2007/21

Notater

Statistics Norway

Statistisk sentralbyrå

Anna-Karin Mevik

Notater

Et usikkerhetsmål for Produksjonsindeksen for industri

Stabsavdeling/Seksjon for metoder og standarder

Innhold

1. Innledning

Produksjonsindeksen for industri er en månedlig volumindeks. Ideelt skal den måle volumendringen i bearbeidingsverdien innen olje- og gassutvinning, industri, bergverksdrift og kraftforsyning (men vi må begrense våre usikkerhetsberegninger til industri og bergverksdrift, se kapittel 4.2). Indeksen er bygd opp som en årlig kjedet Laspeyres indeks.

I dette notatet skal vi lage et usikkerhetsmål for produksjonsindeksen. Vi ønsker å lage et usikkerhetsmål som sier noe om hvor nær vi kan forvente at produksjonsindeksen er den faktiske volumendringen. Med et slikt usikkerhetsmål kan vi lage konfidensintervall, som er en populær måte å fremstille usikkerheten på. (Et konfidensintervall er et intervall som med stor sannsynlighet omslutter den størrelsen som skal estimeres, i dette tilfellet volumendringen i bearbeidingsverdien).

Dermed trenger vi en konkret definisjon på hva som menes med volumendring i bearbeidingsverdien (ellers er det umulig å si noe om hvor godt produksjonsindeksen klarer å treffe denne volumendringen). Som for volumindekser generelt har vi ingen slik definisjon. (Det fins definisjoner på ulike indekser som kan brukes til å måle en volumendring, som f.eks. Laspeyres og Paasches volumindeks, men vi har ikke funnet noen definisjon på selve volumendringen). Derfor skal vi i dette notatet lage oss en definisjon for dette, og måle usikkerheten i produksjonsindeksen ut fra denne.

Vi starter i kapittel 2 med å gi en kort innføring i oppbyggingen av en kjedet Laspeyres volumindeks. Med utgangspunkt i denne indeksen lager vi en definisjon på hva vi skal mene med volumendring. I kapittel 3 presenterer vi produksjonsindeksen, og ser hvordan definisjonen på volumendingen blir i denne situasjonen. I kapittel 4 lager vi et usikkerhetsmål for indeksen. Dette har vi så benyttet til å estimere usikkerheten i produksjonsindeksen for 2005 og 2006. Resultatet av dette gir vi i kapittel 5. Helt til slutt gir vi en oppsummering i kapittel 6.

I arbeidet med usikkerhetsberegningene og skriving av notatet har Øyvind Naustdal og Joaquin Rodriguez på seksjon 240 (økonomiske indikatorer) bidratt med nyttige tilbakemeldinger og viktig informasjon.

2. Laspeyres volumindeks

Volumindekser brukes en del innen økonomisk statistikk. Hensikten med slike indekser er å måle volumendringer. Vi kan ev. si at en volumindeks skal angi hvor stor faktor av en verdiendring som skyldes endring i volum, til forskjell fra endring i pris. De kanskje mest kjente volumindeksene er Paasche og Laspeyre.

Her skal vi gi en kort presentasjon av Laspeyres indeks. Vi skal ta for oss en situasjon hvor det er den årlige volumendringen i produksjonsverdien som skal måles. La oss si at det er endringen fra år *t*-1 til år *t* som skal beregnes (år *t*-1 kalles gjerne basisår mens *t* kalles statistikkår). Da er Laspeyres volumindeks gitt ved

$$\widehat{Q}_{t/t-1} = \frac{\sum_{i \in s} \sum_j p_{ij}^{t-1} q_{ij}^t}{\sum_{i \in s} \sum_j p_{ij}^{t-1} q_{ij}^{t-1}},$$

der *s* er et utvalg av bedrifter, summen $\sum_j$ er over de viktigste varene til den enkelte bedrift, q_{ij}^{t-1} og q_{ij}^t er produsert mengde av vare *j* for bedrift *i*, i årene *t*-1 og t (henholdsvis), og p_{ij}^{t-1} er en pris på vare *j* for bedrift *i*, i år *t*-1. (Merk at summene i teller og nevner skal være over de samme enhetene).

Det kan bemerkes at Laspeyres indeks ikke medfører sammenhengen $endring\ i\ verdi =$ $(endring\ i\ pris)\cdot(endring\ i\ volum)$. Når en verdi endrer seg fra en periode til en annen, skyldes dette normalt en kombinasjon av pris- og volumendring (vi ser bort fra problematikken med kvalitetsendringer). Hvis en volumindeks skal måle den delen av en verdiendring som skyldes endring i volum, mens en prisindeks skal måle den delen som skyldes endring i pris, er dette en sammenheng vi ønsker. Men produktet av Laspeyres pris- og volumindeks er ikke lik verdiendringen. (Fishers indeks er derimot en indeks som tilfredstiller denne sammenhengen (Chevalier, 2003)).

Hva er så $\widehat{Q}_{t/t-1}$ en estimator for? I dette notatet velger vi å si at $\widehat{Q}_{t/t-1}$ er en estimator for den størrelsen vi får på $\widehat{Q}_{t/t-1}$ hvis vi observerer hele populasjonen. Dvs. vi velger å si at $\widehat{Q}_{t/t-1}$ er en estimator for

$$Q_{t/t-1} = \frac{\sum_{i\in U}\sum_j p_{ij}^{t-1}q_{ij}^{t}}{\sum_{i\in U}\sum_j p_{ij}^{t-1}q_{ij}^{t-1}},$$

der U er alle bedrifter i populasjonen, og summen $\sum_j$ er over alle varene til den enkelte bedrift.

Størrelsen $Q_{t/t-1}$ er med andre ord vår definisjon på en volumendring. Merk at summene i teller og nevner skal gå over de samme enhetene. Dette betyr bl.a. at bedrifter som blir etablert i løpet av år *t*, ikke er med i $Q_{t/t-1}$. Bidraget fra disse bedriftene blir dermed ikke regnet med i den samlede volumendringen. Det er også vert å merke seg at vi ville fått en annen definisjon på volumendring hvis vi hadde tatt utgangspunkt i en annen indeks enn Laspeyres.

Den kjedede indeksen til Laspayres, for år *t* og med basisår $t-n$, er gitt ved

$$\hat{I}_{t/t-n} = \widehat{Q}_{t-n+1/t-n}\cdot\ldots\cdot\widehat{Q}_{t-1/t-2}\cdot\widehat{Q}_{t/t-1}\cdot 100.$$

I en slik kjeding kalles $\widehat{Q}_{t/t-1}$ gjerne for korttidsindeks, mens $\hat{I}_{t/t-n}$ kalles langtidsindeks.

Fordi $\hat{I}_{t-1/t-n} = \widehat{Q}_{t-n+1/t-n}\cdot\ldots\cdot\widehat{Q}_{t-1/t-2}\cdot 100$ får vi at $\hat{I}_{t/t-n}$ kan skrives som

$$\hat{I}_{t/t-n} = \hat{I}_{t-1/t-n}\cdot\widehat{Q}_{t/t-1}.$$

Dvs. at vi får langtidsindeksen for år *t* ved å kjede korttidsindeksen $\widehat{Q}_{t/t-1}$ på langtidsindeksen for år $t-1$.

Når vi jobber med langtidsindeksen $\hat{I}_{t/t-n} = \hat{I}_{t-1/t-n}\cdot\widehat{Q}_{t/t-1}$ i dette notatet, skal vi se på $\hat{I}_{t-1/t-n}$ som en gitt konstant. Det er med andre ord bare korttidsindeksen $\widehat{Q}_{t/t-1}$ som er stokastisk i $\hat{I}_{t/t-n}$. Grunnen til at vi vil behandle $\hat{I}_{t-1/t-n}$ som en gitt konstant, er at indekser stort sett benytter seg av det samme utvalget hvert år (dvs. panel). Dette skaper en avhengighet mellom korttidsindeksene, og dermed også mellom $\hat{I}_{t-1/t-n}$ og $\widehat{Q}_{t/t-1}$, som vi ikke klarer å håndtere. (Selv uten avhengighet mellom korttidsindeksene får vi problemer hvis vi ikke behandler $\hat{I}_{t-1/t-n}$ som en konstant, fordi vi da må beregne variansen til et produkt av *n* stokastiske korttidsindekser).

Fordi vi skal se på $\hat{I}_{t-1/t-n}$ som en gitt konstant i langtidsindeksen $\hat{I}_{t/t-n}$, velger vi å si at langtidsindeksen er en estimator for

$$I_{t/t-n} = \hat{I}_{t-1/t-n} \cdot Q_{t/t-1} .$$

Det hadde kanskje vært mer naturlig å se på $\hat{I}_{t/t-n}$ som en estimator for $I_{t/t-n} = I_{t-1/t-n} \cdot Q_{t/t-1}$, der $I_{t-1/t-n} = Q_{t-n+1/t-n} \cdot \ldots \cdot Q_{t-1/t-2} \cdot 100$. Men da vil usikkerheten til $\hat{I}_{t/t-n}$ blant annet avhenge av differansen $\hat{I}_{t-1/t-n} - I_{t-1/t-n}$, og fordi $I_{t-1/t-n}$ er ukjent vil vi ikke klare å estimere denne usikkerheten. Vi kunne selvsagt valgt å estimere den ukjente $I_{t-1/t-n}$ med $\hat{I}_{t-1/t-n}$, men da ender vi opp med det samme usikkerhetsmålet som vi får med $I_{t/t-n} = \hat{I}_{t-1/t-n} \cdot Q_{t/t-1}$. Derfor synes vi det er like greit å se på $\hat{I}_{t/t-n}$ som en estimator for $I_{t/t-n} = \hat{I}_{t-1/t-n} \cdot Q_{t/t-1}$, og måle usikkerheten ut fra dette.

3. Produksjonsindeksen for industri

Produksjonsindeksen er bygd opp som en årlig kjedet Laspeyres indeks, og den skal måle volumendringen i bearbeidingsverdien innen olje- og gassutvinning, industri, bergverksdrift og kraftforsyning.

