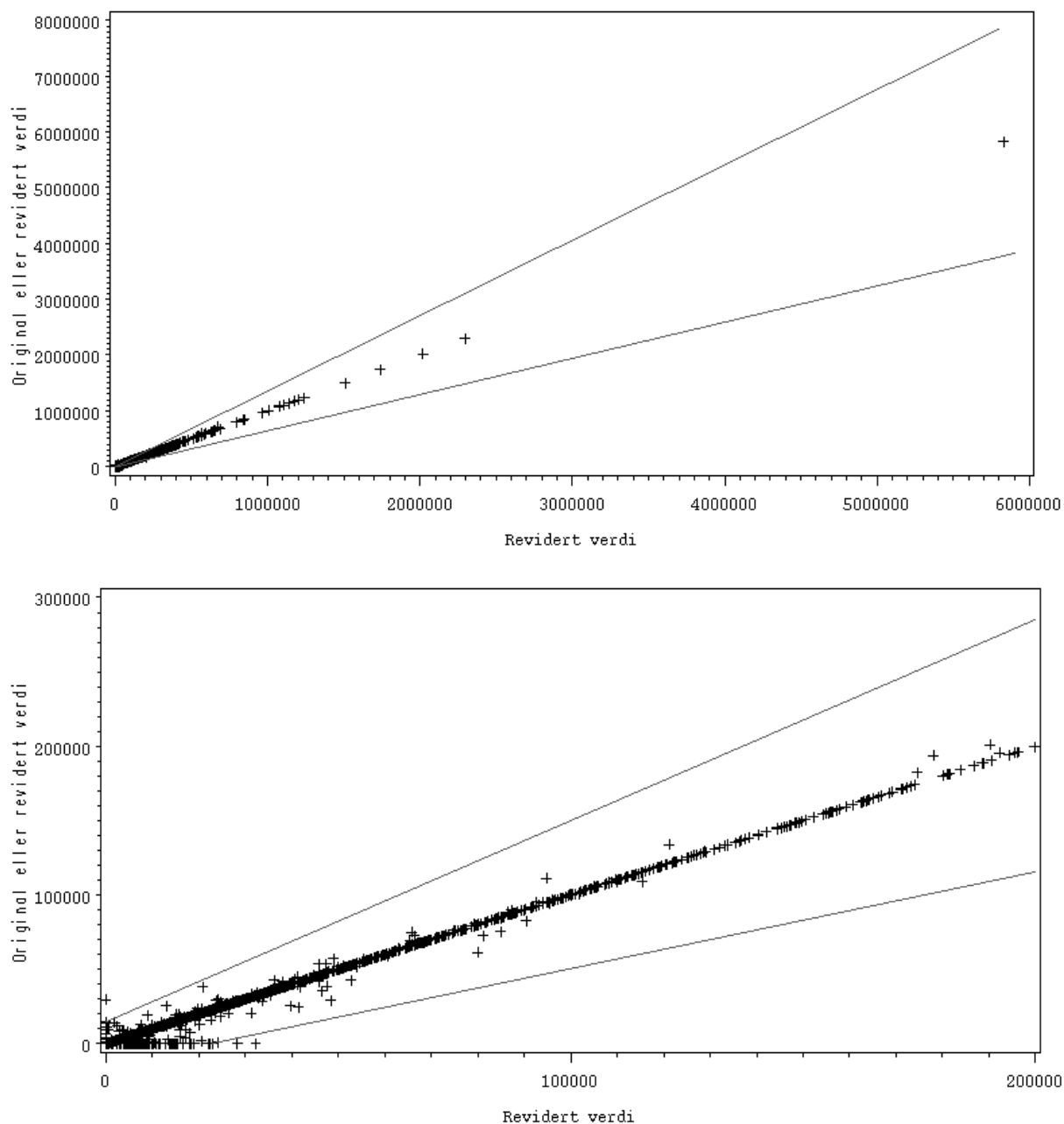
(Den store feilen til v250 som gjorde så stort utslag på $b4$ i forhold til $b2$, er nå plukket ut til revisjon. Derfor er ikke $b5$ veldig mye større enn $b3$. Faktisk er $b5 \approx b3$).

Tabell 6 Sammenligning av delvis reviderte og fullstendig reviderte data, når delvis reviderte data betyr revisjon av de 606 bedriftene som plukkes ut vha metoden. (Kolonne 3 viser størrelsen $a5$, og kolonne 4 viser størrelsen $b5$)

Variabel	Antall store feil som ikke plukkes ut	Forh. mellom summene basert på delvis og fullstendig rev. data	Total endring til gjenværende feil, i % av sum rev. verdi
v210	3	1.00	1.35
v220	0	1.01	4.11
v230	1	0.96	8.03
v240	1	0.99	12.25
v250	3	1.18	26.81
v291	0	0.99	1.16
v260	0	1.00	1.28
v290	2	1.00	0.26
v310	8	0.97	4.69
v314	15	1.24	38.86
v316	1	0.99	1.26
v320	0	1.02	6.52
v330	0	1.00	0.08
v340	0	1.00	0.46
v350	0	1.00	0.44
v360	0	1.00	0.35
v370	0	1.00	0.40
v380	0	1.06	7.42
v390	1	1.00	1.09

For å se hvilke feil som gjenstår når de 606 bedriftene som plukkes ut vha metoden er revidert, har vi plottet delvis reviderte verdier mot reviderte verdier for variabel v210 (se figur 7). Av dette ser vi at de tre store feilene til v210 som ikke blir rettet, er blant de minste av de store feilene. Dvs. vi får rettet de verste av de store feilene. Det samme gjelder for de andre variablene (se vedlegg for figurer tilsvarende figurer 7).

Plott av delvis reviderte verdier mot reviderte verdier, variabel v210



Figur 7 Plott av delvis reviderte verdier mot reviderte verdier for variabelen v210 (i tusen kroner), når det er de 606 bedriftene som plukkes ut vha metoden som blir revidert. Dvs. at delvis revidert verdi = revidert verdi for de 606 bedriftene som plukkes ut vha metoden, og delvis revidert verdi = original verdi for de 1692 bedriftene som ikke plukkes ut. De to heltrukne linjene markerer grensene for når en feil er stor. Det nederste plottet er et utsnitt av det over.

Utprøvingen vi har gjort i dette kapitlet tyder på at seleksjonsmetoden fungerer ganske bra for industristatistikken. Vi får plukket ut flesteparten av de bedriftene vi ønsker, og de vi ikke får plukket ut har ikke de verste feilene. (Men når metoden brukes på en og en variabel om gangen, fungerer den ikke fult så bra. Da blir færre av de store feilene plukket ut, og for v250 blir en av de verste feilene ikke plukket ut).

Metoden kan nok tilpasses bedre ved å ta i bruk flere datakilder⁴, og ved å samarbeide med de som utfører revisjonen. Dette vil i så fall gjøre metoden mer treffsikker.

Det som gjenstår av testing (men som vi ikke gjør i dette notatet), er å se hvordan metoden fungerer på andre datasett.

5. Automatisk korreksjon av tusenfeil

I dette kapitlet skal vi veldig kort se på muligheten for automatisk korreksjon av tusenfeil. Den automatiske korreksjonen skal gjøres før seleksjonen tar til, i den hensikt at færre bedrifter skal bli tatt ut til revisjon.

Tusenfeil betyr at bedriften har oppgitt en verdi i hele kroner, mens den egentlig skal oppgis i hele tusen kroner. Denne feilen ønsker vi å korrigere automatisk ved å dele den originale verdien på 1000, dvs. den automatisk korrigerte verdien er gitt ved

$$\frac{\text{original verdi}}{1000}.$$

Hensikten med denne korrigeringen er som nevnt at færre bedrifter skal bli tatt ut til revisjon. For å få til dette ønsker vi at mange feil skal bli korrigert automatisk. Men samtidig må vi passe på at de korrigerte verdiene er såpass like de faktisk reviderte verdiene at de ikke regnes som store feil. Ut fra disse betraktningene har vi valgt å definere en tusenfeil ved at

$$\left| \frac{\text{original verdi}}{1000} - \text{revidert verdi} \right| < \beta \cdot |\text{revidert verdi}|,$$

der $\beta = 0.35$.

I tillegg vil vi at bare store feil skal kunne være tusenfeil. Dvs. at selv om en småfeil tilfredsstillende utlikheten over, regnes den ikke som en tusenfeil.

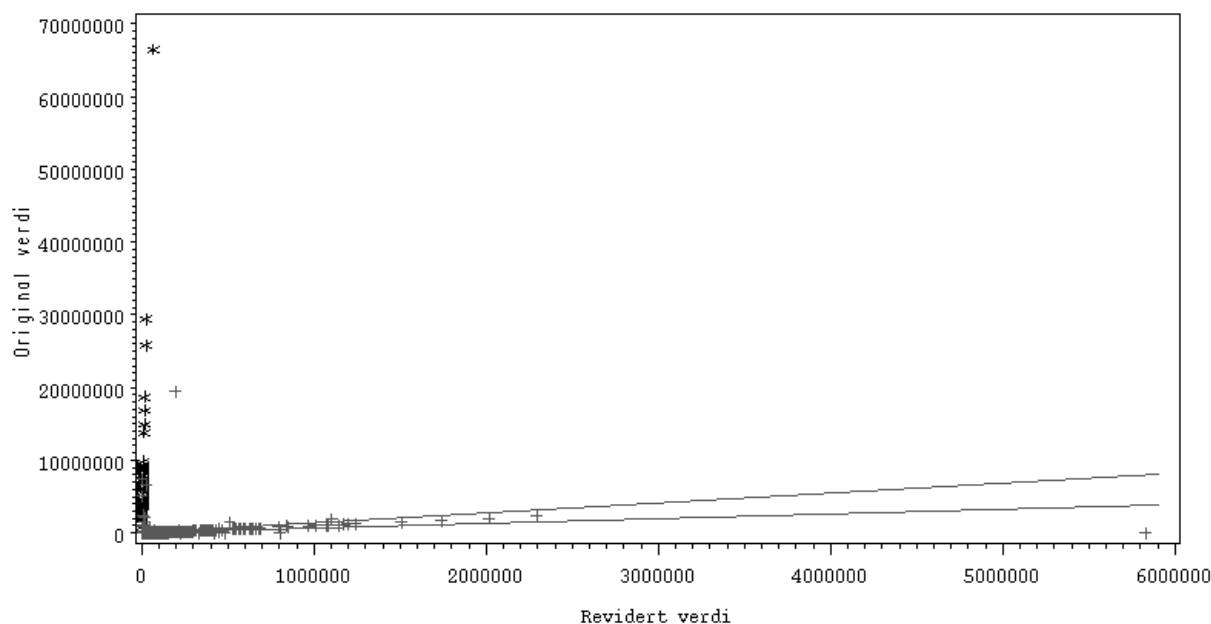
Med denne definisjonen er det variablene v210, v220, v230, v240, v250, v291, v290 og v314 som har tusenfeil (se tabell 6). F.eks. har variabel v210 57 tusenfeil, og dette er de dominerende feilene til variabelen (se figur 8).

Tabell 6 Variabler som har tusenfeil

Variabel	Antall tusenfeil
v210	57
v220	14
v230	5
v240	6
v250	5
v291	52
v290	1
v314	3

⁴ Hvis vi går over fra revisjon av alle bedrifter til selektiv revisjon, har vi ikke lengre fullstendig reviderte data fra foregående år som en datakilde. (Vi har kun reviderte data for de bedriftene som ble plukket ut ved forrige års revisjon). Fordi vi i dette notatet har brukt forrige års reviderte verdi for å anslå den reviderte verdien til mange av variablene, må det lages nye anslag for disse variablene.

Tusenfeil, variabel v210



Figur 8 Plott av original verdi mot revidert verdi for variabelen v210 (i tusen kroner). Tusenfeilene er markert med *. (De to heltrukne linjene markerer grensene for når en feil er stor).

Totalt er det 73 bedrifter som har en eller flere tusenfeil. Men fordi noen av disse også har store feil som ikke er tusenfeil, får vi bare redusert antall bedrifter som vi ønsker å ta ut til revisjon med 55. Dvs. at vi får redusert antall bedrifter som skal tas ut til revisjon fra 524 til 469 bedrifter, hvis vi først gjør en automatisk korrigering av tusenfeil.

Fordi vi ikke kjenner den reviderte verdien på det tidspunktet hvor vi skal korrigere for tusenfeil, bruker vi anslått verdi. Dvs. vi gjør en automatisk korrigering hvis

$$\left| \frac{\text{original verdi}}{1000} - \text{anslått verdi} \right| < 0.35 \cdot |\text{anslått verdi}|$$

(og hvis den anslåtte verdien tilsier at det er en stor feil).

Fordi vi bruker anslått verdi i stedet for revidert verdi, kan vi ikke regne med å få korrigert alle tusenfeil. For variabel v210 blir f.eks. 43 av tusenfeilene korrigert, hvilket betyr at 14 ikke blir korrigert (se tabell 7).

I tillegg til at ikke alle tusenfeil blir korrigert, kan det forekomme feilaktige korrigeringer. Dvs. at det blir gjort en automatisk korrigering der det ikke er noen tusenfeil. Dette er tilfelle for variablene v220 og v314. Begge disse variablene får korrigert en stor feil som ikke er tusenfeil.

Tabell 7 Variabler som får korrigert tusenfeil

Variabel	Antall tusenfeil som blir korrigert	Antall feilaktige korrigeringer
v210	43	0
v220	3	1
v230	1	0
v240	2	0
v250	0	0
v291	50	0
v290	0	0
v314	0	1

Totalt er det 63 bedrifter som får korrigert en eller flere av sine variable. Men når vi tar hensyn til at noen av disse likevel blir plukket ut som følge av en annen variabel, får vi bare redusert antall bedrifter som plukkes ut med 35. Dvs. vi får redusert antall bedrifter som plukkes ut fra 606 til 571 hvis vi først gjør automatisk korrigerings.

De to feilaktige korrigeringsene for v220 og v314 er blant de 571 bedriftene som blir plukket ut til revisjon. Dermed blir disse feilkorrigeringsene rettet opp.

Av de 55 bedriftene som ikke skal tas ut til revisjon når de får korrigert sine tusenfeil, blir 20 likevel plukket ut til revisjon. De gjenværende 35 bedriftene får korrigert sine tusenfeil. Av de 469 bedriftene som skal tas ut til revisjon, blir 445 tatt ut. Dvs. at 24 bedrifter ikke blir tatt ut, og dette er de samme bedriftene som ikke ble plukket ut når vi ikke gjorde noen automatisk korrigerings for tusenfeil.

6. Oppsummering

I dette notatet har vi gitt et forslag på hvordan selektiv revisjon kan gjøres i strukturstatistikken. Seleksjonsmetoden vi har brukt er en metode for å plukke ut bedrifter som har store feil, dvs. bedrifter hvor original verdi er veldig forskjellig fra revidert verdi. (Merk at dette ikke nødvendigvis er de samme feilene som har størst betydning på de oppblåste populasjonstallene). Når metoden er tilpasset og implementert vil den være automatisk.

Generelt kan seleksjonsmetoden brukes i undersøkelser hvor variablene er numeriske, og hvor det fins tilleggdata som kan benyttes til å anslå revidert verdi (før selve revisjonen tar til). Hensikten med metoden er å plukke ut enheter hvor differansen mellom original og revidert verdi er stor, og jo bedre vi klarer å anslå de reviderte verdiene, jo mer treffsikker vil metoden være. Vi kan ikke kontrollere nøyaktig hvor mange enheter som plukkes ut vha metoden, men gjennom valg av parametrene får vi en viss innflytelse på hvor mange som tas ut.

Utprøvingen vi har gjort av metoden tyder på at den kan være brukbar for strukturstatistikken. Med den tilpassingen vi har valgt, klarer vi å plukke ut 500 av 524 bedriftene som har store feil. Dvs. at bare 24 av de bedriftene vi ønsker å ta ut til revisjon, ikke blir tatt ut. I tillegg til de 500 bedriftene med store feil, får vi plukket ut 106 bedrifter som kun har småfeil.

Vi har også veldig kort sett på muligheten for automatisk korreksjon av tusenfeil. Resultatet fra dette viser at vi kan få redusert antall bedrifter som skal tas ut til revisjon noe, ved først å gjøre en automatisk korrigerings for tusenfeil.

Hvis seleksjonsmetoden skal tas i bruk, er det en del ting som må avklares utover selve tilpassingen av metoden. F.eks.:

- Hvor mange ganger skal en bedrift kunne være med i utvalget uten å bli tatt ut til manuell revisjon
- Skal det gjøres noen form for automatisk justering for bedriftene som ikke tas ut til revisjon
- Hvordan skal vi kunne kontrollere at metoden fungerer etter sin hensikt

I tillegg er det viktig å lage et flaggingsopplegg som tar vare på nødvendige opplysninger fra revisjonsprosessen. Disse tingene er ikke omtalt i notatet, men må avklares før vi kan gå over fra full revisjon til selektiv revisjon.

Referanser

Abrahamsen, A. S., Andersen, M. Q. og Ragnarsøn, R. (2001). Project: Quality of Manufacturing Statistics and Evaluation of Methods for Producing Preliminary and Final Figures. Final Report. Documents 2001/8. Statistisk sentralbyrå.

Vedlegg A

Her beskriver vi hvordan revidert verdi for hver enkelt variabel blir anslått. Til dette bruker vi notasjon

a = anslått verdi

o = original verdi

f = forrige års reviderte verdi.

Vi har også laget to figurer for hver variabel. Den ene er et plott av original verdi mot revidert verdi (tilsvarer figur 4). Den andre figuren er et plott av delvis reviderte verdier mot reviderte verdier, når det er de 606 bedriftene som tas ut vha metoden som revideres (tilsvarer figur 7).

Fordi variabel v210 er godt illustrert i notatet, har vi ikke tatt den med i vedlegget.

For noen av variablene avhenger den anslåtte verdien av andre variablers anslåtte og originale verdier:

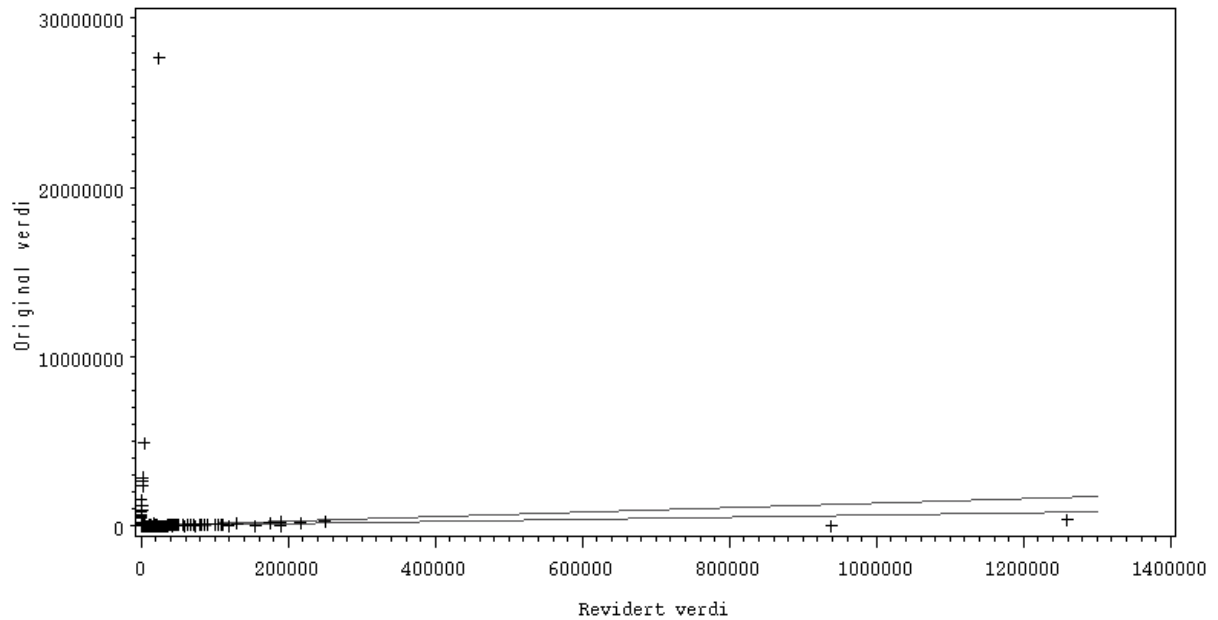
- v290 avhenger av v390
- v291 avhenger av v260 og v290
- v250 avhenger av v230, v240 og v291
- v310 og v314 avhenger av hverandre og av v220

For de fleste av variablene skal revidert verdi være positiv. Derfor må vi passe på at den anslåtte verdien ikke blir negativ for disse variablene. Når vi bruker forrige års reviderte verdi som anslag, er dette ikke noe problem fordi denne verdien alltid er positiv. Men når vi bruker original verdi må vi gjøre justeringer, fordi denne verdien kan være negativ. (Det kan hende at disse justeringene er litt for mye skreddersydd for det aktuelle datasettet vi bruker).

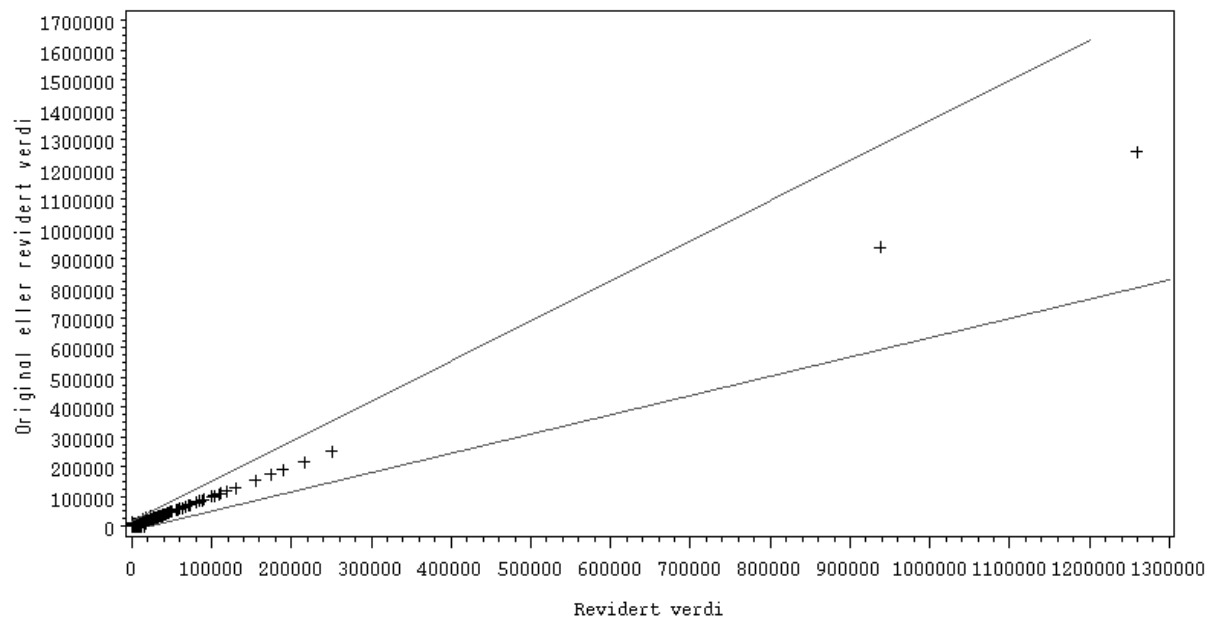
A.1 Variabel v220

Anslår den reviderte verdien med

$$a = f$$



Figur A1 Plott av original verdi mot revidert verdi (i tusen kroner). De to heltrukne linjene markerer grensene for når en feil er stor.

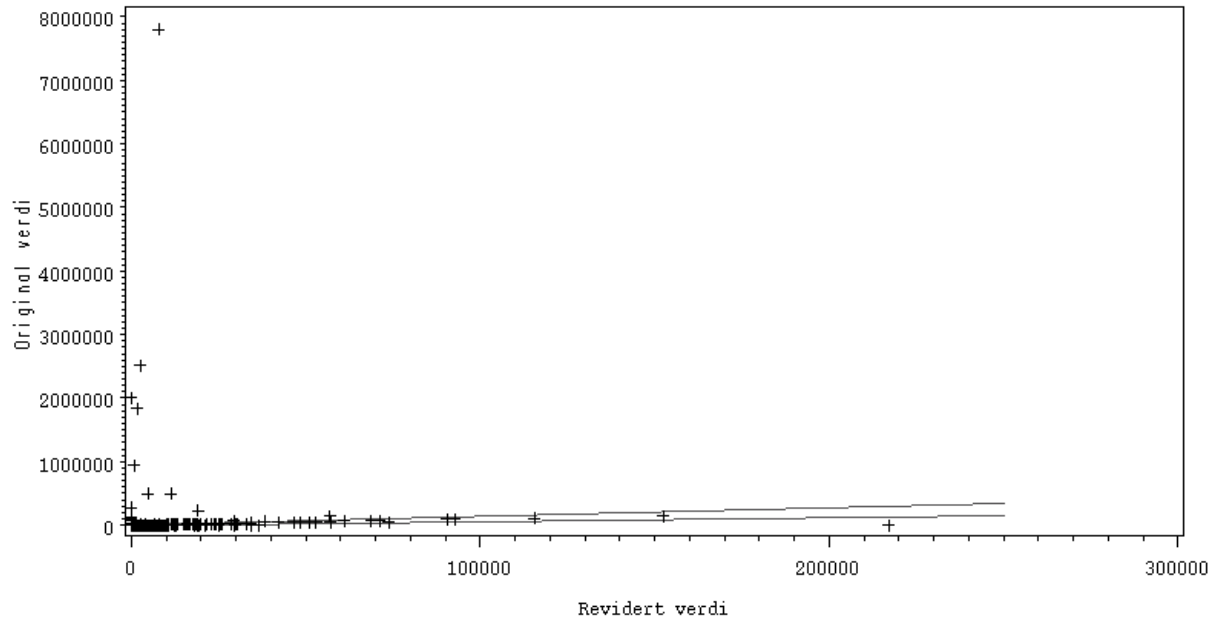


Figur A2 Plott av delvis reviderte verdier mot reviderte verdier (i tusen kroner), når det er de 606 bedriftene som plukkes ut vha metoden som blir revidert. De to heltrukne linjene markerer grensene for når en feil er stor.

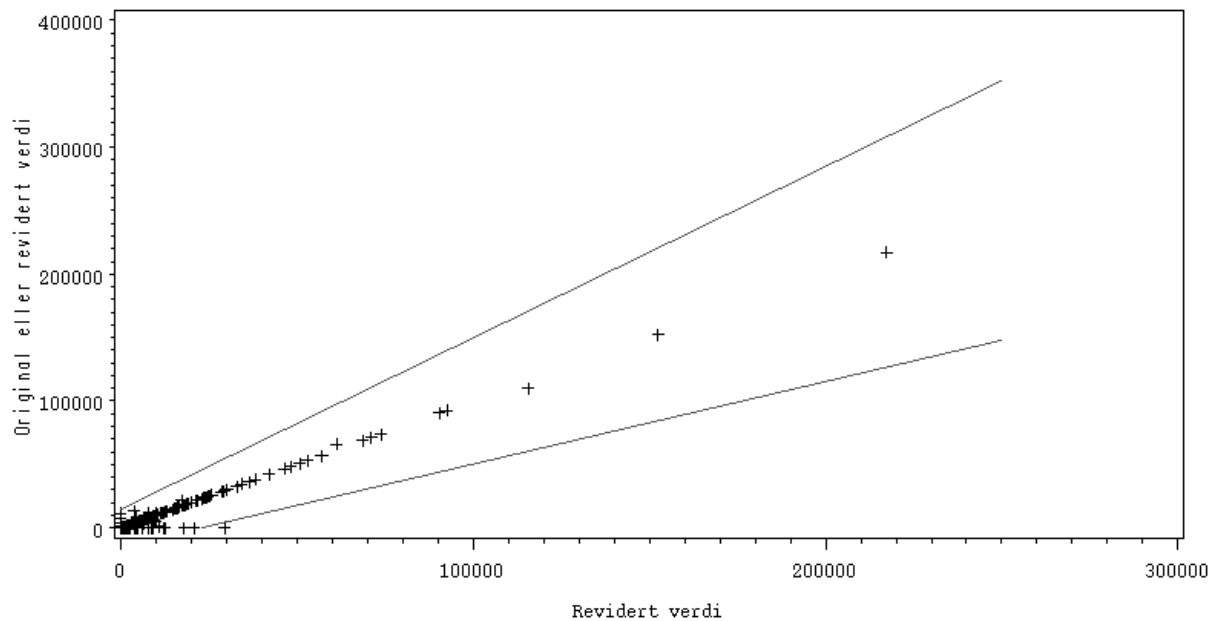
A.2 Variabel v230

Anslår den reviderte verdien med

$$a = f$$



Figur A3 Plott av original verdi mot revidert verdi (i tusen kroner). De to heltrukne linjene markerer grensene for når en feil er stor.

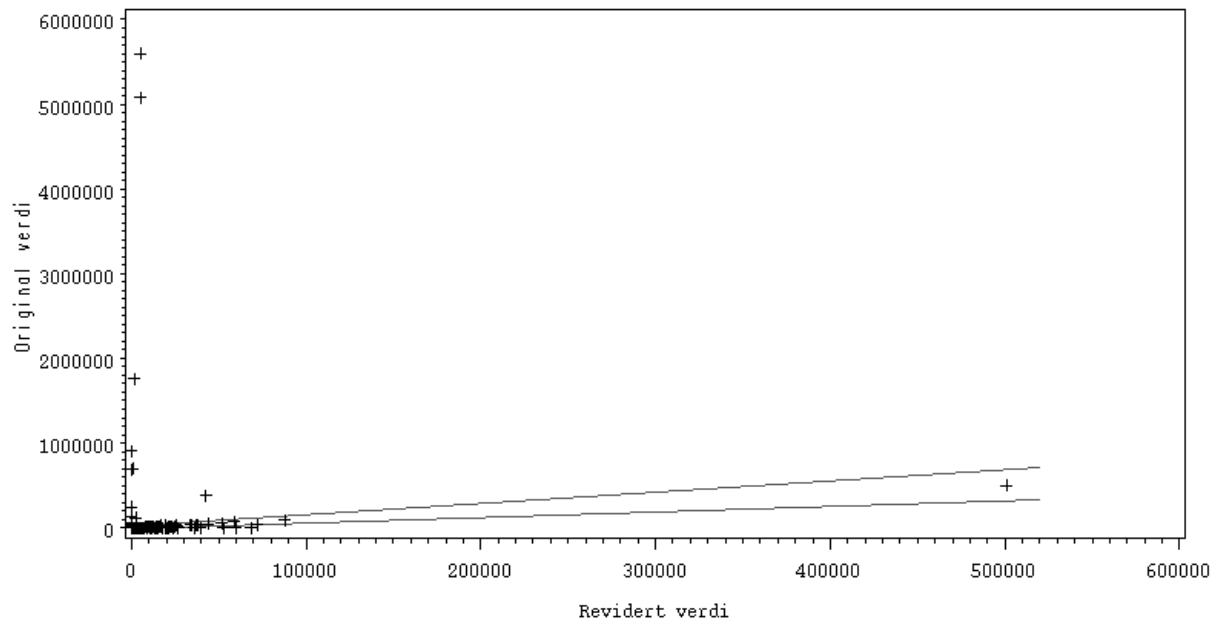


Figur A4 Plott av delvis reviderte verdier mot reviderte verdier (i tusen kroner), når det er de 606 bedriftene som plukkes ut vha metoden som blir revidert. De to heltrukne linjene markerer grensene for når en feil er stor.