Indeksen beregnes på flere aggregeringsnivåer. Det laveste nivået er sektornivå, eller bearbeidingsnivå som det også kalles (men disse publiseres ikke). En sektor tilsvarer i all hovedsak en næringsgruppe, altså 4-sifret NACE-nivå (se Sørensen, 1998, og Standard for næringsgruppering). Det er totalt 132 sektorer innen olje- og gassutvinning, industri, bergverksdrift og kraftforsyning, men produksjonsindeksen beregnes bare for 120 av disse. De 12 sektorene som indeksen ikke beregnes for, har ingen eller veldig lav aktivitet. Alle andre aggregeringsnivå er bygd opp av en eller flere sektorer.

Utvalget som brukes til beregning av indeksen ble trukket i 1996. Det ble benyttet stratifisert enkel tilfeldig trekking, der strataene ble bestemt ut fra omsetning, sysselsetting og næring. Stratifisering og allokering av utvalgsenheter er gjort slik at alle bedrifter med minst 100 sysselsatte, og alle med minst 10% av omsetningen på sektornivå, er tatt med i utvalget. Videre er det ikke trukket ut bedrifter blant de som har færre enn 10 sysselsatte. Utvalgsstørrelsen for de resterende strataene er bestem ved optimal allokering (basert på bedriftenes omsetning). Dette gjør at dekningsgraden for utvalget, med hensyn til omsetning, er veldig stor. I 1998 var f.eks. dekningsgraden på 80%. Det gjøres årlige oppdateringer av utvalget.

Indeksen skal som nevnt måle endringen i volumet av bearbeidingsverdi. Bearbeidingsverdien til en bedrift er gitt ved brutto produksjonsverdi minus verdien på innsatsvarene. Hvis $p_{i,j,t}$ og $q_{i,j,t}$ er henholdsvis pris og mengde for en produsert vare *j*, for bedrift *i* i periode *t*, mens $\alpha_{i,l,t}$ og $\delta_{i,l,t}$ er henholdsvis pris og mengde for en innsatsvare *l*, for bedrift *i* i perioden *t*, kan bearbeidingsverdien for bedrift *i* skrives som

$$ba_{i,t} = \sum_j p_{i,j,t} \cdot q_{i,j,t} - \sum_l \alpha_{i,l,t} \cdot \delta_{i,l,t} .$$

Summen $\sum_j$ er over alle varer som bedriften produserer, og $\sum_l$ er over alle innsatsvarer som bedriften benytter. (Fordi det praktisk talt er umulig å få disse opplysningene på månedsbasis, må vi bruke andre data for å måle volumendringen i bearbeidingsverdien. Dette kommer vi tilbake til i avsnitt 3.1).

3.1. Korttidsindeksen

Korttidsindeksen skal måle volumendringen for en bestemt måned *m*, sett i forhold til en gjennomsnittlig måned året før. Med vår definisjon av volumendring fra kapittel 2 får vi at korttidsindeksen er en estimator for volumendringen

$$Q_{t,m} = \frac{\sum_{i \in U} ba^{*}_{i,t,m}}{\sum_{i \in U} ba_{i,t-1} \big/ 12}.$$

Her er U mengden av alle bedrifter som hører til gjeldende aggregeringsnivå, $ba_{i,t-1}$ er bearbeidingsverdien til bedrift *i* i år *t*-1, og $ba^{*}_{i,t,m} = \sum_j p_{i,j,t-1} \cdot q_{i,j,t,m} - \sum_l \alpha_{i,l,t-1} \cdot \delta_{i,l,t,m}$ (dvs. bearbeidingsverdien for måned *m* i år *t*, beregnet med samme priser som brukes i år *t*-1).

Volumendringen $Q_{t,m}$ kan skrives som en vektet sum av volumendringene på sektornivå. Hvis vi lar $Q^{b}_{t,m}$ betegne volumendringen for en sektor *b*, har vi at

$$Q_{t,m} = \sum_b w^b_t Q^b_{t,m} ,$$

der $w^b_t = \dfrac{\sum_{i \in U_b} ba_{i,t-1}}{\sum_b \sum_{i \in U_b} ba_{i,t-1}}$, U_b er mengden av bedrifter i sektor *b*, og summen $\sum_b$ er over sektorene som inngår i det nivået $Q_{t,m}$ gjelder. Vekten w^b_t til en sektor tilsvarer sektorens andel av bearbeidingsverdien på gjeldende nivå. Vekten avhenger med andre ord av hvilket aggregeringsnivå $Q_{t,m}$ gjelder.

Volumendringen estimeres med indeksen

$$\widehat{Q}_{t,m} = \sum_b w^b_t \widehat{Q}^b_{t,m} ,$$

der $\widehat{Q}^b_{t,m}$ er indeks for sektor *b*.

Ideelt sett skal $\widehat{Q}^b_{t,m}$ være en indeks for volumendringen i bearbeidingsverdien. Men de nødvendige dataene for å beregne bearbeidingsverdi er normalt bare tilgjenglig på årlig basis og med betydelig forsinkelse. Det er derfor ikke mulig å lage en slik indeks. I stedet er det valgt å lage en indeks som måler endringen i enten produksjonsverdi i faste priser eller timeverk, i håp om at dette skal gi en god tilnærming av volumendringen i bearbeidingsverdien.

Faktisk har vi ikke tilgang til bearbeidingsverdien fra foregående år heller, selv ikke aggregerte verdier. Dette betyr at vektene, som er andeler av bearbeidingsverdi fra foregående år, ikke er kjent. Derfor brukes det vekter som er beregnet ut fra framskrevne bearbeidingsverdier. I våre analyser har vi valgt å se bort fra dette, hvilket betyr at vi underslår noe av usikkerheten i produksjonsindeksen.

Sektorene hvor vi benytter indeks for volumendring i produksjonsverdien tilhører næringene 10-21, 23-27, 29.7, 31.3, 36, 37 og 40 (se Standard for næringsgruppering). Det er totalt 81 sektorer, og vi velger å kalle disse for produksjonssektorer. I disse sektorene antas det at volumendringen i produksjonsverdien (basert på de viktigste varene til bedriftene) gir en god tilnærming av volumendringen i bearbeidingsverdien. Dvs. at

$$Q_{t,m}^{b} \approx \frac{\sum_{i \in U_b} \sum_{j} p_{i,j,t-1} q_{i,j,t,m}}{\sum_{i \in U_b} \sum_{j} p_{i,j,t-1} q_{i,j,t-1} / 12},$$

der $p_{i,j,t-1}$ og $q_{i,j,t-1}$ er henholdsvis pris og produsert mengde av vare *j* for bedrift *i* i år *t*-1, og $q_{i,j,t,m}$ er produsert mengde av vare *j* for bedrift *i* i måned *m* i år *t*. Summen $\sum_{j}$ går over de viktigste varene til bedriften.

I produksjonssektorene blir dermed volumendringen i bearbeidingsverdi estimert med indeksen

$$\widehat{Q}_{t,m}^{b} = \frac{\sum_{i \in s_b} \sum_{j} p_{i,j,t-1} q_{i,j,t,m}}{\sum_{i \in s_b} \sum_{j} p_{i,j,t-1} q_{i,j,t-1} / 12},$$

der s_b er utvalget av bedrifter fra sektor *b*.

Sektorene hvor vi benytter en indeks for timeverk tilhører næringene 22, 28-35 (unntatt 29.7 og 31.3). Det er totalt 39 sektorer, og vi velger å kalle dem timeverkssektorer. I disse sektorene har det vært vanskelig å finne gode produkter å innhente tall for, og produksjonen av et produkt kan ta flere måneder. Derfor kan vi ikke bruke volumendring i månedlig produksjonsverdi som en tilnærming av volumendringen i månedlig bearbeidingsverdi. I stedet antas det at endring i brukte timeverk kan gi en tilfredsstillende tilnærming, hvis det korrigeres for produktivitet. Mer presist antar vi sammenhengen

$$Q_{t,m}^{b} \approx f_{m}^{b} \frac{\sum_{i \in U_b} t_{i,t,m}}{\sum_{i \in U_b} t_{i,t-1} / 12},$$

der $t_{i,t-1}$ er brukte timeverk for bedrift *i* i år *t*-1, mens $t_{i,t,m}$ er brukte timeverk i måned *m* i år *t*. Faktoren f_{m}^{b} er en produktivitetsfaktor som skal korrigere for at bedre produksjonsutstyr, ny teknikk og lignende øker produktiviteten per timeverk. Fordi produktiviteten kan variere fra sektor til sektor, og fra måned til måned, er det lagd produktivitetsfaktorer som avhenger av sektor og måned.

I timeverkssektorene blir dermed volumendringen i bearbeidingsverdi estimert med

$$\widehat{Q}_{t,m}^{b} = f_{m}^{b} \frac{\sum_{i \in s_b} t_{i,t,m}}{\sum_{i \in s_b} t_{i,t-1} / 12}.$$

3.2. Langtidsindeksen

I produksjonsindeksen brukes det årlig kjeding med 1995 som referanseperiode. Fordi korttidsindeksene måler endringen for en måned, sett i forhold til en gjennomsnittlig måned året før, kan disse ikke kjedes samme direkte. I stedet lages det årlige korttidsindekser som måler endringen for et år sett i forhold til foregående år, og så kjedes disse sammen. De årlige korttidsindeksene er gitt ved $\widehat{Q}_t = \frac{1}{12}\sum_{m=1}^{12}\widehat{Q}_{t,m}$, dvs. gjennomnittet av de månedlige korttidsindeksene.

Langtidsindeksen for en måned *m* i år *t* er nå gitt ved

$$\begin{aligned}\hat{I}_{t,m} &= \widehat{Q}_{96}\cdot\widehat{Q}_{97}\cdot\ldots\cdot\widehat{Q}_{t-1}\cdot\widehat{Q}_{t,m}\cdot 100\\ &= \hat{I}_{t-1}\cdot\widehat{Q}_{t,m}\ ,\end{aligned}$$

der $\hat{I}_{t-1} = \widehat{Q}_{96}\cdot\widehat{Q}_{97}\cdot\ldots\cdot\widehat{Q}_{t-1}\cdot 100$. ($\hat{I}_{t-1}$ kan sees på som en langtidsindeks for år *t*-1).

Som vi argumenterte for i kapittel 2, skal vi se på $\hat{I}_{t,m}$ som en estimator for

$$I_{t,m} = \hat{I}_{t-1}\cdot Q_{t,m}\ ,$$

og vi skal behandle $\hat{I}_{t-1}$ som en gitt konstant.

Altså, produksjonsindeksen er gitt ved

$$\hat{I}_{t,m} = \hat{I}_{t-1}\cdot\widehat{Q}_{t,m}\ ,$$

og vi skal lage et usikkerhetsmål som sier noe om hvor nær vi kan forvente at produksjonsindeksen er volumendringen $I_{t,m} = \hat{I}_{t-1}\cdot Q_{t,m}$. Videre skal vi behandle $\hat{I}_{t-1}$ som en gitt konstant.

For mer informasjon om produksjonsindeksen, se Sørensen (1998).

4. Et usikkerhetsmål for produksjonsindeksen

4.1. Utledning av usikkerhetsmål

I dette kapittelet skal vi lage et usikkerhetsmål for produksjonsindeksen $\hat{I}_{t,m}$. Vi skal lage det vi kaller et modellbasert usikkerhetsmål. Det betyr at vi ser på utvalget som gitt, mens bedriftenes timeverk $t_{i,t,m}$ og produksjonsverdi $\sum_j p_{i,j,t-1}q_{i,j,t,m}$ er stokastiske variable.[1]

[1] Det er $q_{i,j,t,m}$ (produsert mengde av vare *j* for bedrift *i* i måned *m*, år *t*) som er stokastisk. Fordi en og samme bedrift rapporterer for flere varer, velger vi å modellere summen $\sum_j p_{i,j,t-1}q_{i,j,t,m}$ fremfor hver $q_{i,j,t,m}$ for seg (summen $\sum_j$ er over alle varer bedriften rapporterer for). På den måten slipper vi å tenke på avhengighetsstrukturen mellom varene til en bedrift.

Vi har valgt å se bort fra usikkerhet som skyldes målefeil og frafall. I tillegg til har vi valgt å se bort fra at endring i produksjonsverdi og timeverk bare er tilnærminger til volumendring i bearbeidingsverdi. I stedet lager vi et usikkerhetsmål som om det er eksakt likhet, dvs.

$$Q_{t,m}^{b} = \frac{\sum_{i \in U_b} \sum_j p_{i,j,t-1} q_{i,j,t,m}}{\sum_{i \in U_b} \sum_j p_{i,j,t-1} q_{i,j,t-1} / 12}$$

for produksjonssektorene, og

$$Q_{t,m}^{b} = f_m^b \frac{\sum_{i \in U_b} t_{i,t,m}}{\sum_{i \in U_b} t_{i,t-1} / 12}$$

for timeverkssektorene. Hvor mye vi underslår usikkerheten med dette, avhenger av hvor stor forskjell det er mellom volumendring i bearbeidingsverdi og endring i produksjonsverdi eller timeverk.

For å kunne lage det modellbaserte usikkerhetsmålet trenger vi modell for produksjonsverdien $\sum_j p_{i,j,t-1} q_{i,j,t,m}$ og brukte timeverk $t_{i,t,m}$ for en bedrift. Det viser seg at begge disse størrelsene stort sett lar seg beskrive veldig bra med ratemodellen, når forrige års verdier er forklaringsvariabel. Dvs., for bedrifter i produksjonssektorene har vi modellen

$$\mathrm{E}\left[\sum_j p_{i,j,t-1} q_{i,j,t,m}\right] = \beta_g \cdot \sum_j p_{i,j,t-1} q_{i,j,t-1} / 12$$

$$\mathrm{V}\left(\sum_j p_{i,j,t-1} q_{i,j,t,m}\right) = \sigma_g^2 \cdot \sum_j p_{i,j,t-1} q_{i,j,t-1} / 12 ,$$

der *g* deler inn bedriftene i 6 grupper, avhenging av hvilken sektor bedriften tilhører.

For bedrifter i timeverkssektorene har vi

$$\mathrm{E}\left[t_{i,t,m}\right] = \beta_g \cdot t_{i,t-1} / 12$$

$$\mathrm{V}\left(t_{i,t,m}\right) = \sigma_g^2 \cdot t_{i,t-1} / 12 ,$$

der *g* deler inn bedriftene i 3 grupper. (Se vedlegg for mer om modelltilpassingen).

For å gjøre notasjonen litt lettere innfører vi variablene y_i og x_i for en bedrift *i*, der

$$y_i = \begin{cases} \sum_j p_{i,j,t-1} q_{i,j,t,m} & \text{hvis bedriften tilhører en produksjonssektor} \\ t_{i,t,m} & \text{hvis bedriften tilhører en timeverkssektor} \end{cases}$$

og

$$x_i = \begin{cases} \sum_j p_{i,j,t-1} q_{i,j,t-1} / 12 & \text{hvis bedriften tilhører en produksjonssektor} \\ t_{i,t-1} / 12 & \text{hvis bedriften tilhører en timeverkssektor} \end{cases} .$$

Dermed har vi modellen

$$\mathrm{E}\left[y_i\right] = \beta_g x_i$$

$$\mathrm{V}\left(y_i\right) = \sigma_g^2 x_i \,,$$

for både produksjonsverdi og timeverk. I tillegg antar vi at bedriftene er uavhengige av hverandre.

Vi skal nå sjekke om produksjonsindeksen $\hat{I}_{t,m}$ er forventningsrett for $I_{t,m}$ under denne modellen, dvs. om forventet verdi til differansen $\hat{I}_{t,m} - I_{t,m}$ er lik null. Først merker vi oss at

$$\begin{aligned}
\hat{I}_{t,m} - I_{t,m} &= \hat{I}_{t-1} \cdot \widehat{Q}_{t,m} - \hat{I}_{t-1} \cdot Q_{t,m} \\
&= \hat{I}_{t-1}\left(\widehat{Q}_{t,m} - Q_{t,m}\right) \\
&= \hat{I}_{t-1} \cdot \left(\sum_b w_t^b \widehat{Q}_{t,m}^b - \sum_b w_t^b Q_{t,m}^b\right) \\
&= \hat{I}_{t-1} \cdot \sum_b w_t^b \left(\widehat{Q}_{t,m}^b - Q_{t,m}^b\right) .
\end{aligned}$$

Dermed har vi, ved å huske at $\hat{I}_{t-1}$ er en konstant,

$$\mathrm{E}\left[\hat{I}_{t,m} - I_{t,m}\right] = \hat{I}_{t-1} \cdot \sum_b w_t^b \cdot \mathrm{E}\left[\widehat{Q}_{t,m}^b - Q_{t,m}^b\right].$$

For å finne forventningen til $\widehat{Q}_{t,m}^b - Q_{t,m}^b$ bruker vi definisjonen av y_i og x_i, og får at

$$\widehat{Q}_{t,m}^b = f_m^b \cdot \frac{\sum_{i \in s_b} y_i}{\sum_{i \in s_b} x_i}$$

og

$$Q_{t,m}^b = f_m^b \cdot \frac{\sum_{i \in U_b} y_i}{\sum_{i \in U_b} x_i} \,,$$

for både produksjonssektorene og timeverkssektorene. Her er $f_m^b = 1$ for produksjonssektorene (for timeverkssektorene er f_m^b produktivitetsfaktoren). Av dette får vi

$$
\begin{aligned}
\mathrm{E}\left[\hat{I}_{t,m} - I_{t,m}\right] &= \hat{I}_{t-1} \cdot \sum_b w_t^b \cdot \mathrm{E}\left[\widehat{Q}_{t,m}^b - Q_{t,m}^b\right] \\
&= \hat{I}_{t-1} \cdot \sum_b w_t^b \cdot \mathrm{E}\left[f_m^b \cdot \frac{\sum_{i \in s_b} y_i}{\sum_{i \in s_b} x_i} - f_m^b \cdot \frac{\sum_{i \in U_b} y_i}{\sum_{i \in U_b} x_i}\right] \\
&= \hat{I}_{t-1} \cdot \sum_b w_t^b f_m^b \left(\frac{\sum_{i \in s_b} \mathrm{E}[y_i]}{\sum_{i \in s_b} x_i} - \frac{\sum_{i \in U_b} \mathrm{E}[y_i]}{\sum_{i \in U_b} x_i}\right) \\
&= \hat{I}_{t-1} \cdot \sum_b w_t^b f_m^b \left(\beta_g - \beta_g\right) \\
&= 0\ .
\end{aligned}
$$

Produksjonsindeksen er med andre ord forventingsrett.

Fordi produksjonsindeksen er forventningsrett, er kvadratroten av prediksjonsvariansen et fornuftig mål på usikkerheten. Dvs. vi bruker standardavviket

$$
\mathrm{st}\left(\hat{I}_{t,m} - I_{t,m}\right) = \sqrt{\mathrm{V}\left(\hat{I}_{t,m} - I_{t,m}\right)}
$$

som mål på usikkerheten til produksjonsindeksen $\hat{I}_{t,m}$.

For å beregne prediksjonsvariansen til $\hat{I}_{t,m}$ benytter vi at $\hat{I}_{t,m} - I_{t,m} = \hat{I}_{t-1} \cdot \sum_b w_t^b \left(\widehat{Q}_{t,m}^b - Q_{t,m}^b\right)$. På grunn av uavhengighet og at $\hat{I}_{t-1}$ er en konstant får vi

$$
\mathrm{V}\left(\hat{I}_{t,m} - I_{t,m}\right) = \hat{I}_{t-1}^2 \cdot \sum_b \left(w_t^b\right)^2 \mathrm{V}\left(\widehat{Q}_{t,m}^b - Q_{t,m}^b\right).
$$

Prediksjonsvariansen til $\hat{I}_{t,m}$ er med andre ord en vektet sum av prediksjonsvariansene til korttidsindeksen på sektornivå.