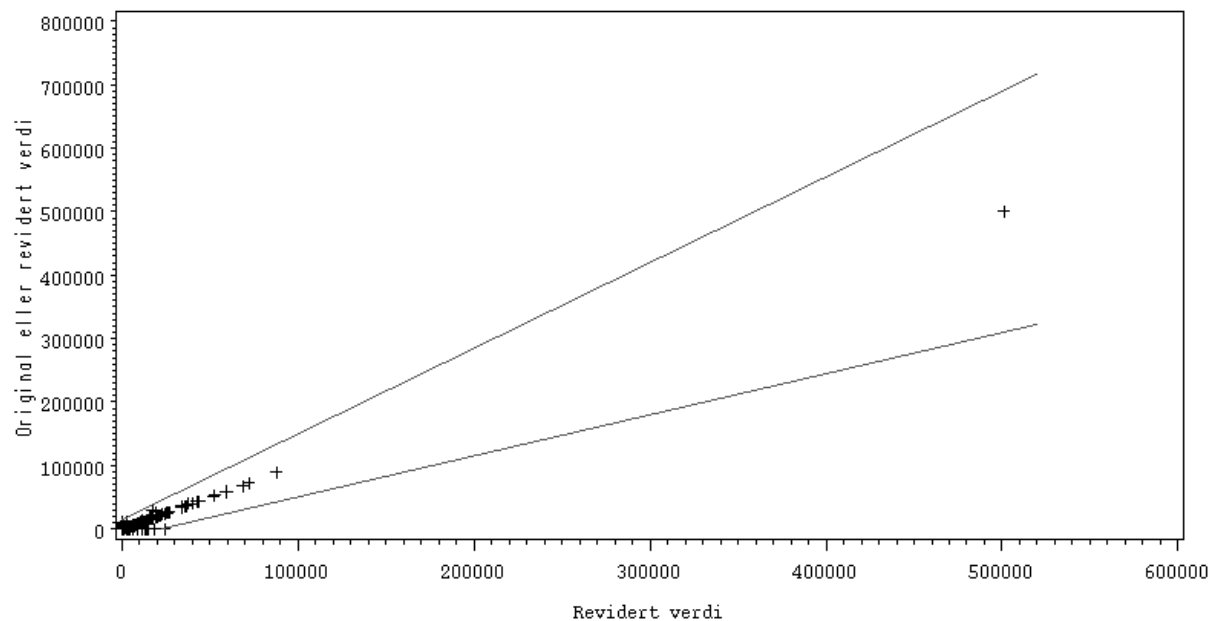
A.3 Variabel v240

Anslår den reviderte verdien med

$$a = f$$



Figur A5 Plott av original verdi mot revidert verdi (i tusen kroner). De to heltrukne linjene markerer grensene for når en feil er stor.



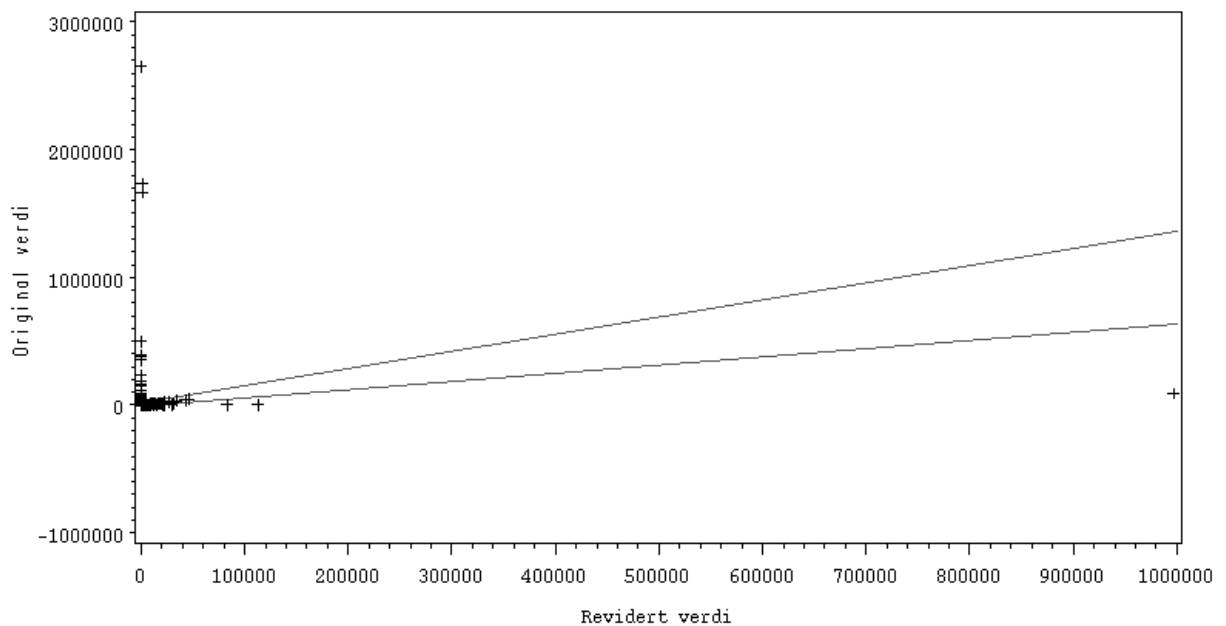
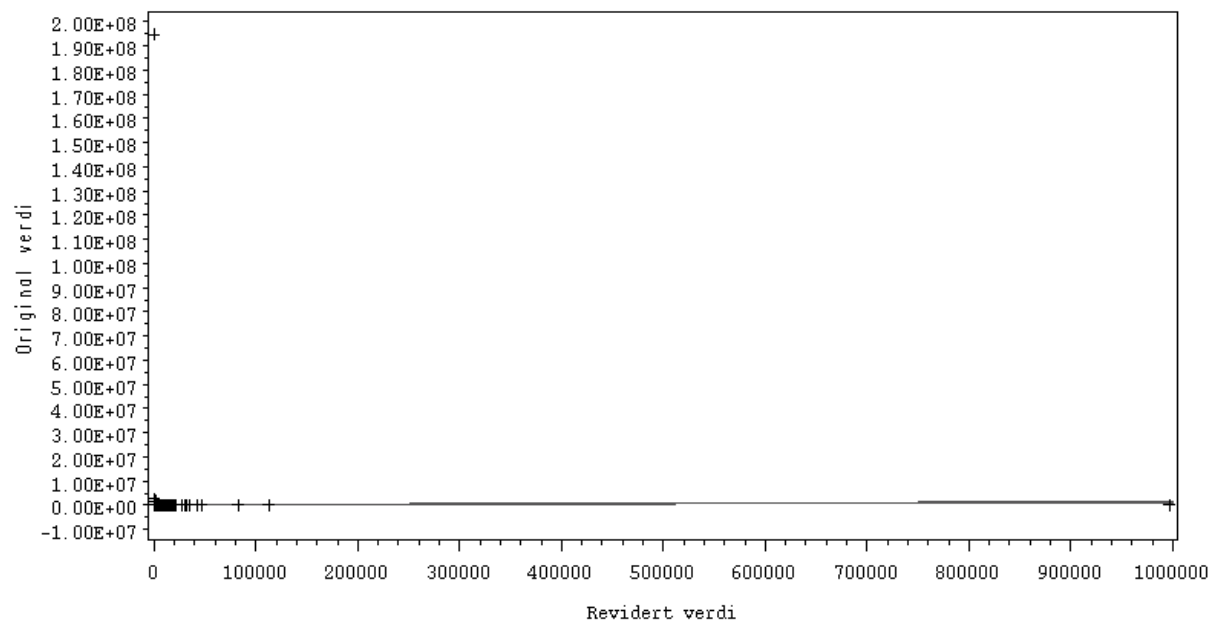
Figur A6 Plott av delvis reviderte verdier mot reviderte verdier (i tusen kroner), når det er de 606 bedriftene som plukkes ut vha metoden som blir revidert. De to heltrukne linjene markerer grensene for når en feil er stor.

A.4 Variabel v250

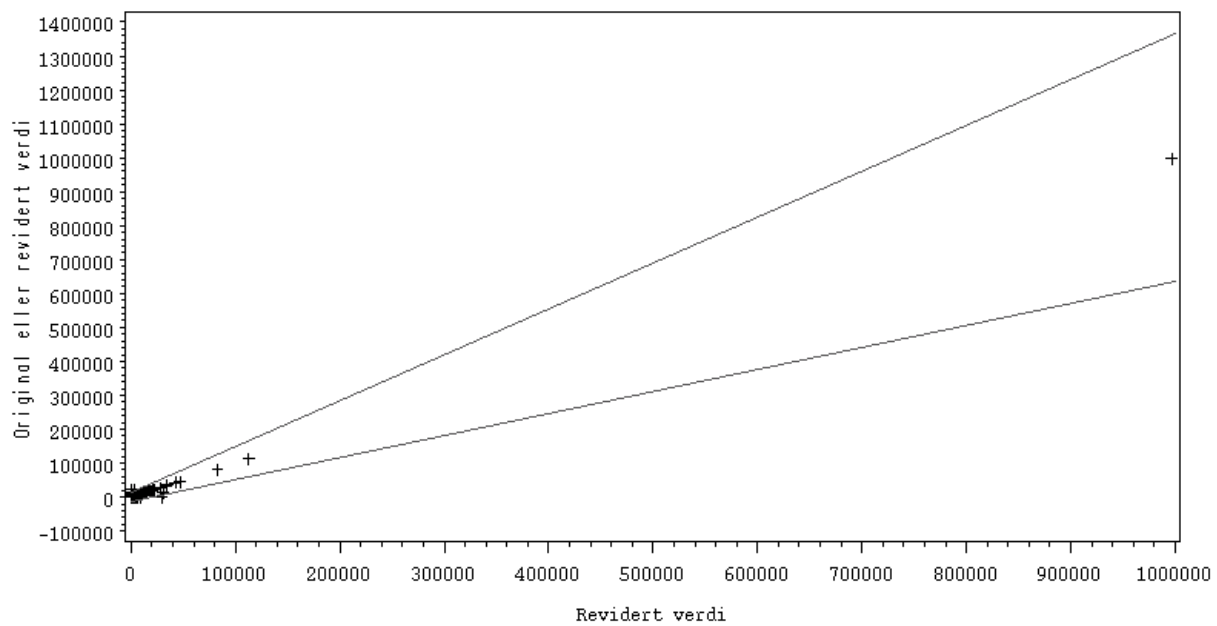
Anslår den reviderte verdien med

$$a = \begin{cases} 0 & \text{hvis } f = 0 \text{ eller } x = 0 \text{ eller } o \leq 0 \\ \min(f, x) & \text{hvis } o \geq x \\ o & \text{ellers} \end{cases},$$

der $x = \max(a_{v291} - a_{v230} - a_{v240}, 0)$, og a_{v230} , a_{v240} og a_{v291} er anslått verdi for henholdsvis v230, v240 og v291.



Figur A7 Plott av original verdi mot revidert verdi (i tusen kroner). De to heltrukne linjene markerer grensene for når en feil er stor. Det nederste plottet er et utsnitt av det over.



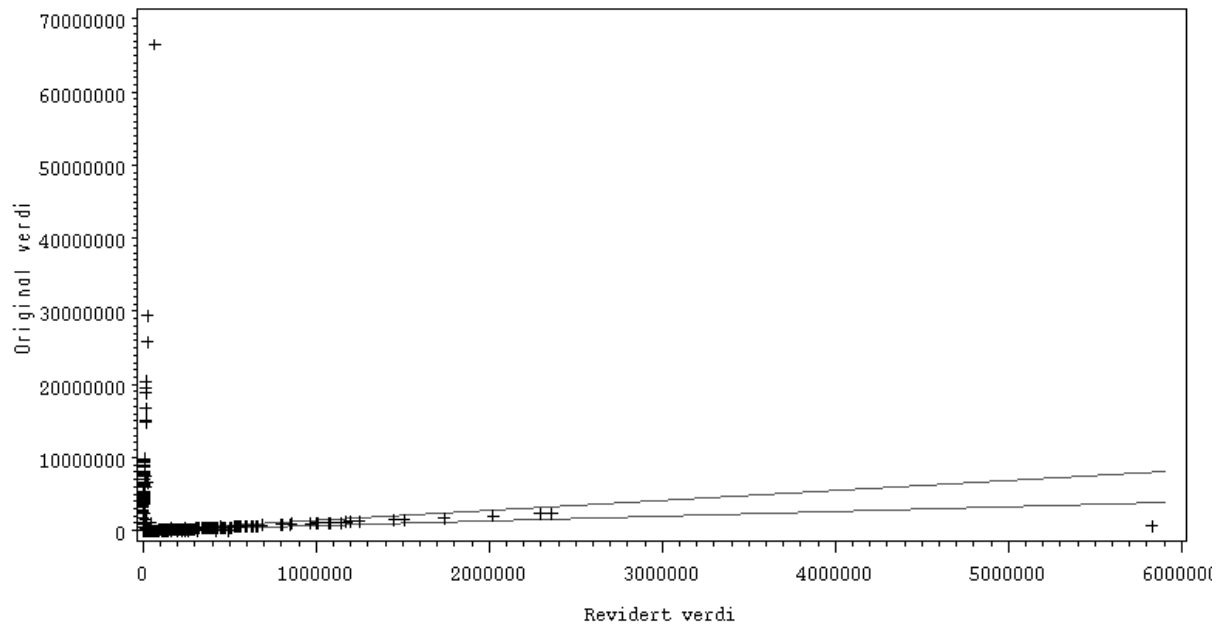
Figur A8 *Plott av delvis reviderte verdier mot reviderte verdier (i tusen kroner), når det er de 606 bedriftene som plukkes ut vha metoden som blir revidert. De to heltrukne linjene markerer grensene for når en feil er stor.*

A.5 Variabel v291

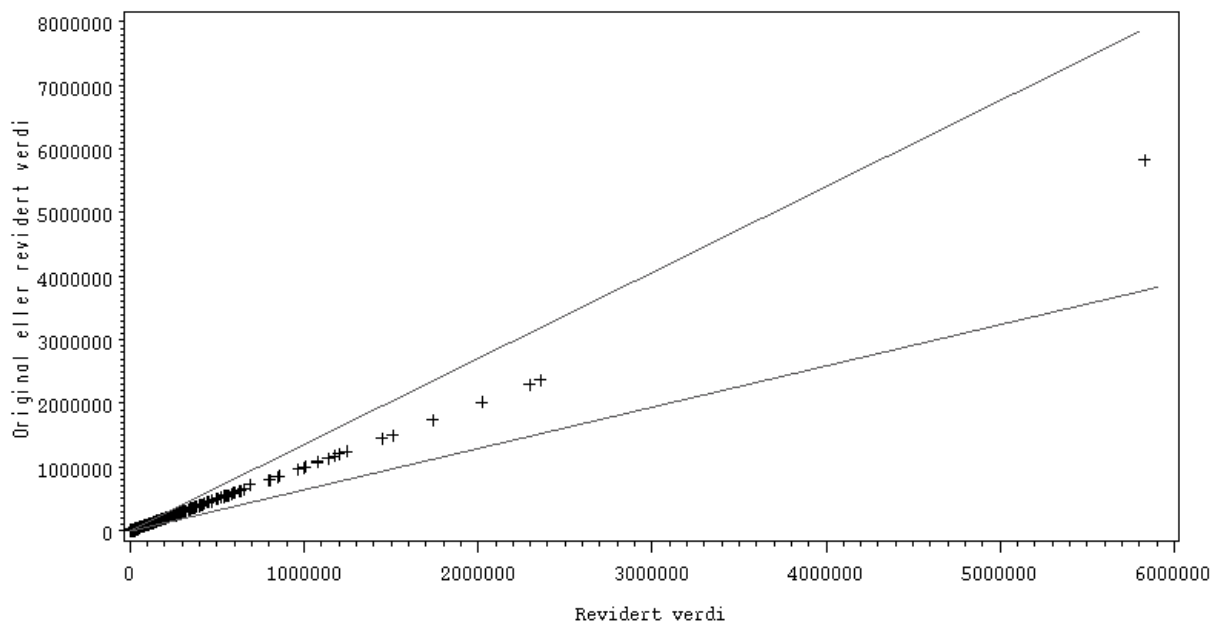
Anslår den reviderte verdien med

$$a = \text{maks}(a_{v290} - a_{v260}, 0),$$

der a_{v260} og a_{v290} er anslått verdi for henholdsvis v260 og v290.