For prediksjonsvariansene til korttidsindeksen på sektornivå har vi

$$
\begin{aligned}
\mathrm{V}\left(\hat{Q}_{t,m}^{b}-Q_{t,m}^{b}\right) &= \mathrm{V}\left(\hat{Q}_{t,m}^{b}\right)-2\mathrm{C}\left(\hat{Q}_{t,m}^{b},Q_{t,m}^{b}\right)+\mathrm{V}\left(Q_{t,m}^{b}\right)\\
&= \mathrm{V}\left(f_m^b\frac{\sum_{i\in s_b}y_i}{\sum_{i\in s_b}x_i}\right)-2\mathrm{C}\left(f_m^b\frac{\sum_{i\in s_b}y_i}{\sum_{i\in s_b}x_i},f_m^b\frac{\sum_{i\in U_b}y_i}{\sum_{i\in U_b}x_i}\right)+\mathrm{V}\left(f_m^b\frac{\sum_{i\in U_b}y_i}{\sum_{i\in U_b}x_i}\right)\\
&= \left(f_m^b\right)^2\left(\frac{\sigma_g^2}{\sum_{i\in s_b}x_i}-2\frac{\mathrm{C}\left(\sum_{i\in s_b}y_i,\sum_{i\in U_b}y_i\right)}{\left(\sum_{i\in s_b}x_i\right)\left(\sum_{i\in U_b}x_i\right)}+\frac{\sigma_g^2}{\sum_{i\in U_b}x_i}\right)\\
&= \left(f_m^b\right)^2\left(\frac{\sigma_g^2}{\sum_{i\in s_b}x_i}-2\frac{\mathrm{V}\left(\sum_{i\in s_b}y_i\right)}{\left(\sum_{i\in s_b}x_i\right)\left(\sum_{i\in U_b}x_i\right)}+\frac{\sigma_g^2}{\sum_{i\in U_b}x_i}\right)\\
&= \left(f_m^b\right)^2\left(\frac{\sigma_g^2}{\sum_{i\in s_b}x_i}-2\frac{\sigma_g^2}{\sum_{i\in U_b}x_i}+\frac{\sigma_g^2}{\sum_{i\in U_b}x_i}\right)\\
&= \left(f_m^b\right)^2\sigma_g^2\left(\frac{1}{\sum_{i\in s_b}x_i}-\frac{1}{\sum_{i\in U_b}x_i}\right).
\end{aligned}
$$

Standardavviket $\mathrm{st}\left(\hat{I}_{t,m}-I_{t,m}\right)$ kan dermed skrives som

$$\mathrm{st}\left(\hat{I}_{t,m}-I_{t,m}\right)=\hat{I}_{t-1}\cdot\sqrt{\sum_b\left(w_t^b\right)^2\mathrm{V}\left(\hat{Q}_{t,m}^{b}-Q_{t,m}^{b}\right)},$$

der

$$\mathrm{V}\left(\hat{Q}_{t,m}^{b}-Q_{t,m}^{b}\right)=\left(f_m^b\right)^2\sigma_g^2\left(\frac{1}{\sum_{i\in s_b}x_i}-\frac{1}{\sum_{i\in U_b}x_i}\right).$$

Fordi dette standardavviket er ukjent må det estimeres før vi kan bruke det til å måle usikkerheten. Det gjør vi ved å estimere prediksjonsvariansen til korttidsindeksene på sektornivå.

I prediksjonsvariansen til korttidsindeksen er σ_g^2 ukjent. I tillegg er x_i ukjent for bedriftene som ikke er med i utvalget. Vi må derfor estimere σ_g^2 og x_i for $i\notin s_b$.

Som estimator for σ_g^2 bruker vi

$$\hat{\sigma}_g^2=\frac{1}{n_g-1}\sum_{i\in s_g}\frac{1}{x_i}\left(y_i-\hat{\beta}x_i\right)^2,$$

der $\hat{\beta} = \sum_{i \in s_g} y_i \Big/ \sum_{i \in s_g} x_i$, s_g er bedrifter i utvalget som tilhører gruppe *g*, og n_g er antall bedrifter i s_g . Dette er en forventningsrett estimator for σ_g^2 .

Ved estimering av x_i har vi brukt forskjellige estimatorer i produksjonssektorene og timeverkssektorene. I timeverkssektorene er $x_i = t_{i,t-1}/12$, dvs. x_i er bedriftens gjennomsnittlig brukte timeverk per måned. Innen hver sektor er det en proporsjonal sammenheng mellom x_i og bedriftens sysselsetting. Fordi vi kjenner sysselsettingen til alle bedriftene fra register, har vi utnyttet denne sammenhengen ved å estimere x_i med $\hat{x}_i = \left(\sum_{s_b} x_i \Big/ \sum_{s_b} syss_i \right) \cdot syss_i$, $i \in U_b \setminus s_b$, der $syss_i$ er sysselsetting til bedrift *i*.

I produksjonssektorene er $x_i = \sum_j p_{i,j,t-1} q_{i,j,t-1} / 12$, dvs. x_i er bedriftens gjennomsnittlige produksjonsverdi per måned. I noen av sektorene er det en proporsjonal sammenheng mellom x_i og bedriftens omsetning. For disse sektorene har vi derfor estimert x_i med $\hat{x}_i = \left(\sum_{s_b} x_i \Big/ \sum_{s_b} oms_i \right) \cdot oms_i$, $i \in U_b \setminus s_b$, der oms_i er omsetningen til bedrift *i* (vi kjenner omsetningen til alle bedrifter fra register). I resten av sektorene har vi brukt gjennomsnittlig x_i -verdi som estimator, dvs. $\hat{x}_i = \overline{x}_{s_b}$.

Med disse estimatorene for σ_g^2 og x_i får vi følgende estimator for standardavviket $\mathrm{st}\left(\hat{I}_{t,m} - I_{t,m}\right)$:

(1) $$\widehat{\mathrm{st}}\left(\hat{I}_{t,m} - I_{t,m}\right) = \hat{I}_{t-1} \cdot \sqrt{\sum_b \left(w_t^b\right)^2 \widehat{\mathrm{V}}\left(\widehat{Q}_{t,m}^b - Q_{t,m}^b\right)} ,$$

der

$$\widehat{\mathrm{V}}\left(\widehat{Q}_{t,m}^b - Q_{t,m}^b\right) = \left(f_m^b \right)^2 \hat{\sigma}_g^2 \left(\frac{1}{\sum_{i \in s_b} x_i} - \frac{1}{\sum_{i \in s_b} x_i + \sum_{i \in U_b \setminus s_b} \hat{x}_i} \right).$$

Det er med andre ord dette estimerte standardavviket vi bruker til å måle usikkerheten i produksjonsindeksen.

Vi skal nå lage et konfidensintervall for volumendringen $I_{t,m}$. Vi har at

$$\frac{\hat{I}_{t,m} - I_{t,m}}{\mathrm{st}\left(\hat{I}_{t,m} - I_{t,m}\right)} = \frac{\sum_b w_t^b \left(\widehat{Q}_{t,m}^b - Q_{t,m}^b\right)}{\sqrt{\sum_b \mathrm{V}\left(w_t^b \left(\widehat{Q}_{t,m}^b - Q_{t,m}^b\right)\right)}} .$$

Under visse betingelser på $\mathrm{V}\left(w_t^b \left(\widehat{Q}_{t,m}^b - Q_{t,m}^b\right)\right)$ har vi derfor fra Lindebergsetningen at $\left(\hat{I}_{t,m} - I_{t,m}\right) \Big/ \mathrm{st}\left(\hat{I}_{t,m} - I_{t,m}\right)$ er tilnærmet standardnormaltfordelt (Bjørnstad, 1986), hvis aggregeringsnivået er over tilstrekkelig mange sektorer. Et tilnærmet 95% konfidensintervall for volumendringen $I_{t,m}$ er dermed gitt ved $\hat{I}_{t,m} \pm 1.96 \cdot \mathrm{st}\left(\hat{I}_{t,m} - I_{t,m}\right)$. Fordi $\mathrm{st}\left(\hat{I}_{t,m} - I_{t,m}\right)$ er ukjent putter vi inn estimatoren $\widehat{\mathrm{st}}\left(\hat{I}_{t,m} - I_{t,m}\right)$ for den, og bruker

$$\hat{I}_{t,m} \pm 1.96 \cdot \widehat{\mathrm{st}}\left(\hat{I}_{t,m} - I_{t,m}\right)$$

som et 95% konfidensintervall for volumendringen $I_{t,m}$.

Det bør bemerkes at hvis aggregeringsnivået til indeksen ikke er over tilstrekkelig mange sektorer, så er det ikke sikkert dette intervallet kan brukes som et 95% konfidensintervall (men det vil kunne brukes som et 75% konfidensintervall).

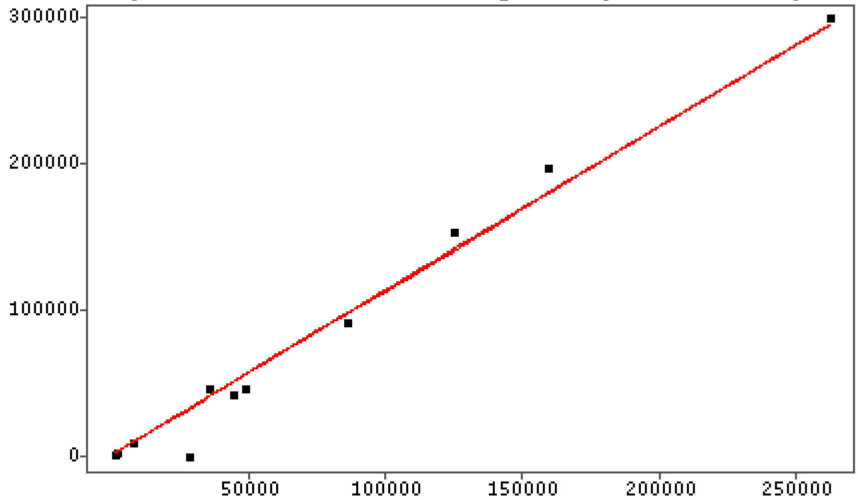
4.2. Hva påvirker den målte usikkerheten

Usikkerhetsmålet $\widehat{\mathrm{st}}\left(\hat{I}_{t,m} - I_{t,m}\right)$ sier noe om hvor nær vi kan forvente at produksjonsindeksen $\hat{I}_{t,m}$ er $I_{t,m}$. Som vi ser av (1) fremkommer denne usikkerheten som en vektet sum av prediksjonsvariansene til korttidsindeksen på sektornivå. Fordi prediksjonsvariansen er et mål på usikkerhet, betyr det at vi får usikkerheten til produksjonsindeksen $\hat{I}_{t,m}$ ved å summere sammen de vektede usikkerhetene til korttidsindeksen på sektornivå.

Av dette får vi at usikkerheten til $\hat{I}_{t,m}$ avhenger av hvilke sektorer som inngår i aggregeringsnivået, og hvor stor usikkerhet og vekt disse sektorene har. Det typiske er at jo høyere aggregeringsnivå, dvs. jo flere sektorer som er med, jo mindre usikkerhet. Men hvis noen sektorer har en stor usikkerhet kombinert med en stor vekt, sammenlignet med de andre sektorene, kan det også hende at usikkerheten går opp når vi aggregerer over flere sektorer. For å få lav usikkerhet på aggregert nivå er det derfor viktig at vi har lav usikkerhet i de sektorene som har stor vekt, og som vi skal nevne i neste avsnittet kan vi få lav usikkerhet ved å øke dekningsgraden i sektoren. Dvs. høy dekningsgrad i sektorene som har stor vekt er viktig for å få lav usikkerhet.

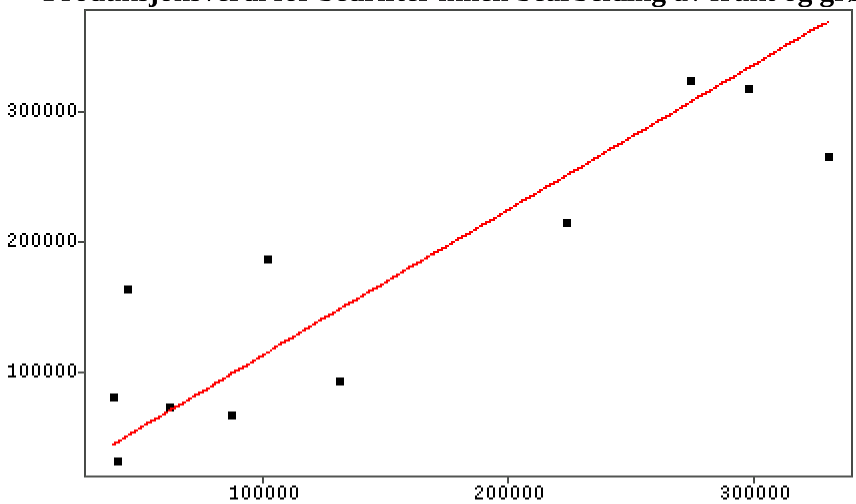
Usikkerheten til korttidsindeksen på sektornivå sier noe om hvor nær vi kan forvente at korttidsindeksen er det tallet vi ville fått på indeksen hvis vi har fulltelling i sektoren. Vi skal nevne noen forhold som påvirker denne usikkerheten. Et forhold er utvalgsstørrelsen. Jo flere bedrifter vi tar med i utvalget, jo mindre usikkerhet. Et annet er utvalgets dekningsgrad målt ved x_i, dvs. målt ved forrige års produksjonsverdi eller timeverk. Jo større dekningsgrad vi har i utvalget, jo mindre usikkerhet. Til sist skal vi nevne et forhold som har med hvordan produksjonsverdien, eller timeverksbruken, til en bedrift har endret seg i forhold til foregående år. Hvis alle bedrifter innen en sektor har endret seg relativt likt, så bidrar dette til liten usikkerhet i denne sektoren (se figur 1). Her derimot bedriftene endret seg veldig forskjellig, så bidrar dette til stor usikkerhet (se figur 2).

Produksjonsverdi for bedrifter innen produksjon av finer, kryssfiner o.l.



Figur 1 *Figuren viser hvordan produksjonsverdien har endret seg fra foregående år for bedrifter som driver med produksjon av finer, kryssfiner o.l. (dvs. sektor 2020). Usikkerheten til korttidsindeksen for denne sektoren er 2.9. Den horisontale aksen viser gjennomsnittlig produksjonsverdi i 2005, mens den vertikale aksen viser produksjonsverdi for januar 2006 (beregnet med 2005-priser). Den heltrukne linjen er regresjonslinjen vi får med ratemodellen.*

Produksjonsverdi for bedrifter innen bearbeiding av frukt og grønnsaker



Figur 2 *Figuren viser hvordan produksjonsverdien har endret seg fra foregående år for bedrifter som driver med bearbeiding av frukt og grønnsaker (dvs. sektor 1530). Usikkerheten til korttidsindeksen for denne sektoren er 9.4. Den horisontale aksen viser gjennomsnittlig produksjonsverdi i 2005, mens den vertikale aksen viser produksjonsverdi for januar 2006 (beregnet med 2005-priser). Den heltrukne linjen er regresjonslinjen vi får med ratemodellen.*

4.3. En begrensing for usikkerhetsmålet

I noen få av sektorene (ca. 15) blir det samlet inn aggregerte opplysninger per sektor. Disse opplysningene kan ikke brytes ned på bedriftsnivå. Fordi usikkerhetsmålet (1) forutsetter opplysinger på bedriftsnivå, kan det ikke brukes på disse sektorene. Fordi vi ikke har funnet noen annen måte å måle usikkerheten i disse sektorene, har vi valgt å si at de ikke har noen usikkerhet, altså at de har usikkerhet lik null. (Disse sektorene har fått $\widehat{\mathrm{V}}\left(\widehat{Q}_{t,m}^{b}-Q_{t,m}^{b}\right)=0$ i usikkerhetsmålet (1)).

Dette betyr at vi underslår usikkerheten fra disse sektorene. Men for de fleste av sektorene er vekten ved aggregering så liten at dette ikke betyr noe nevneverdig på aggregert nivå (forutsatt at usikkerheten til sektoren ikke er ekstremt stor). Men for fire av sektorene hvor vi underslår usikkerheten, er vektene veldig store. Dette gjelder sektorene fra olje- og gassutvinning og kraftforsyning (næring 11 og 40). Usikkerhetsmålet (1) vil derfor underestimere usikkerheten betydelig hvis disse fire sektorene er med i aggregeringsnivået til $\hat{I}_{t,m}$.

Konklusjonen er dermed at usikkerhetsmålet (1) ikke kan brukes til å måle usikkerheten til produksjonsindeksen for olje- og gassutvinning og kraftforsyning, eller andre aggregeringsnivåer hvor olje- og gassutvinning og kraftforsyning inngår.

5. Talleksempel

Vi skal nå presentere resultatene vi fikk når vi benyttet usikkerhetsmålet fra forrige kapittel til å måle usikkerheten i produksjonsindeksen. Perioden vi har sett på er januar-desember 2005 og januar-desember 2006. For hver av disse 24 månedene har vi beregnet det estimerte standardavviket $\widehat{\mathrm{st}}\left(\hat{I}_{t,m}-I_{t,m}\right)$ gitt ved (1).

I tillegg til å beregne $\widehat{\mathrm{st}}\left(\hat{I}_{t,m}-I_{t,m}\right)$ har vi beregnet 95% konfidensintervallet $\hat{I}_{t,m}\pm 1.96\cdot\widehat{\mathrm{st}}\left(\hat{I}_{t,m}-I_{t,m}\right)$ og variasjonskoeffisienten. Variasjonskoeffisienten angir hvor stor usikkerheten er i prosent av indeksen, og er gitt ved

$$\frac{\widehat{\mathrm{st}}\left(\hat{I}_{t,m}-I_{t,m}\right)}{\hat{I}_{t,m}}\cdot 100\,.$$

Vi har gjort beregningene på 5 aggregerte nivåer (se tabell 1).[2]

Tabell 1 Aggregeringsnivå

Aggregeringsnivå	Antall sektorer	Utvalgets dekningsgrad i 2006, målt ved omsetning (%)
Industri	108	68.4
Innsatsvarer	52	65.0
Investeringsvarer	23	64.1
Varige konsumvarer	8	65.2
Ikke-varige konsumvarer	30	67.9

Av figur 3, 5, 7, 9 og 11 kan vi tydelig se at for 2006 så er produksjonsindeksen, for alle aggregeringsnivå, en del lavere i juli enn i resten av året. Det samme er tilfelle for 2005. Dette

[2] Vi har også gjort beregningene for aggregeringsnivået industri og bergverksdrift. Resultatet av dette viser at usikkerheten innen industri og bergverksdrift er så å si den samme som innen industri.

indikerer at aktiviteten er en del lavere i juli enn i resten av året, noe som kanskje kan forklares med ferieavvikling.

Av tabell 2 kan vi se at estimert standardavvik for industri ligger mellom 0.59 og 0.87 i 2005, og mellom 0.49 og 0.79 i 2006. Variasjonskoeffisienten ligger mellom 0.59 og 1.05 i 2005, og mellom 0.45 og 0.94 i 2006. Ut fra disse tallene kan vi nok si at det er veldig liten usikkerhet i produksjonsindeksen innen industri. Sammenligner vi usikkerheten mellom de to årene, finner vi at det er lavere usikkerhet i 2006 enn i 2005 (se figur 4).

Ser vi på utviklingen i det estimerte standardavviket utover året, ser vi at $\widehat{\mathrm{st}}\left(\hat{I}_{t,m} - I_{t,m}\right)$ totalt sett øker fram til november, for så å gå litt ned igjen i desember (se figur 4). Denne tendensen ser også ut til å gjelde for de andre aggregeringsnivåene (unntatt for varige konsumvarer). En mulig forklaring på dette kan være at bedriftene har litt forskjellig utvikling utover året. Denne forskjellen vil bli tydeliggere jo lengre ut i året vi kommer, fordi vi da får lengre avstand til perioden vi sammenligner mot. Dette betyr i så fall at det blir større forskjell utover året med tanke på hvordan en bedrifts produksjonsverdi/timeverk har endret seg i forhold til foregående år. Som vi nevnte i kapittel 4.1 får vi større usikkerhet jo større forskjellen er mellom bedriftene. At usikkerheten går ned igjen i desember, må i så fall bety at bedriftene igjen blir mer like. (Fordi utvalg og populasjon er konstant gjennom året, kan ikke endring i disse være med på å forklare endringen i $\widehat{\mathrm{st}}\left(\hat{I}_{t,m} - I_{t,m}\right)$).