Figur A9 Plott av original verdi mot revidert verdi (i tusen kroner). De to heltrukne linjene markerer grensene for når en feil er stor.

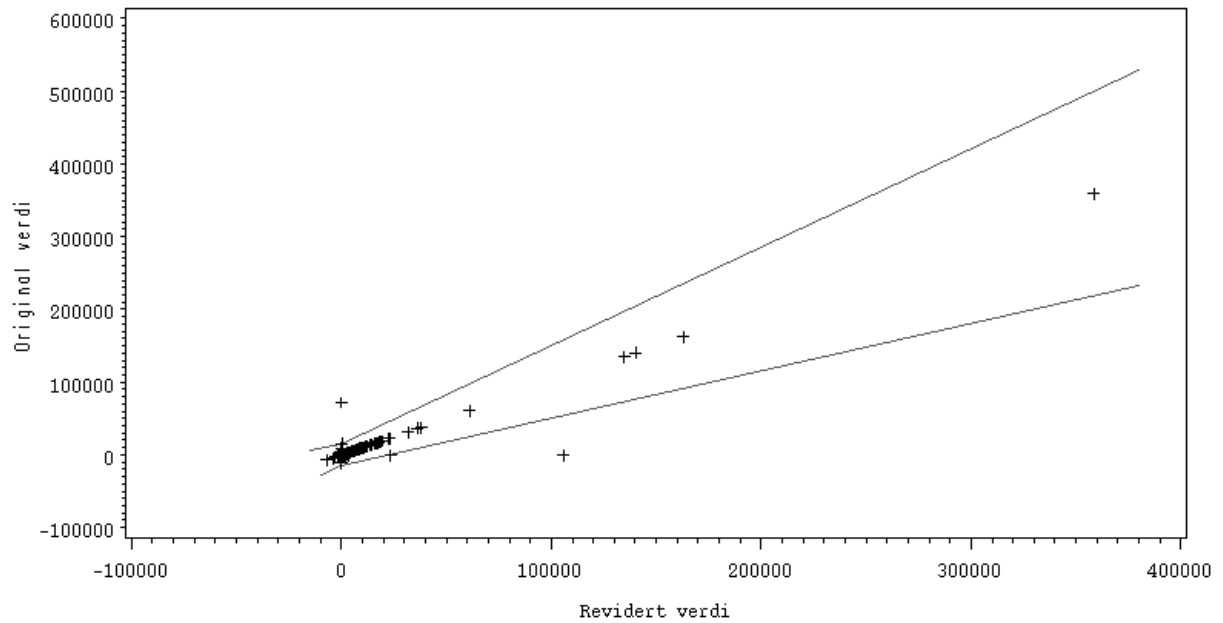


Figur A10 Plott av delvis reviderte verdier mot reviderte verdier (i tusen kroner), når det er de 606 bedriftene som plukkes ut vha metoden som blir revidert. De to heltrukne linjene markerer grensene for når en feil er stor.

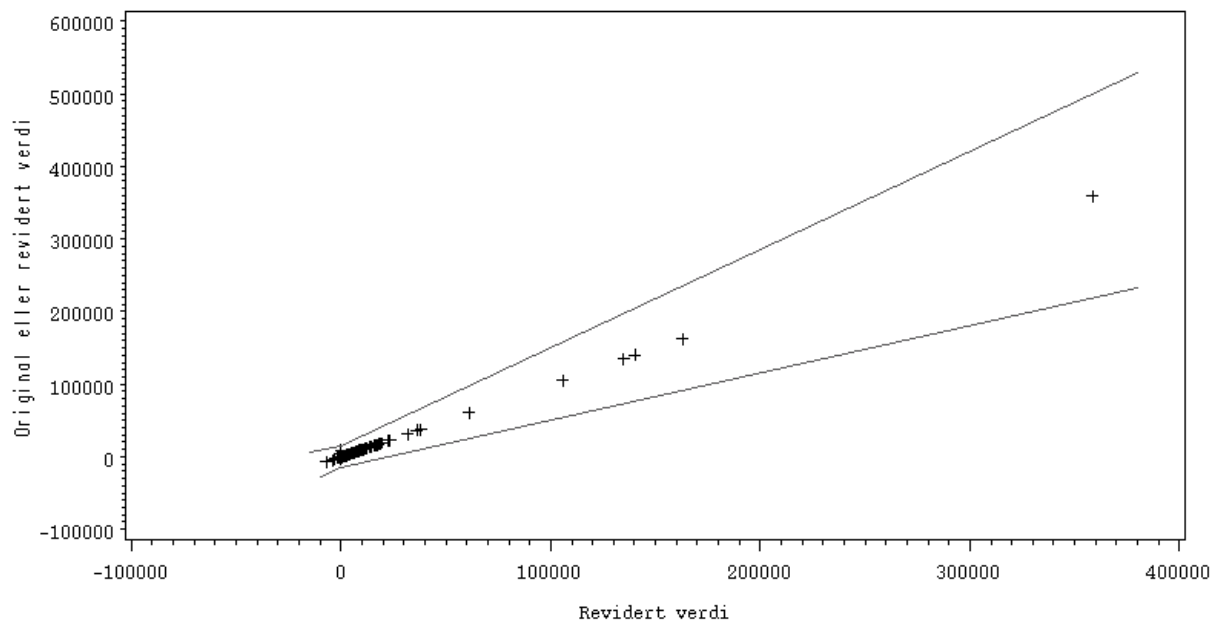
A.6 Variabel v260

Anslår den reviderte verdien med

$$a = 0$$



Figur A11 Plott av original verdi mot revidert verdi (i tusen kroner). De to heltrukne linjene markerer grensene for når en feil er stor.



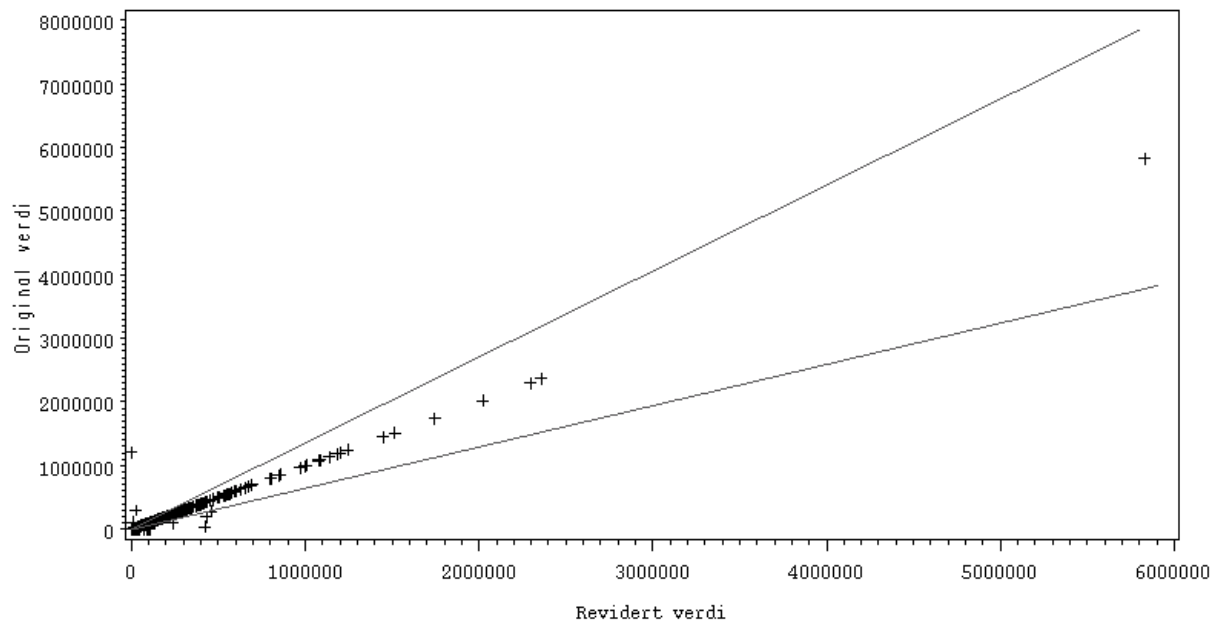
Figur A12 Plott av delvis reviderte verdier mot reviderte verdier (i tusen kroner), når det er de 606 bedriftene som plukkes ut vha metoden som blir revidert. De to heltrukne linjene markerer grensene for når en feil er stor.

A.7 Variabel v290

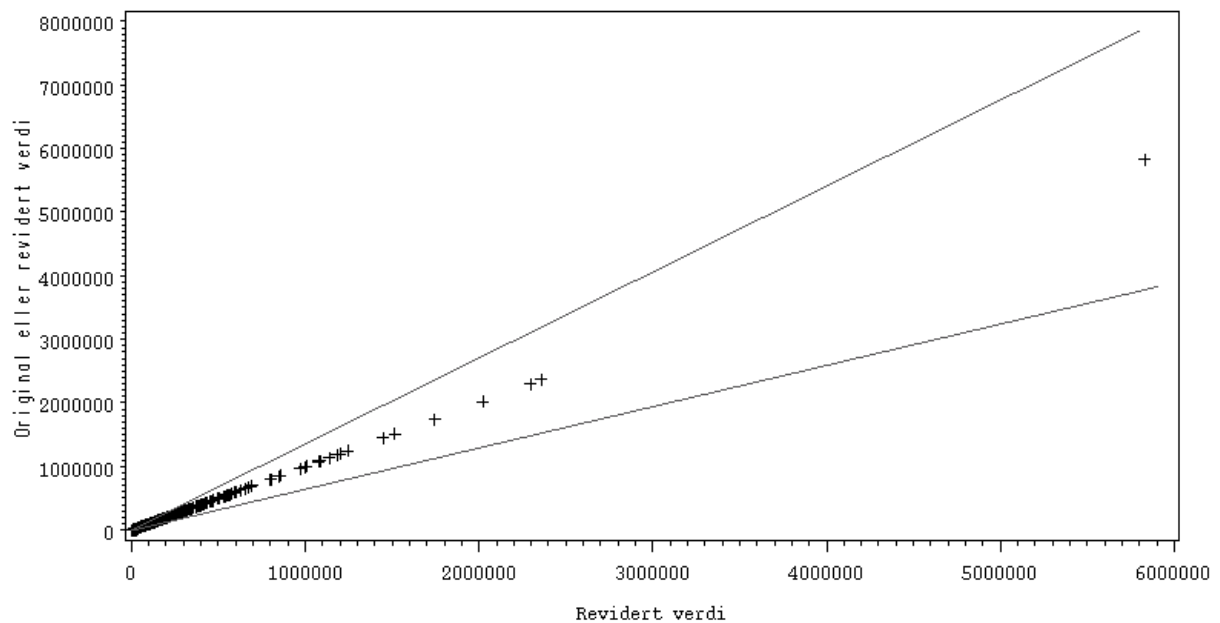
Anslår den reviderte verdien med

$$a = \begin{cases} o & \text{hvis } 0.7 \cdot a_{v390} \leq o \leq 4 \cdot a_{v390} \\ f & \text{ellers} \end{cases},$$

der a_{v390} er anslått verdi for v390.



Figur A13 Plott av original verdi mot revidert verdi (i tusen kroner). De to heltrukne linjene markerer grensene for når en feil er stor.



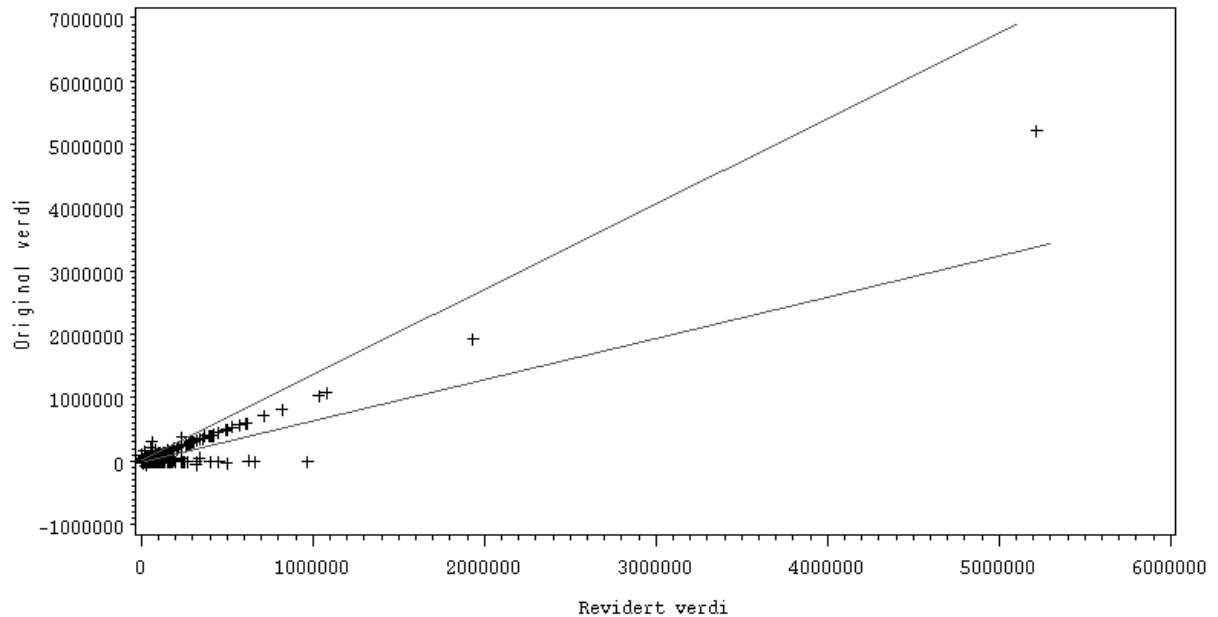
Figur A14 Plott av delvis reviderte verdier mot reviderte verdier (i tusen kroner), når det er de 606 bedriftene som plukkes ut vha metoden som blir revidert. De to heltrukne linjene markerer grensene for når en feil er stor.