Industri er det aggregeringsnivået som har minst usikkerhet. Størst usikkerhet har vi for varige konsumvarer. Her ligger det estimerte standardavviket mellom 2.90 og 7.07 i 2005, og 2.61 og 4.48 i 2006. Variasjonskoeffisienten varierer mellom 2.60 – 6.12 og 2.66 – 7.44 i henholdsvis 2005 og 2006. Dette tyder på at det er en del usikkerhet innen varige konsumvarer, men ikke veldig stor.

At det er industri som har minst usikkerhet og varige konsumvarer som har størst usikkerhet er ikke overraskende. Industri er det høyeste aggregeringsnivået vi ser på, mens varige konsumvarer er det laveste, og som vi nevnte i kapittel 4.1 er det typisk at jo høyere aggregeringsnivå jo mindre usikkerhet.

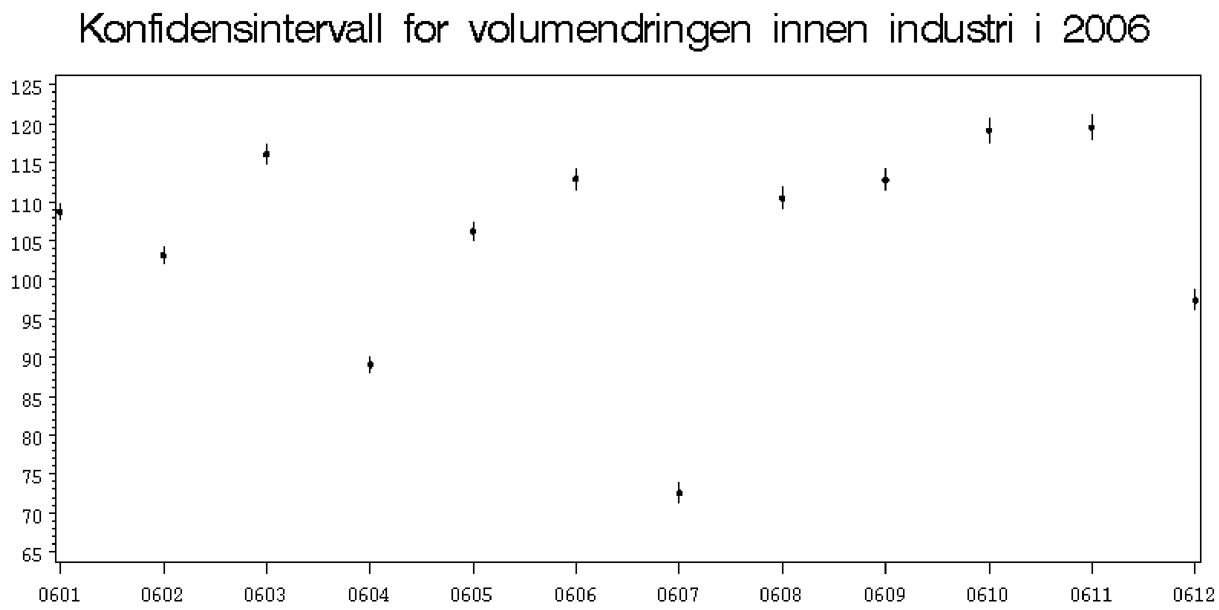
For de tre aggregeringsnivåene innsattsvarer, investeringsvarer og ikke-varige konsumvarer har vi nokså liten usikkerhet. I 2005 ligger f.eks. det estimerte standardavviket mellom 0.90 – 1.57, 0.92 – 1.41 og 1.26 – 1.81, respektivt. De tilsvarende tallene for variasjonskoeffisienten er 0.93 – 1.91, 0.88 – 1.53 og 1.35 – 2.19. I figur 12 ser vi at usikkerheten for ikke-varige konsumvarer er klart lavere i 2006 enn i 2005. For investeringsvarer derimot har vi at usikkerheten er minst i 2005, med unntak av to måneder (se figur 8).

Sammenligner vi usikkerheten til ikke-varige konsumvarer, innsattsvarer og investeringsvarer, finner vi at i 2005 så er det ikke-varige konsumvarer som har størst usikkerhet. Minst usikkerhet har investeringsvarer (med unntak av tre måneder). For 2006 derimot er det ikke-varige konsumvarer som har minst usikkerhet (med unntak av 3 måneder), mens innsatsvarer og investeringsvarer har størst usikkerhet i ca. 6 måneder hver.

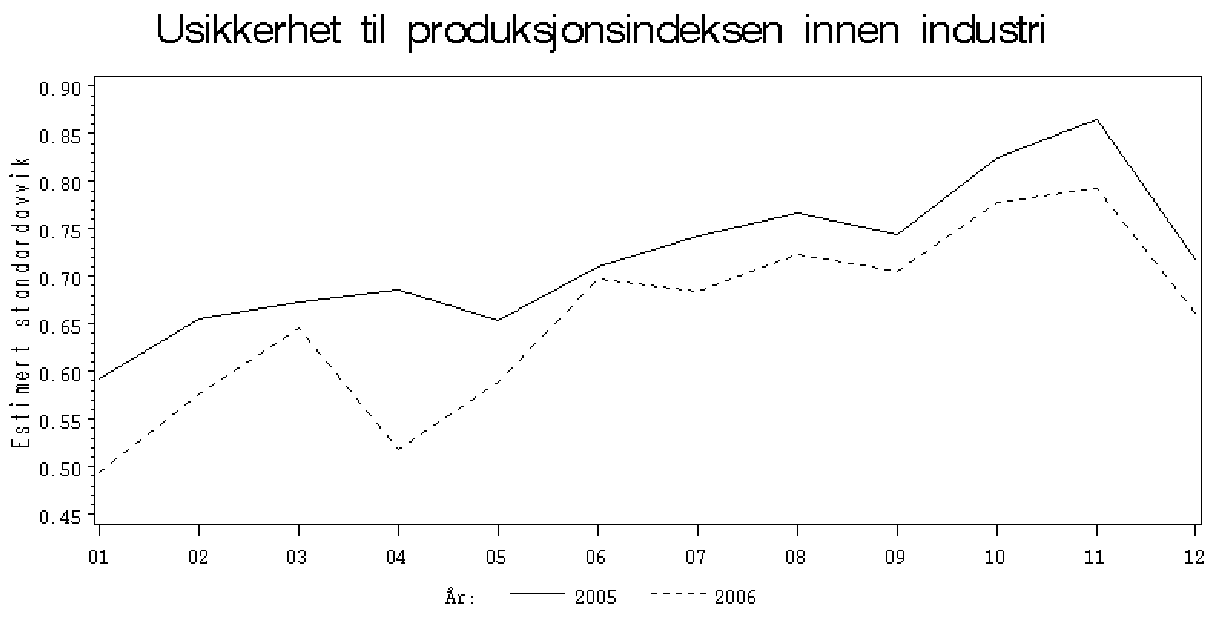
Med et par unntak har vi at variasjonskoeffisienten er en del større i juli enn i de andre månedene. Dette skyldes at produksjonsindeksen er en del mindre i denne måneden sammenlignet med de andre månedene. Dermed blir variasjonskoeffisienten, som angir usikkerheten i prosent av indeksen, en del større uten at det estimerte standardavviket er tilsvarende større.

Tabell 2 Usikkerhet til produksjonsindeksen innen industri

	Produksjons-indeksen	Usikkerhet (estimert standardavvik)	Variasjonskoeffisient (usikkerhet i % av indeksen)	95% konfidensintervall	
				Nedre grense	Øvre grense
2005 01	100.63	0.59	0.59	99.47	101.79
2005 02	99.01	0.65	0.66	97.73	100.29
2005 03	97.52	0.67	0.69	96.20	98.84
2005 04	104.09	0.69	0.66	102.74	105.43
2005 05	97.94	0.65	0.67	96.66	99.22
2005 06	111.09	0.71	0.64	109.70	112.48
2005 07	70.76	0.74	1.05	69.30	72.21
2005 08	103.84	0.77	0.74	102.33	105.34
2005 09	110.92	0.74	0.67	109.47	112.38
2005 10	109.02	0.82	0.76	107.40	110.63
2005 11	113.77	0.87	0.76	112.07	115.46
2005 12	94.40	0.72	0.76	92.99	95.81
2006 01	108.77	0.49	0.45	107.80	109.74
2006 02	103.14	0.58	0.56	102.01	104.27
2006 03	116.11	0.65	0.56	114.84	117.38
2006 04	89.09	0.52	0.58	88.07	90.11
2006 05	106.22	0.59	0.56	105.07	107.38
2006 06	112.95	0.70	0.62	111.58	114.31
2006 07	72.57	0.68	0.94	71.23	73.92
2006 08	110.51	0.72	0.65	109.09	111.93
2006 09	112.83	0.71	0.63	111.45	114.22
2006 10	119.20	0.78	0.65	117.67	120.72
2006 11	119.59	0.79	0.66	118.04	121.14
2006 12	97.36	0.66	0.68	96.06	98.66



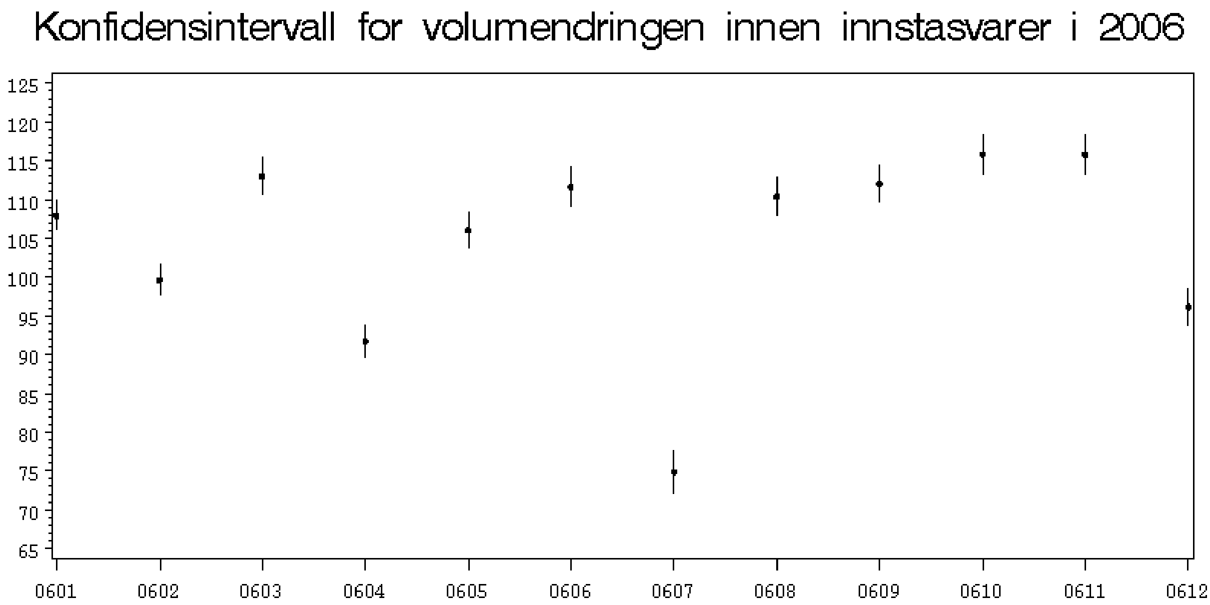
Figur 3 *Figur som viser konfidensintervallet til volumendringen* $I_{t,m}$ *innen industri. Produksjonsindeksen er markert med • og intervallet er markert med |. Langs den horisontale aksen er datoen gitt på formen ååmm.*



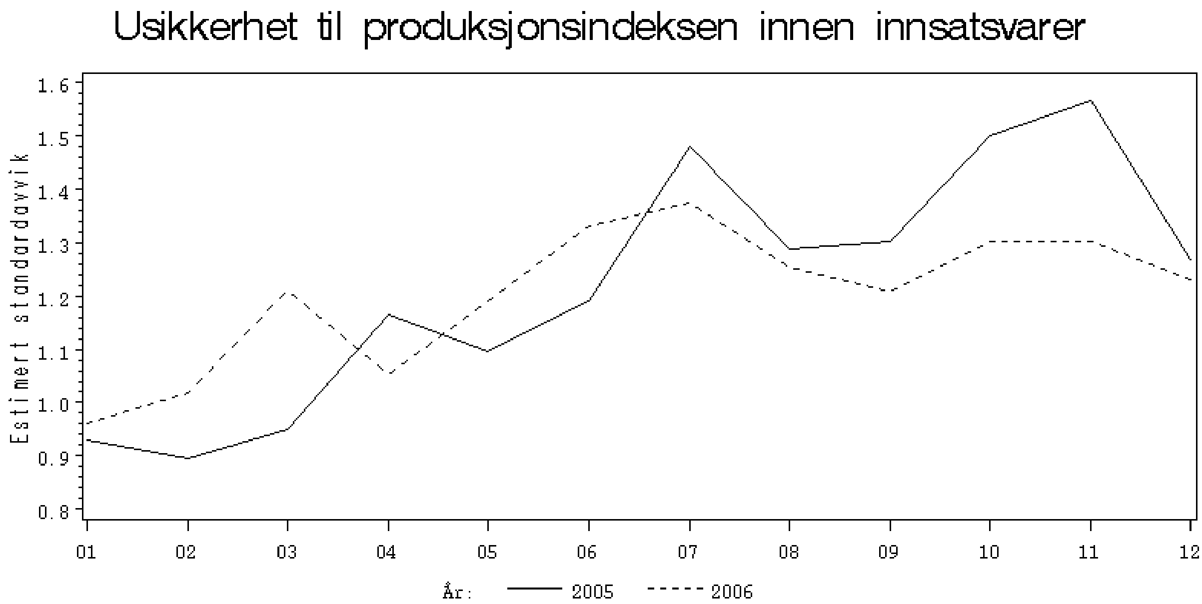
Figur 4 *Figuren viser usikkerheten, dvs. estimert standardavvik, til produksjonsindeksen innen industri. For hver måned (horisontal akse) er usikkerheten i 2005 og 2006 gitt ved henholdsvis den heltrukne og den stiplede linjen.*

Tabell 3 Usikkerhet til produksjonsindeksen innen innsatsvarer

	Produksjons-indeksen	Usikkerhet (estimert standardavvik)	Variasjonskoeffisient (usikkerhet i % av indeksen)	95% konfidensintervall	
				Nedre grense	Øvre grense
2005 01	100.55	0.93	0.93	98.73	102.37
2005 02	95.60	0.90	0.94	93.84	97.36
2005 03	94.88	0.95	1.00	93.02	96.74
2005 04	105.48	1.17	1.10	103.20	107.77
2005 05	99.35	1.09	1.10	97.20	101.49
2005 06	112.32	1.19	1.06	109.99	114.65
2005 07	77.48	1.48	1.91	74.57	80.38
2005 08	105.80	1.29	1.22	103.28	108.33
2005 09	109.79	1.30	1.19	107.23	112.34
2005 10	107.13	1.50	1.40	104.19	110.07
2005 11	111.95	1.57	1.40	108.87	115.02
2005 12	94.01	1.27	1.35	91.53	96.50
2006 01	107.96	0.96	0.89	106.07	109.84
2006 02	99.68	1.02	1.02	97.69	101.68
2006 03	113.02	1.21	1.07	110.64	115.39
2006 04	91.76	1.05	1.15	89.70	93.83
2006 05	106.04	1.19	1.12	103.71	108.37
2006 06	111.65	1.33	1.19	109.04	114.25
2006 07	74.89	1.37	1.84	72.20	77.58
2006 08	110.41	1.26	1.14	107.95	112.87
2006 09	111.98	1.21	1.08	109.62	114.35
2006 10	115.86	1.30	1.12	113.30	118.41
2006 11	115.82	1.30	1.12	113.27	118.37
2006 12	96.18	1.23	1.28	93.77	98.59



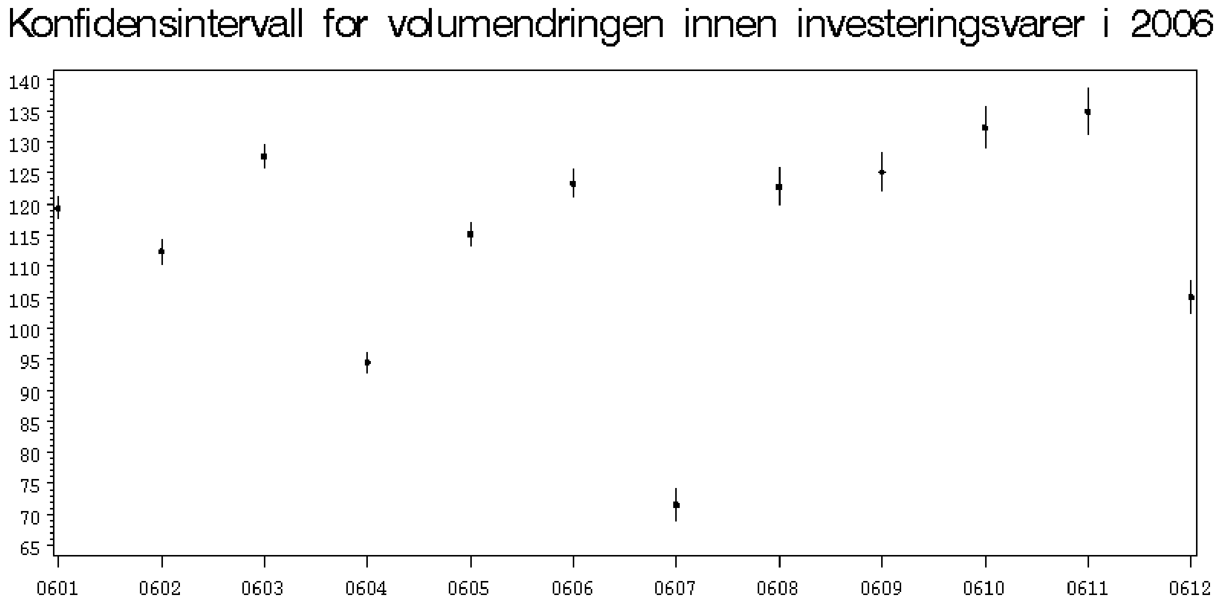
Figur 5 *Figur som viser konfidensintervallet til volumendringen* $I_{t,m}$ *innen innsatsvarer. Produksjonsindeksen er markert med • og intervallet er markert med |. Langs den horisontale aksen er datoen gitt på formen ååmm.*



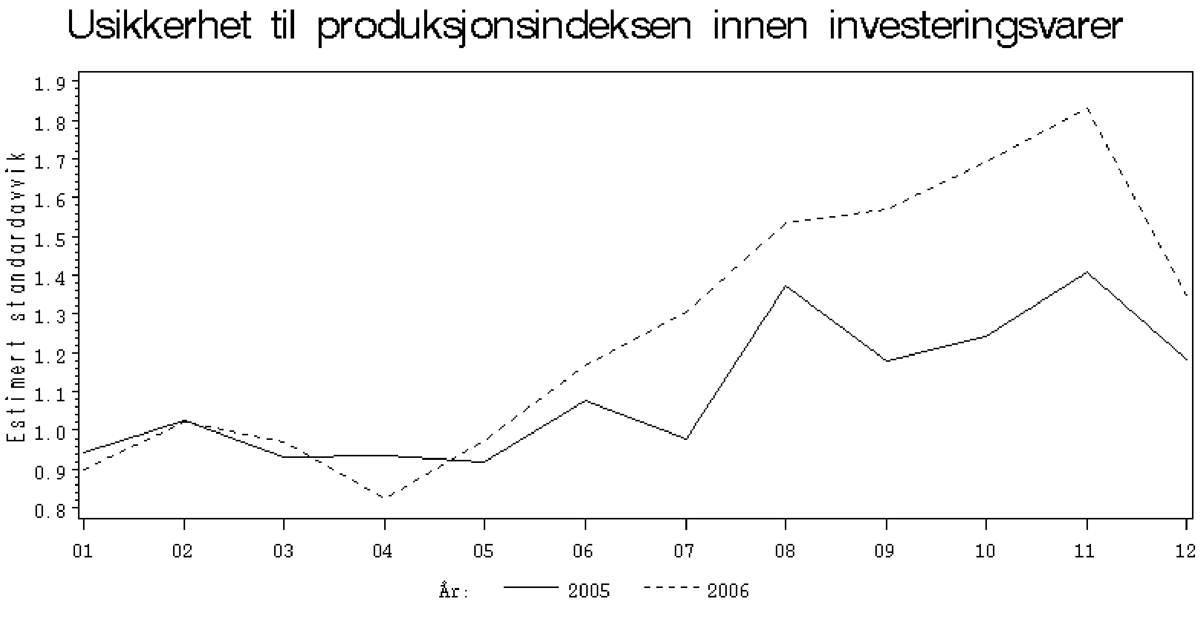
Figur 6 *Figuren viser usikkerheten, dvs. estimert standardavvik, til produksjonsindeksen innen innsatsvarer. For hver måned (horisontal akse) er usikkerheten i 2005 og 2006 gitt ved henholdsvis den heltrukne og den stiplede linjen.*