A.8 Variabel v310

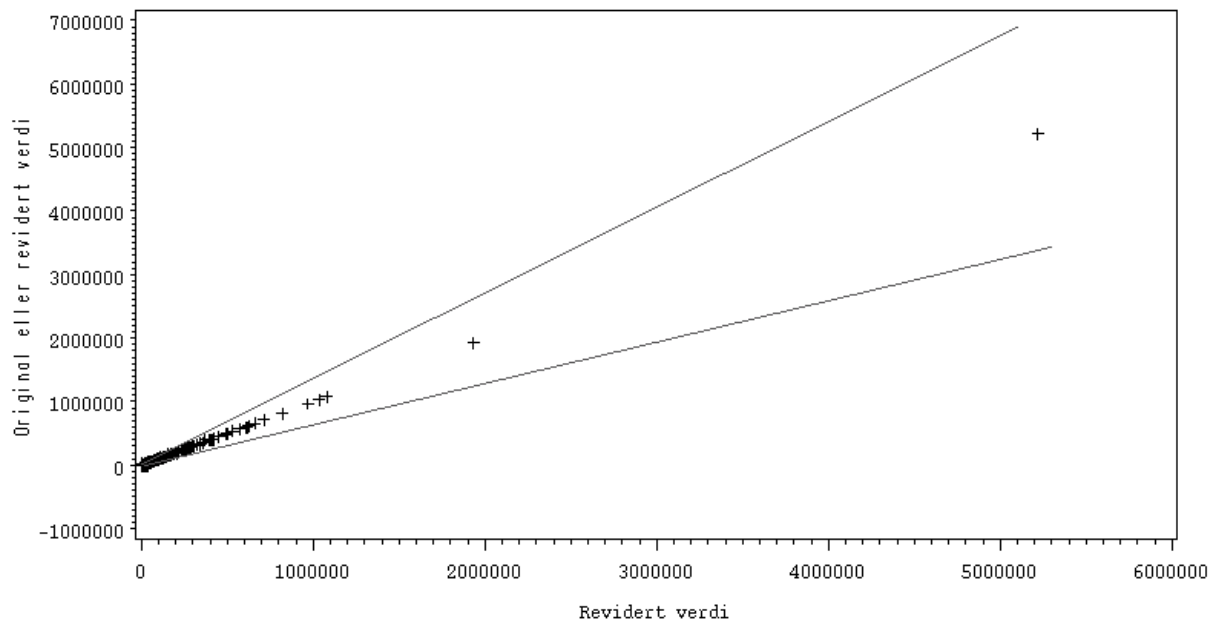
Anslår den reviderte verdien med

$$a = \begin{cases} 0 & \text{hvis } o_{v310} + o_{v314} \leq 0 \\ o_{v310} + o_{v314} & \text{hvis } 0 < o_{v310} + o_{v314} < 0.8 \cdot a_{v220} \\ o_{v310} + o_{v314} - 0.8 \cdot a_{v220} & \text{ellers} \end{cases},$$

der o_{v310} og o_{v314} er original verdi for henholdsvis v310 og v314, og a_{v220} er anslått verdi for v220.



Figur A15 Plott av original verdi mot revidert verdi (i tusen kroner). De to heltrukne linjene markerer grensene for når en feil er stor.



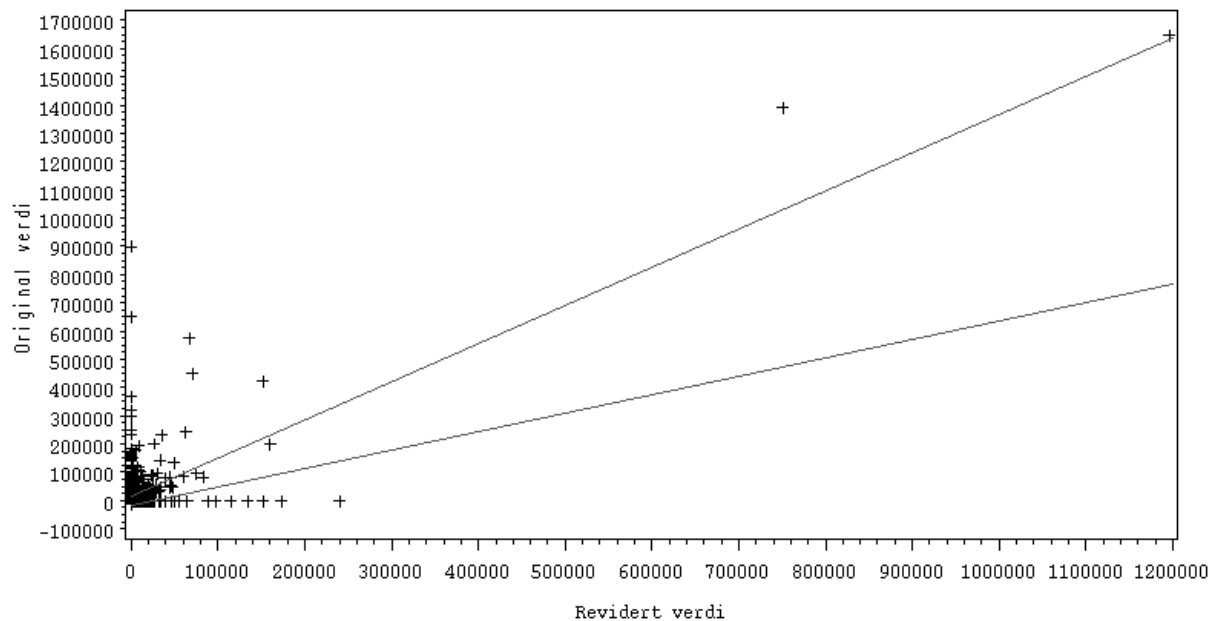
Figur A16 Plott av delvis reviderte verdier mot reviderte verdier (i tusen kroner), når det er de 606 bedriftene som plukkes ut vha metoden som blir revidert. De to heltrukne linjene markerer grensene for når en feil er stor.

A.9 Variabel v314

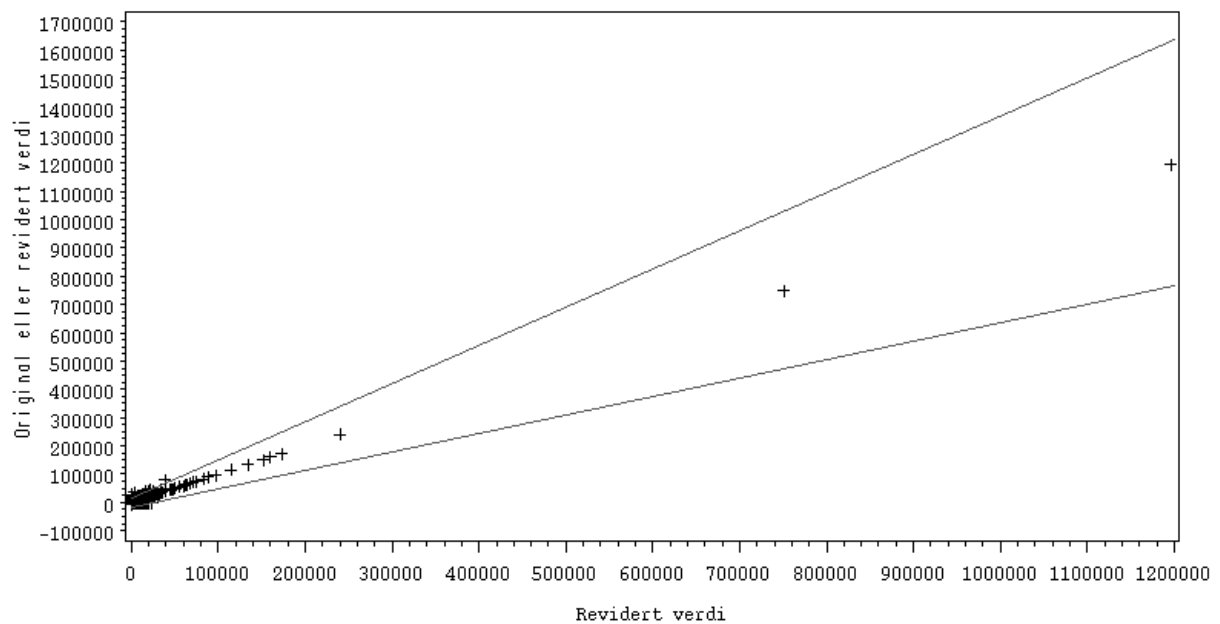
Anslår den reviderte verdien med

$$a = \begin{cases} 0.8 \cdot a_{v220} & \text{hvis } o_{v310} + o_{v314} \geq 0.8 \cdot a_{v220} \\ 0 & \text{ellers} \end{cases},$$

der o_{v310} og o_{v314} er original verdi for henholdsvis v310 og v314, og a_{v220} er anslått verdi for v220.



Figur A17 Plott av original verdi mot revidert verdi (i tusen kroner). De to heltrukne linjene markerer grensene for når en feil er stor.

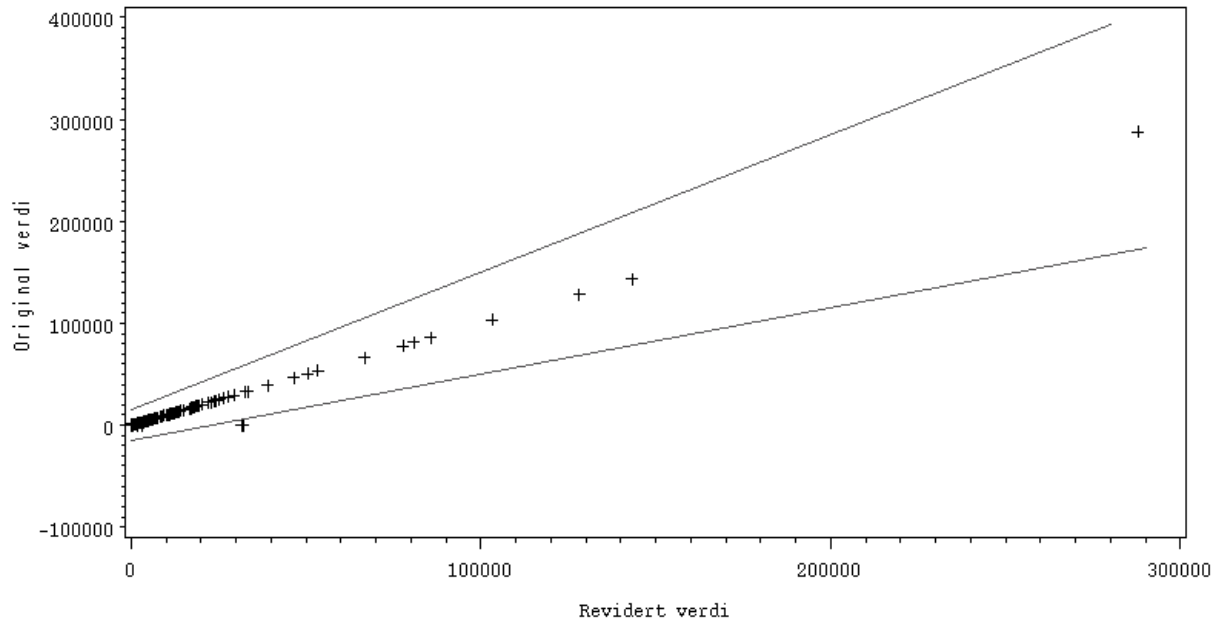


Figur A18 Plott av delvis reviderte verdier mot reviderte verdier (i tusen kroner), når det er de 606 bedriftene som plukkes ut vha metoden som blir revidert. De to heltrukne linjene markerer grensene for når en feil er stor.

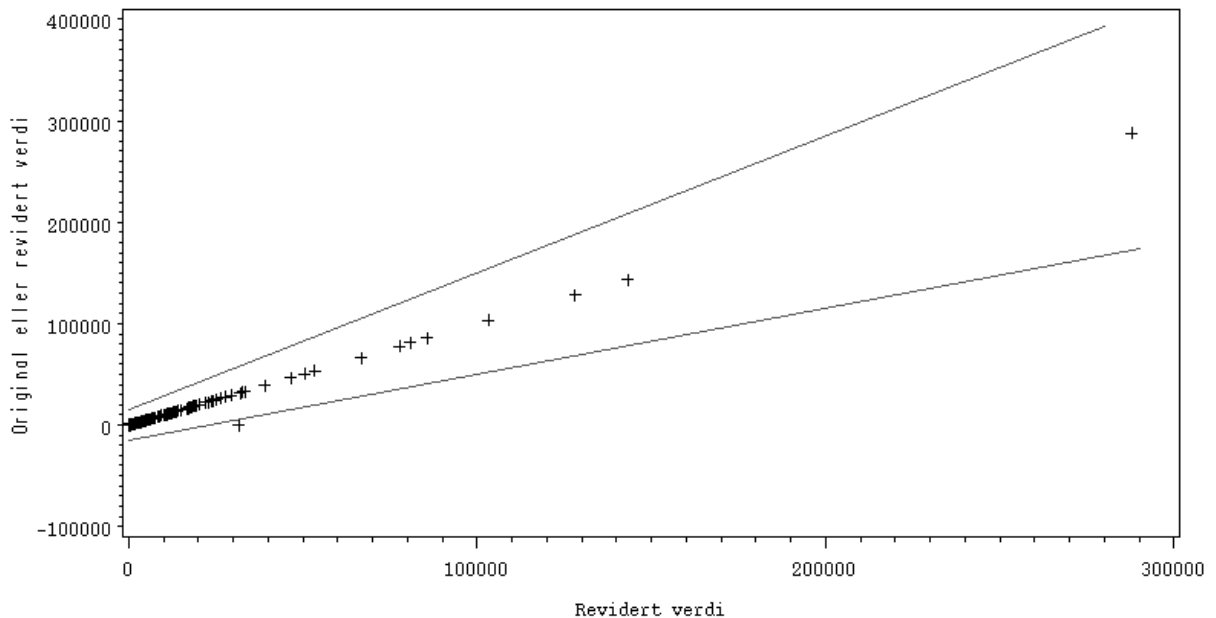
A.10 Variabel v316

Anslår den reviderte verdien med

$$a = \text{maks}(o, 0)$$



Figur A19 Plott av original verdi mot revidert verdi (i tusen kroner). De to heltrukne linjene markerer grensene for når en feil er stor.

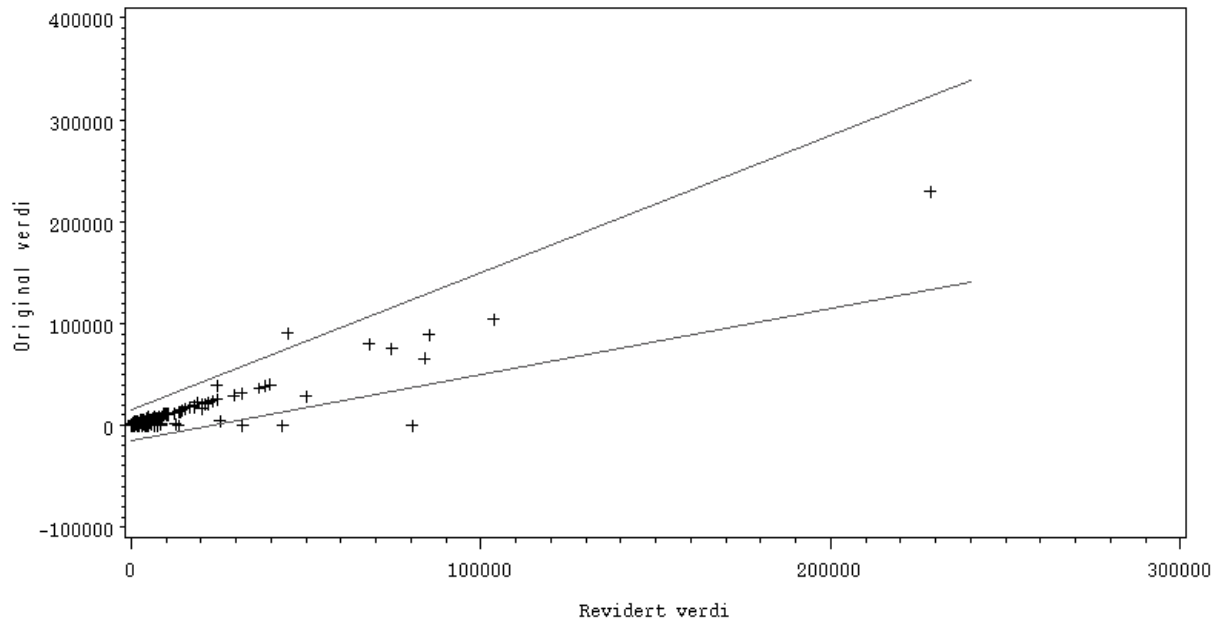


Figur A20 Plott av delvis reviderte verdier mot reviderte verdier (i tusen kroner), når det er de 606 bedriftene som plukkes ut vha metoden som blir revidert. De to heltrukne linjene markerer grensene for når en feil er stor.

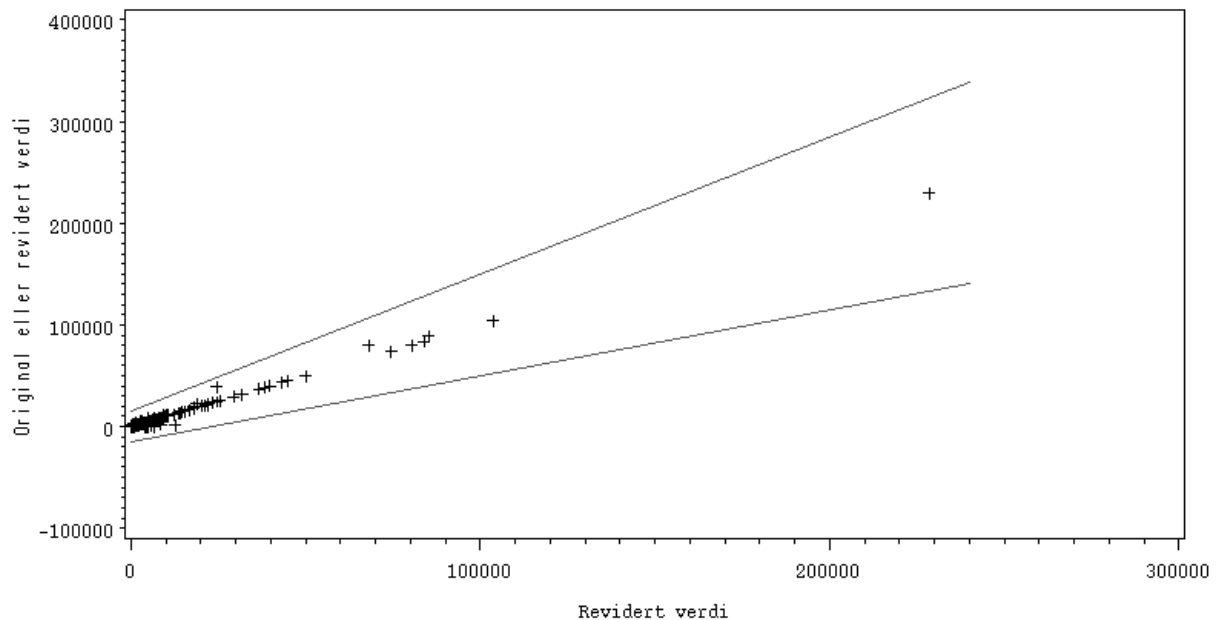
A.11 Variabel v320

Anslår den reviderte verdien med

$$a = f$$



Figur A21 Plott av original verdi mot revidert verdi (i tusen kroner). De to heltrukne linjene markerer grensene for når en feil er stor.

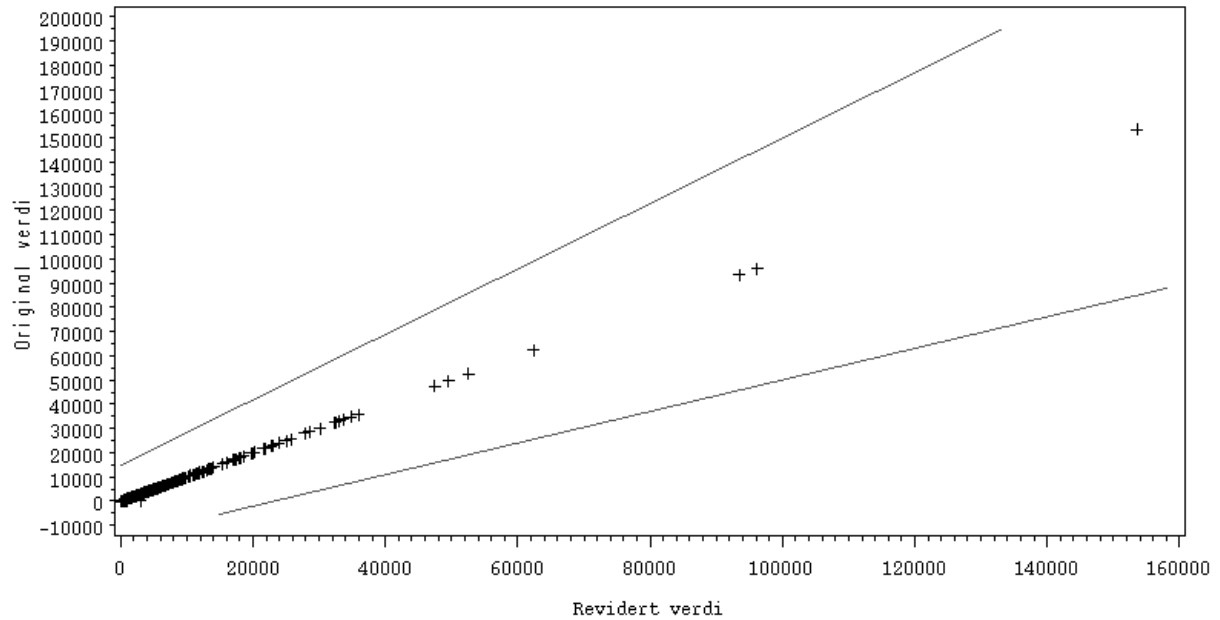


Figur A22 Plott av delvis reviderte verdier mot reviderte verdier (i tusen kroner), når det er de 606 bedriftene som plukkes ut vha metoden som blir revidert. De to heltrukne linjene markerer grensene for når en feil er stor.

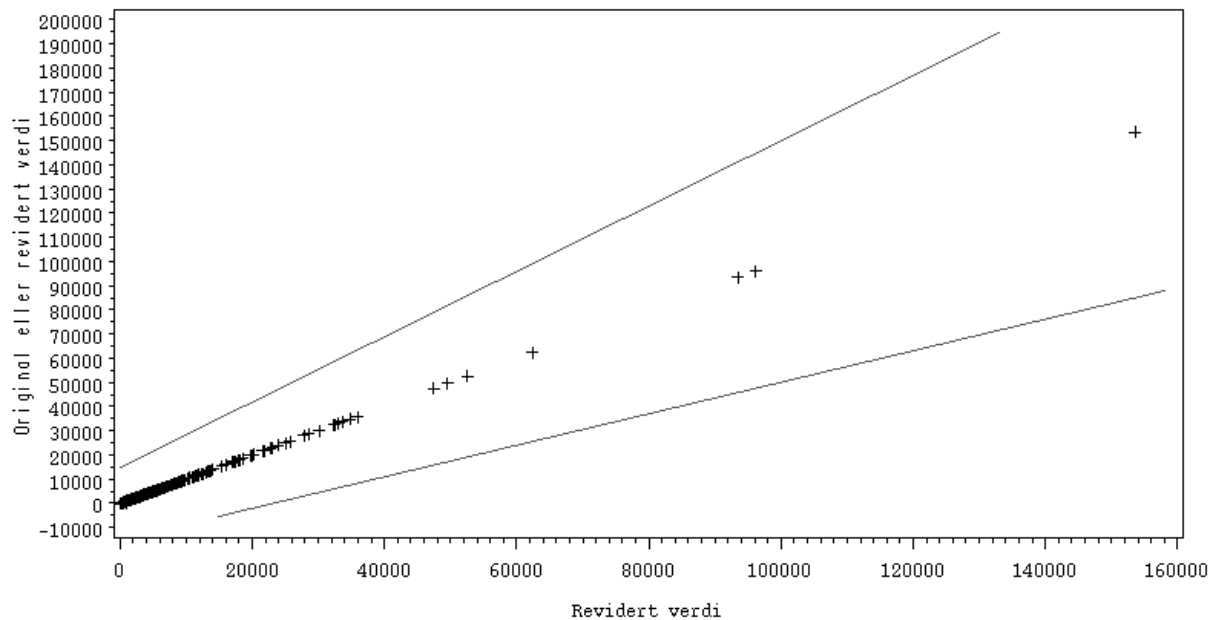
A.12 Variabel v330

Anslår den reviderte verdien med

$$a = \text{maks}(o, 0)$$



Figur A23 Plott av original verdi mot revidert verdi (i tusen kroner). De to heltrukne linjene markerer grensene for når en feil er stor.

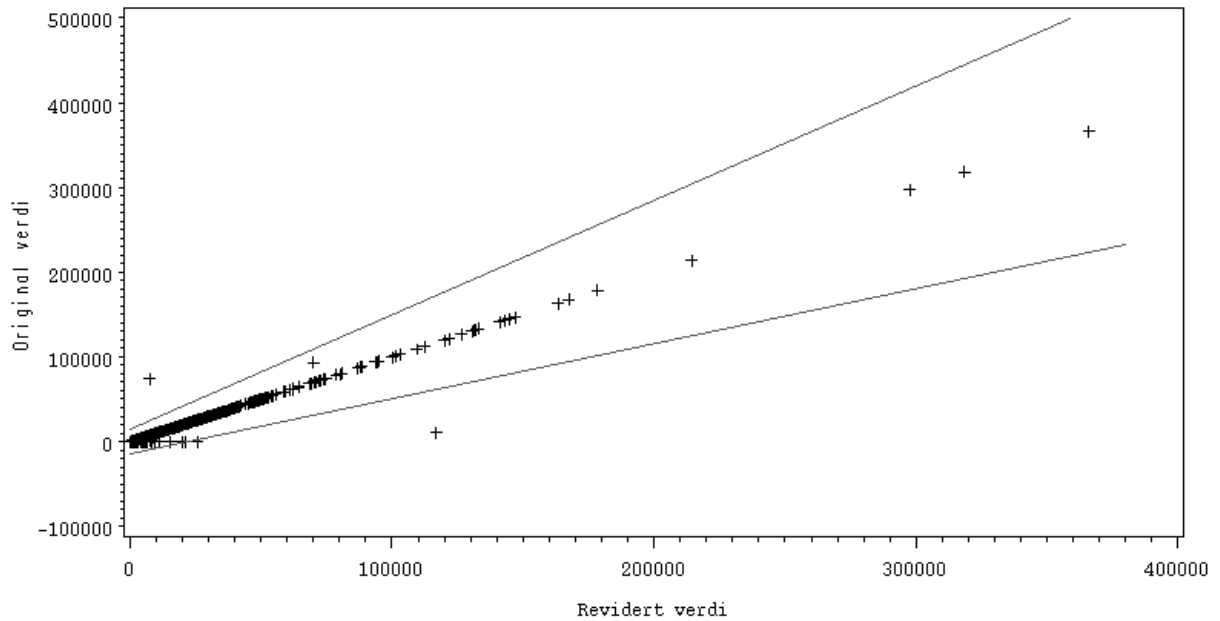


Figur A24 Plott av delvis reviderte verdier mot reviderte verdier (i tusen kroner), når det er de 606 bedriftene som plukkes ut vha metoden som blir revidert. De to heltrukne linjene markerer grensene for når en feil er stor.

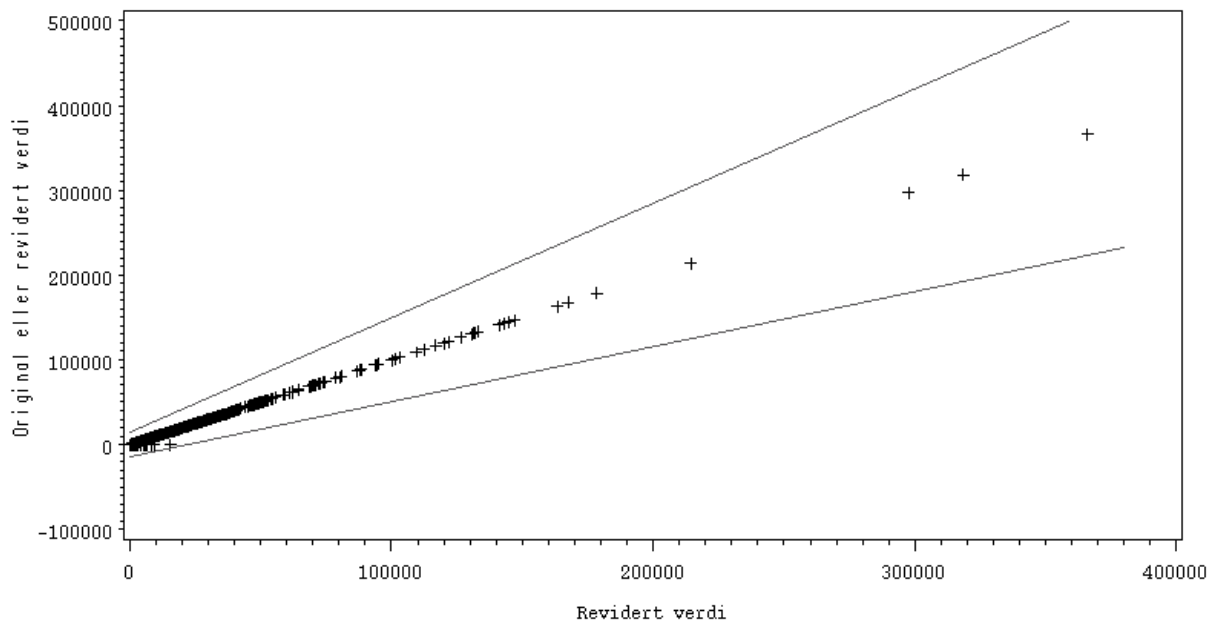
A.13 Variabel v340

Anslår den reviderte verdien med

$$a = \begin{cases} f & \text{hvis } o \leq 0 \\ o & \text{ellers} \end{cases}$$



Figur A25 Plott av original verdi mot revidert verdi (i tusen kroner). De to heltrukne linjene markerer grensene for når en feil er stor.

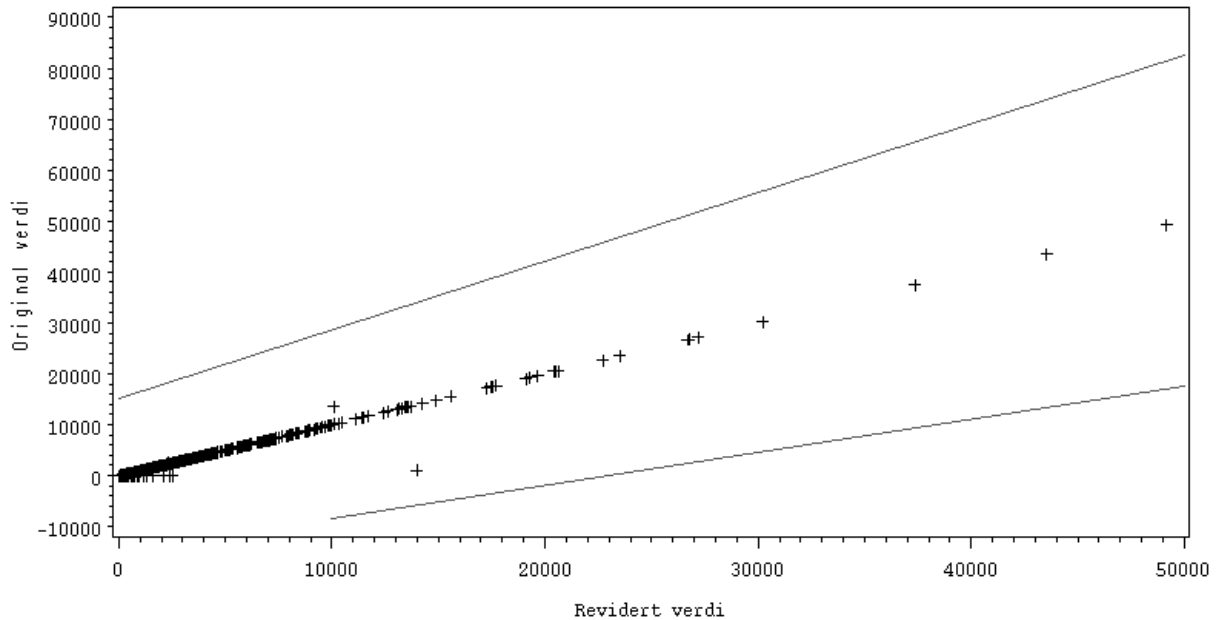


Figur A26 Plott av delvis reviderte verdier mot reviderte verdier (i tusen kroner), når det er de 606 bedriftene som plukkes ut vha metoden som blir revidert. De to heltrukne linjene markerer grensene for når en feil er stor.

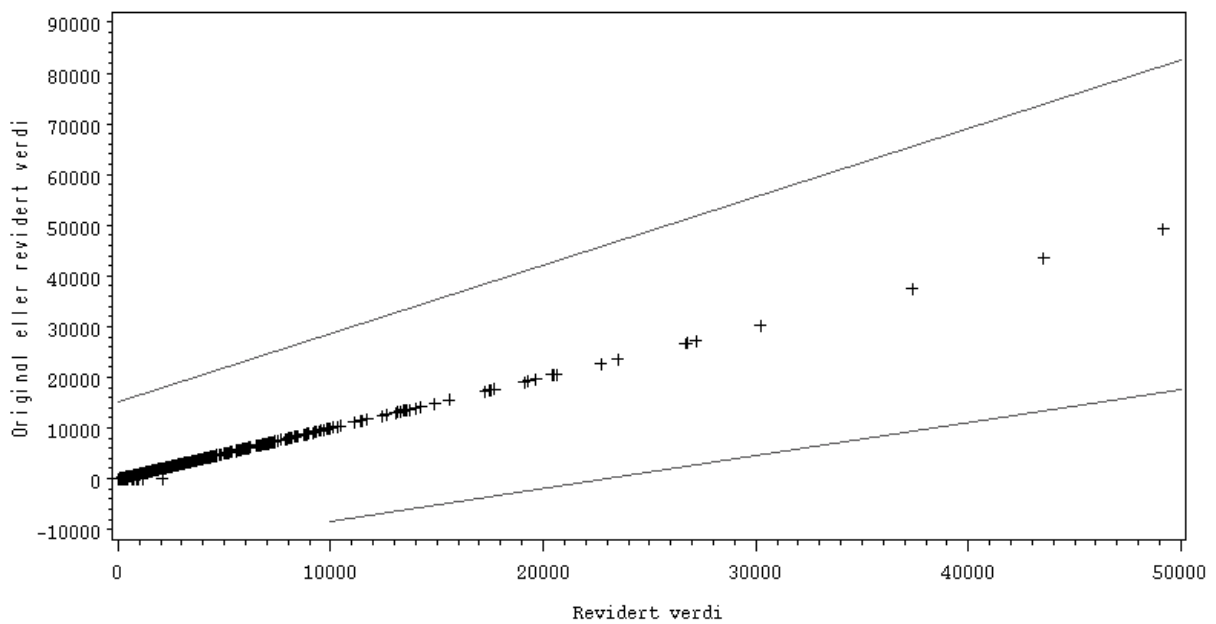
A.14 Variabel v350

Anslår den reviderte verdien med

$$a = \begin{cases} f & \text{hvis } o \leq 0 \\ o & \text{ellers} \end{cases}$$



Figur A27 Plott av original verdi mot revidert verdi (i tusen kroner). De to heltrukne linjene markerer grensene for når en feil er stor.

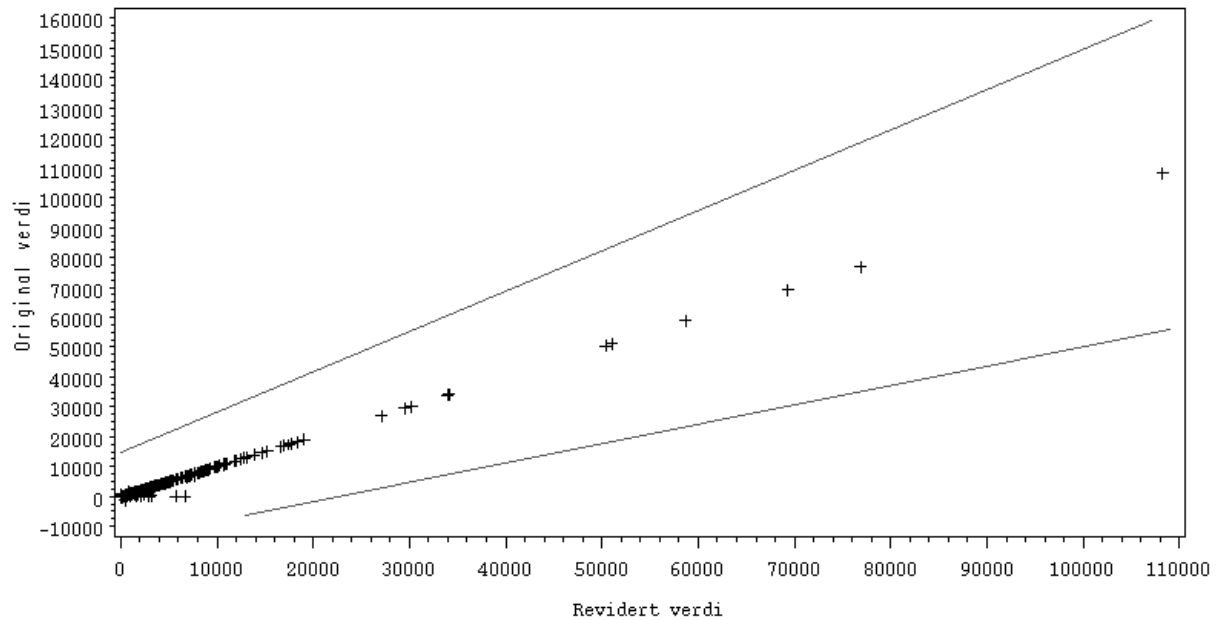


Figur A28 Plott av delvis reviderte verdier mot reviderte verdier (i tusen kroner), når det er de 606 bedriftene som plukkes ut vha metoden som blir revidert. De to heltrukne linjene markerer grensene for når en feil er stor.

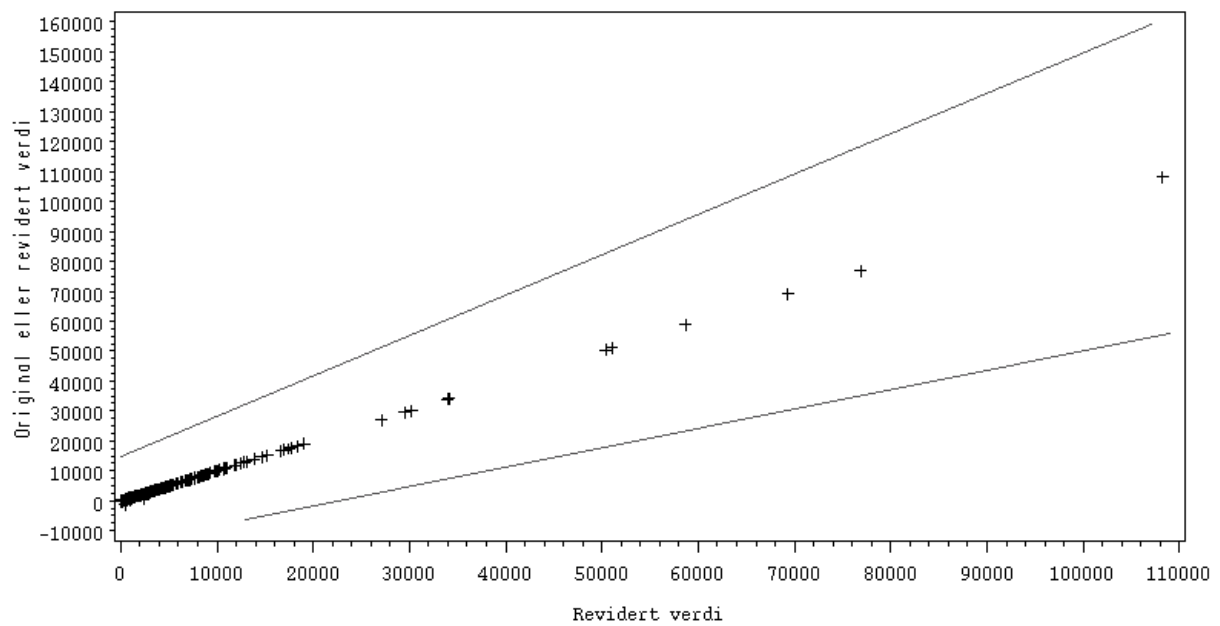
A.15 Variabel v360

Anslår den reviderte verdien med

$$a = \begin{cases} f & \text{hvis } o < 0 \\ o & \text{ellers} \end{cases}$$



Figur A29 Plott av original verdi mot revidert verdi (i tusen kroner). De to heltrukne linjene markerer grensene for når en feil er stor.

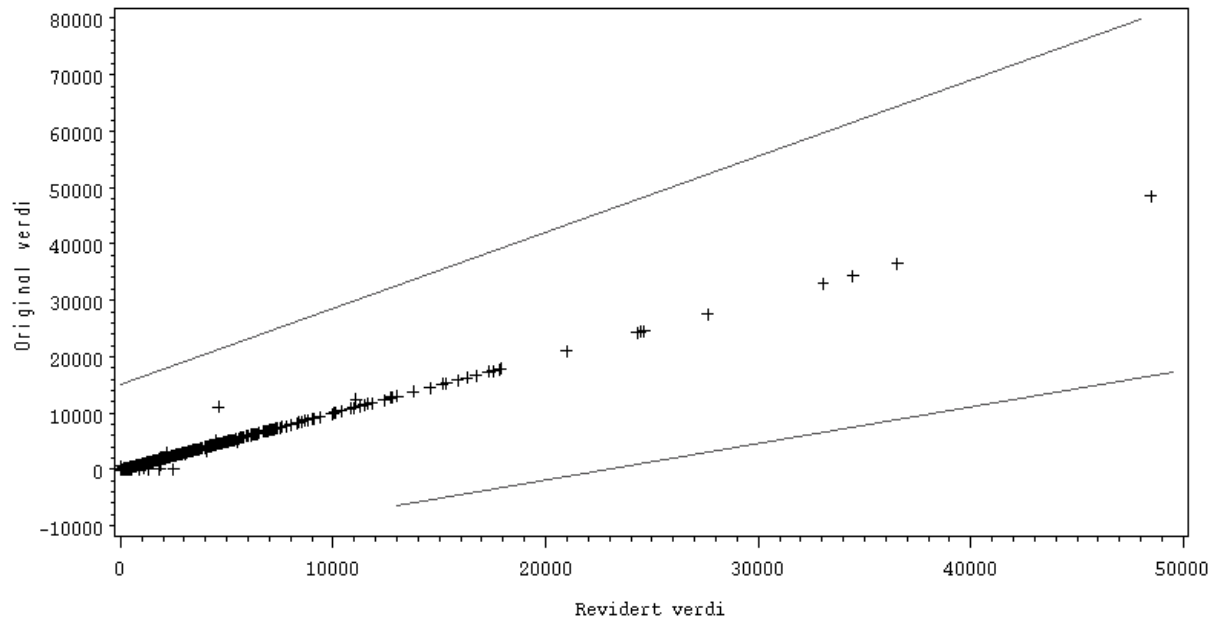


Figur A30 Plott av delvis reviderte verdier mot reviderte verdier (i tusen kroner), når det er de 606 bedriftene som plukkes ut vha metoden som blir revidert. De to heltrukne linjene markerer grensene for når en feil er stor.

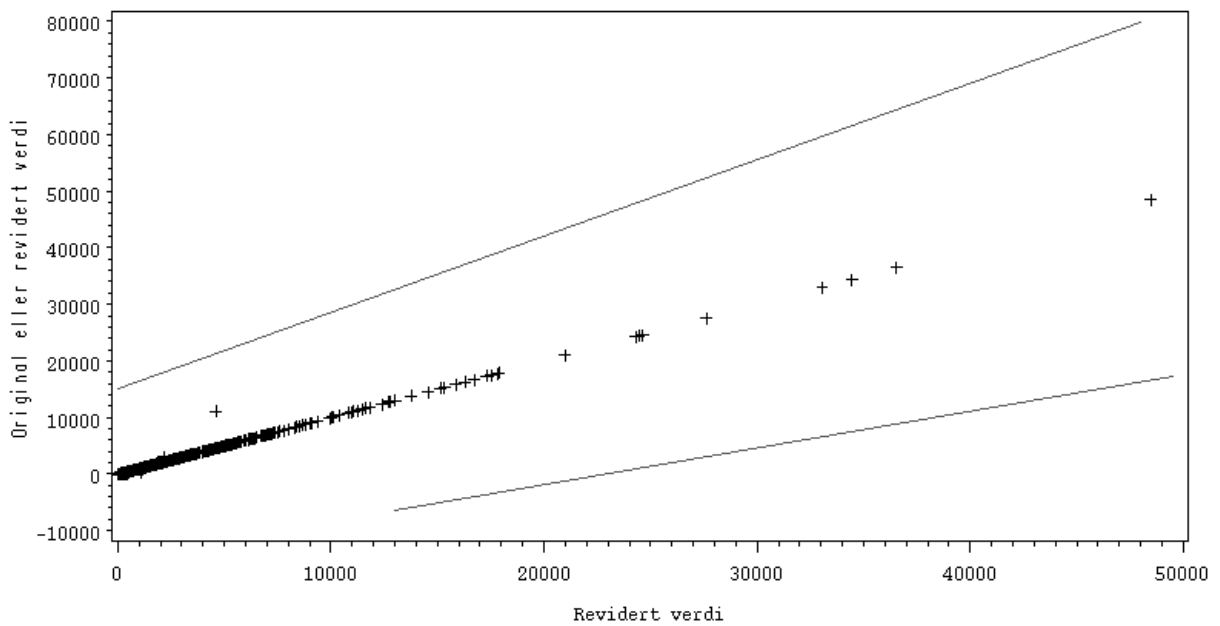
A.16 Variabel v370

Anslår den reviderte verdien med

$$a = \begin{cases} f & \text{hvis } o < 0 \\ o & \text{ellers} \end{cases}$$



Figur A31 Plott av original verdi mot revidert verdi (i tusen kroner). De to heltrukne linjene markerer grensene for når en feil er stor.

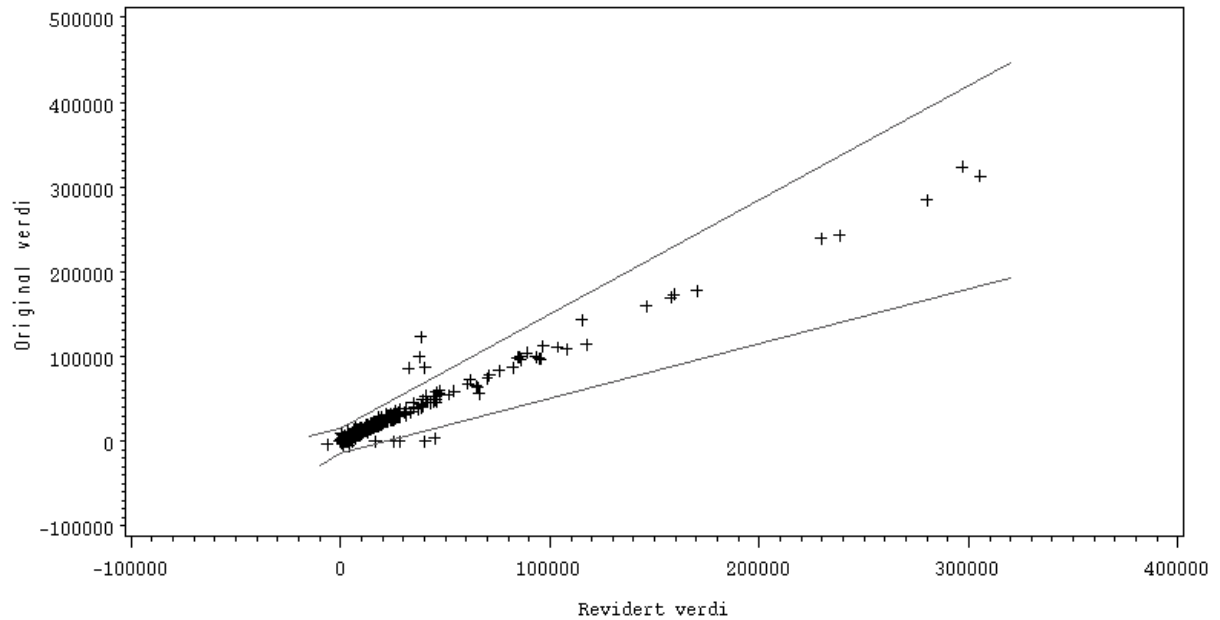


Figur A32 Plott av delvis reviderte verdier mot reviderte verdier (i tusen kroner), når det er de 606 bedriftene som plukkes ut vha metoden som blir revidert. De to heltrukne linjene markerer grensene for når en feil er stor.

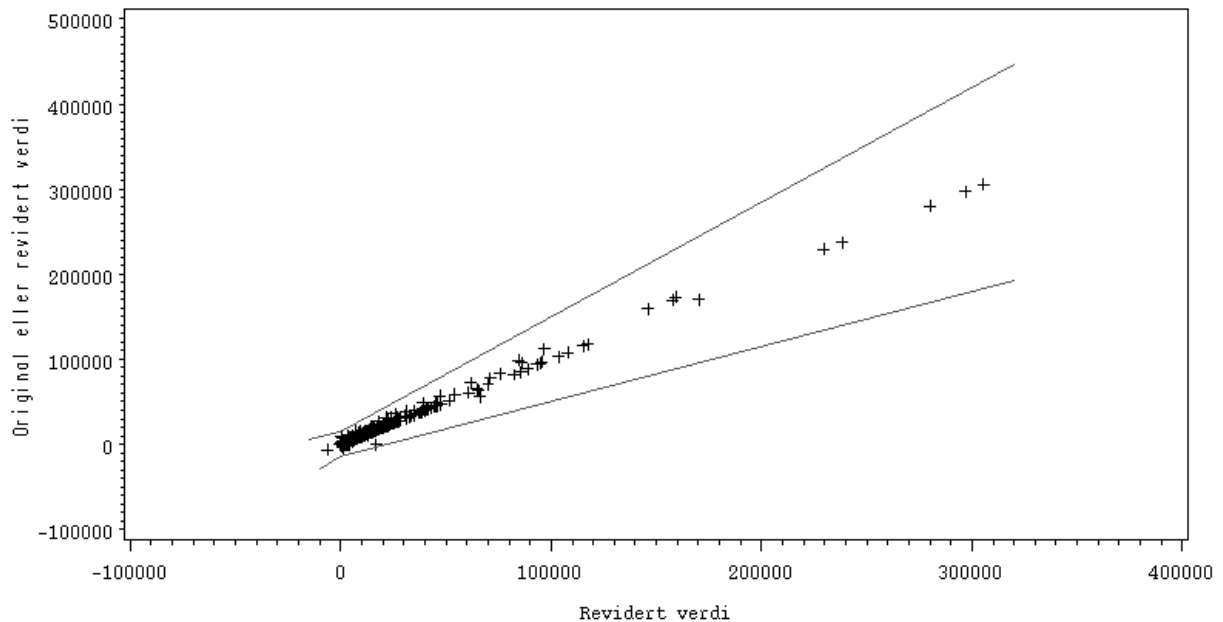
A.17 Variabel v380

Anslår den reviderte verdien med

$$a = f$$



Figur A33 Plott av original verdi mot revidert verdi (i tusen kroner). De to heltrukne linjene markerer grensene for når en feil er stor.

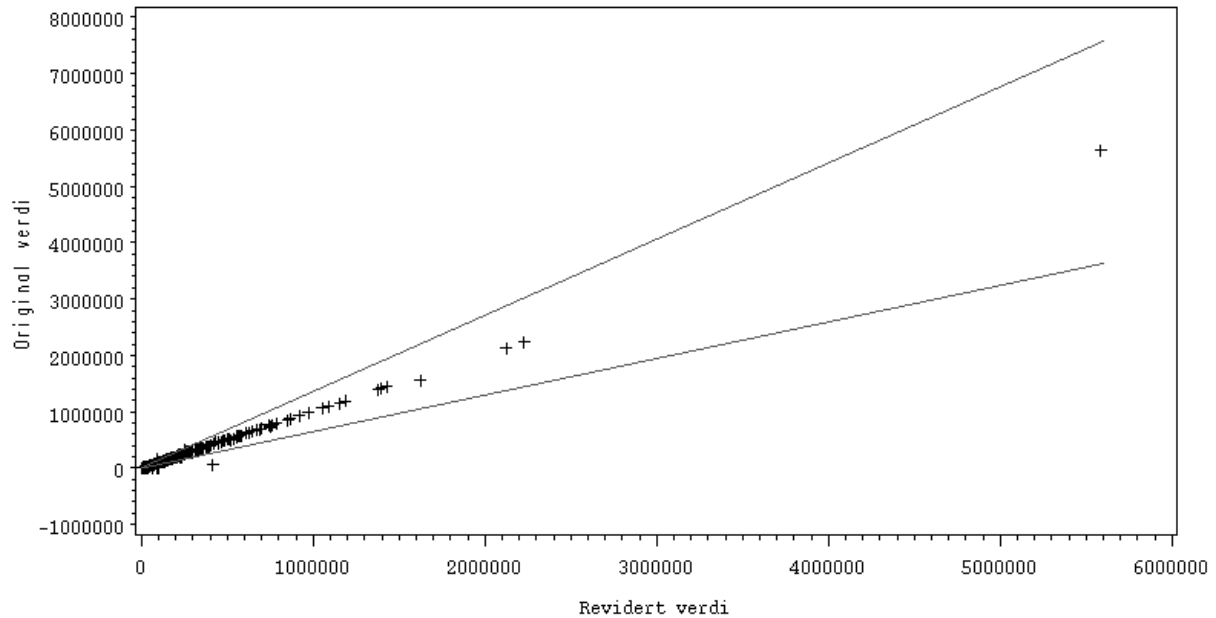


Figur A34 Plott av delvis reviderte verdier mot reviderte verdier (i tusen kroner), når det er de 606 bedriftene som plukkes ut vha metoden som blir revidert. De to heltrukne linjene markerer grensene for når en feil er stor.

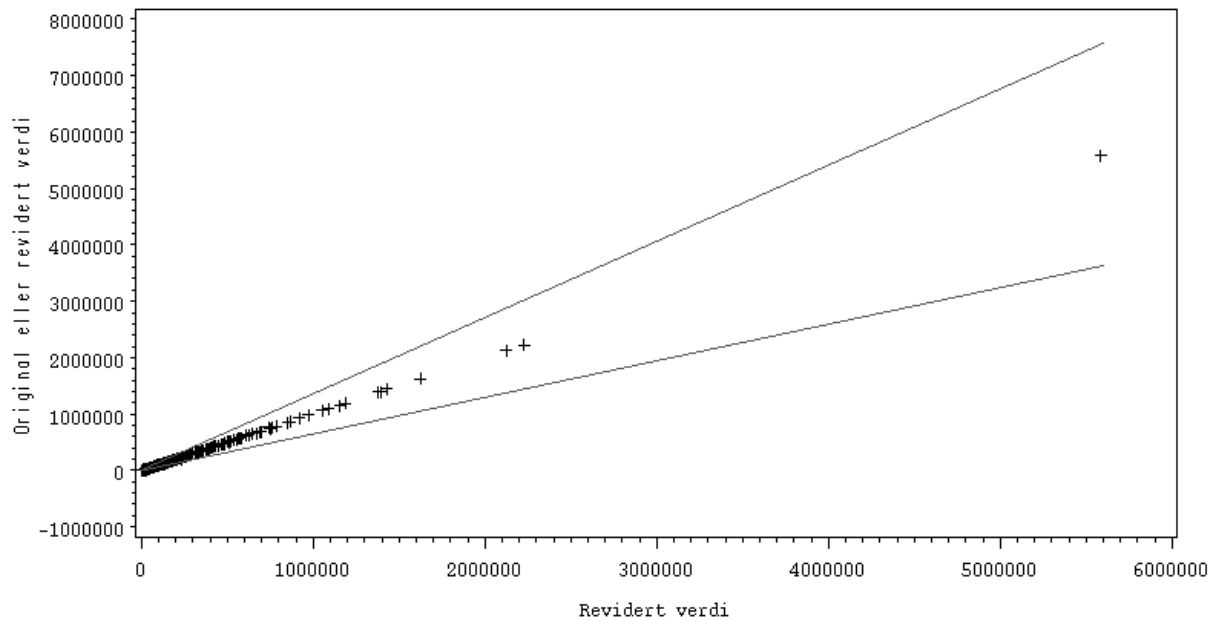
A.18 Variabel v390

Anslår den reviderte verdien med

$$a = \begin{cases} f & \text{hvis } o \leq 0 \\ o & \text{ellers} \end{cases}$$



Figur A35 Plott av original verdi mot revidert verdi (i tusen kroner). De to heltrukne linjene markerer grensene for når en feil er stor.



Figur A36 Plott av delvis reviderte verdier mot reviderte verdier (i tusen kroner), når det er de 606 bedriftene som plukkes ut vha metoden som blir revidert. De to heltrukne linjene markerer grensene for når en feil er stor.

De sist utgitte publikasjonene i serien Notater

- 2005/19 H. Tønnseth. Årsrapport 2004. Kontaktutvalget for helse- og sosialstatistikk 10s.
- 2005/20 N.K. Buskoven: Vertskommunekompensasjon - kartlegging av kommunenes utgifter til asylmottak. 49s.
- 2005/21 H.C. Hougen: Omnibusundersøkelsen oktober/november 2004. Dokumentasjonsrapport. 52s.
- 2005/22 D. Sve, L. Solheim og G. Haraldsen: Eldres kvalitet. Dokumentasjon av datafangsten. 64s.
- 2005/23 E. Rauan: Undersøking om foreldrebetaling i barnehagar, januar 2005. 45s.
- 2005/24 L. Østby: Bruk av velferdsordninger blant nyankomne innvandrere fra de nye EØS-medlemslandene. 36s.
- 2005/25 A. Fagereng: Reestimering av faktoretterspørselen i KVARTS. 72s.
- 2005/26 O. Haugen: Utrekning av vekter til inntekts og formuesundersøkingane 2000, 2001 og 2002. 56s.
- 2005/27 M. Bråthen, J.I. Hamre og T. Pedersen: Evaluering av ordinære arbeidsmarkedstiltak. Beskrivende analyse av deltakerne i 2002 og forslag til ny evalueringsmetode. 33s.
- 2005/28 M. Høstmark: Forundersøkelse om kommunale helseutgifter knyttet til bosetting av flyktninger. 48s.
- 2005/29 A. Vedø: Analyse av revisjon. Lønn i bygge- og anleggsvirksomhet. 43s.
- 2005/30 H.C. Hougen: Samordnet levekårsundersøkelse 2004 - tverrsnittsundersøkelsen. Dokumentasjonsrapport. 139s.
- 2005/31 T. Hægeland, L.J. Kirkebøen og O. Raaum: Skolerresultater 2004. En kartlegging av karakterer fra grunn- og videregående skoler i Norge. 89s.
- 2005/32 A. Rolland: Brukertilfredshetsmålinger i offentlig sektor. Utredning for Moderniseringsdepartementet og regjeringens handlingsplan for modernisering. 96s.
- 2005/33 K. Aasestad, A. Finstad og K. Loe Hansen: Bruk av helsefarlige produkter i grafisk industri. 27s.
- 2005/34 S.W. Bogen, K. Digre, A. Hedum, T. Hægeland, T.K. Schjerven og B. Vold: Et system for statistikk omstatlig virksomhet. Forprosjektnotat. 44s.
- 2005/35 Kostra. Arbeidsgrupperapporter 2005. 230s.
- 2005/36 D. Rafat: Produksjonsopplegg for foreløpige tall i industristatistikken. 46s.
- 2005/37 T. Dale og B. Hole: Evaluering av elektroniske skjemaer i KOSTRA. Case: Skjema 20 - Fysisk planlegging, kulturminner, natur og nærmiljø. 55s.
- 2005/38 A. Sundvoll: Kirkelig tjenestestatistikk i KOSTRA-drakt. Et pilotprosjekt. 48s.
- 2005/39 G.I. Gundersen, B. Hoem, P. Løkkevik og D. Splide. Gjennomgang av metoder og datakilder i energiregnskapet. 50s.
- 2005/40 K. Loe Hansen: Bruk av helsefarlige produkter i båtbyggerbransjen. 27s.
- 2005/41 S. Skaare: Undersøkelsen om samvær og bidrag 2004. 67s.
- 2005/42 A. Haglund, A. Hedum, T. Schjerven og K.Ø. Sørensen: Offentlig sektor og BoF. 63s.
- 2005/43 O. Villund: Yrkesdata for selvstendig næringsdrivende. Dokumentasjonsnotat. 44s.
- 2005/44 O. Villund: Alder i AKU endring av definisjoner og trekkgrunnlag. 27s.
- 2005/45 J.I. Hamre: Estimering av fylkesfordelte og sektorfordelte tall for egenmeldt sykefravær. Dokumentasjon av metode og system, og resultater. 67s.