Tabell 4 Usikkerhet til produksjonsindeksen innen investeringsvarer

	Produksjons-indeksen	Usikkerhet (estimert standardavvik)	Variasjonskoeffisient (usikkerhet i % av indeksen)	95% konfidensintervall Nedre grense	Øvre grense
2005 01	106.83	0.94	0.88	104.98	108.67
2005 02	102.92	1.03	1.00	100.91	104.93
2005 03	99.82	0.93	0.93	98.00	101.64
2005 04	107.23	0.93	0.87	105.40	109.07
2005 05	101.17	0.92	0.91	99.37	102.97
2005 06	114.39	1.08	0.94	112.28	116.50
2005 07	63.99	0.98	1.53	62.08	65.90
2005 08	108.88	1.37	1.26	106.19	111.57
2005 09	116.24	1.18	1.02	113.93	118.56
2005 10	115.02	1.24	1.08	112.59	117.45
2005 11	120.64	1.41	1.17	117.87	123.40
2005 12	99.02	1.18	1.20	96.70	101.34
2006 01	119.43	0.90	0.75	117.66	121.19
2006 02	112.35	1.02	0.91	110.35	114.35
2006 03	127.72	0.97	0.76	125.82	129.63
2006 04	94.43	0.83	0.87	92.81	96.04
2006 05	115.16	0.97	0.85	113.25	117.07
2006 06	123.36	1.17	0.94	121.07	125.64
2006 07	71.54	1.30	1.82	68.99	74.10
2006 08	122.81	1.53	1.25	119.80	125.81
2006 09	125.14	1.57	1.25	122.07	128.22
2006 10	132.36	1.69	1.28	129.05	135.67
2006 11	134.91	1.83	1.36	131.33	138.50
2006 12	105.04	1.35	1.28	102.40	107.68



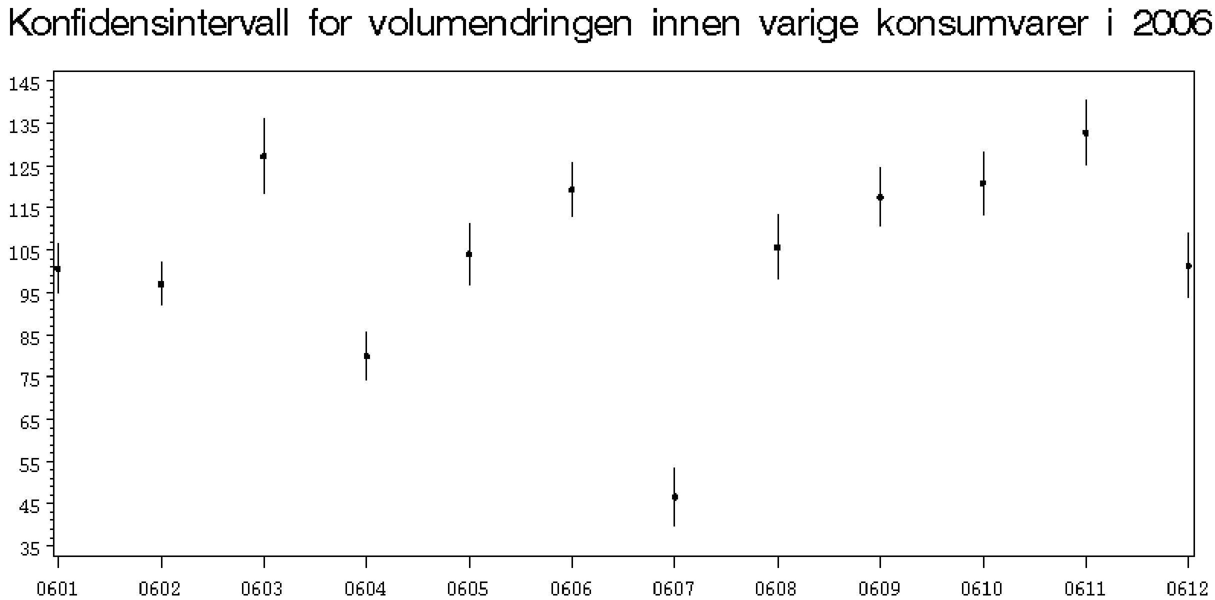
Figur 7 *Figur som viser konfidensintervallet til volumendringen* $I_{t,m}$ *innen investeringsvarer. Produksjonsindeksen er markert med • og intervallet er markert med |. Langs den horisontale aksen er datoen gitt på formen ååmm.*



Figur 8 *Figuren viser usikkerheten, dvs. estimert standardavvik, til produksjonsindeksen innen investeringsvarer. For hver måned (horisontal akse) er usikkerheten i 2005 og 2006 gitt ved henholdsvis den heltrukne og den stiplede linjen.*

Tabell 5 Usikkerhet til produksjonsindeksen innen varige konsumvarer

	Produksjons-indeksen	Usikkerhet (estimert standardavvik)	Variasjonskoeffisient (usikkerhet i % av indeksen)	95% konfidensintervall	
				Nedre grense	Øvre grense
2005 01	105.59	4.39	4.16	96.99	114.19
2005 02	105.29	6.57	6.24	92.41	118.16
2005 03	101.31	6.11	6.03	89.34	113.29
2005 04	116.44	7.07	6.07	102.58	130.30
2005 05	100.37	2.90	2.89	94.70	106.05
2005 06	127.27	4.28	3.36	118.88	135.66
2005 07	54.33	3.32	6.12	47.81	60.84
2005 08	110.79	3.73	3.37	103.47	118.11
2005 09	125.44	3.26	2.60	119.06	131.83
2005 10	106.97	4.35	4.06	98.45	115.49
2005 11	122.54	4.64	3.79	113.44	131.64
2005 12	95.91	4.44	4.63	87.21	104.62
2006 01	100.80	2.99	2.97	94.93	106.67
2006 02	97.06	2.61	2.69	91.95	102.17
2006 03	127.22	4.48	3.52	118.44	135.99
2006 04	79.93	2.87	3.59	74.30	85.56
2006 05	104.11	3.69	3.54	96.88	111.34
2006 06	119.31	3.18	2.66	113.09	125.54
2006 07	46.69	3.47	7.44	39.88	53.49
2006 08	105.69	3.89	3.68	98.06	113.32
2006 09	117.59	3.51	2.99	110.71	124.48
2006 10	120.91	3.77	3.12	113.53	128.29
2006 11	132.86	3.81	2.87	125.39	140.33
2006 12	101.40	3.81	3.76	93.93	108.87

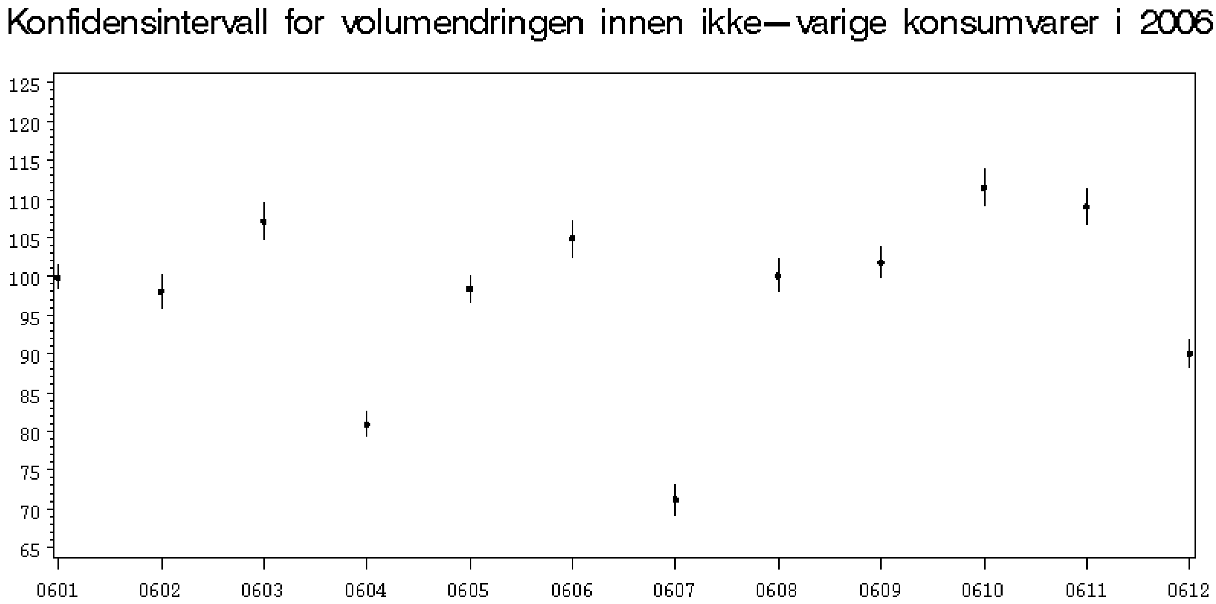


Figur 9 *Figur som viser konfidensintervallet til volumendringen* $I_{t,m}$ *innen varige konsumvarer. Produksjonsindeksen er markert med • og intervallet er markert med |. Langs den horisontale aksen er datoen gitt på formen ååmm.*

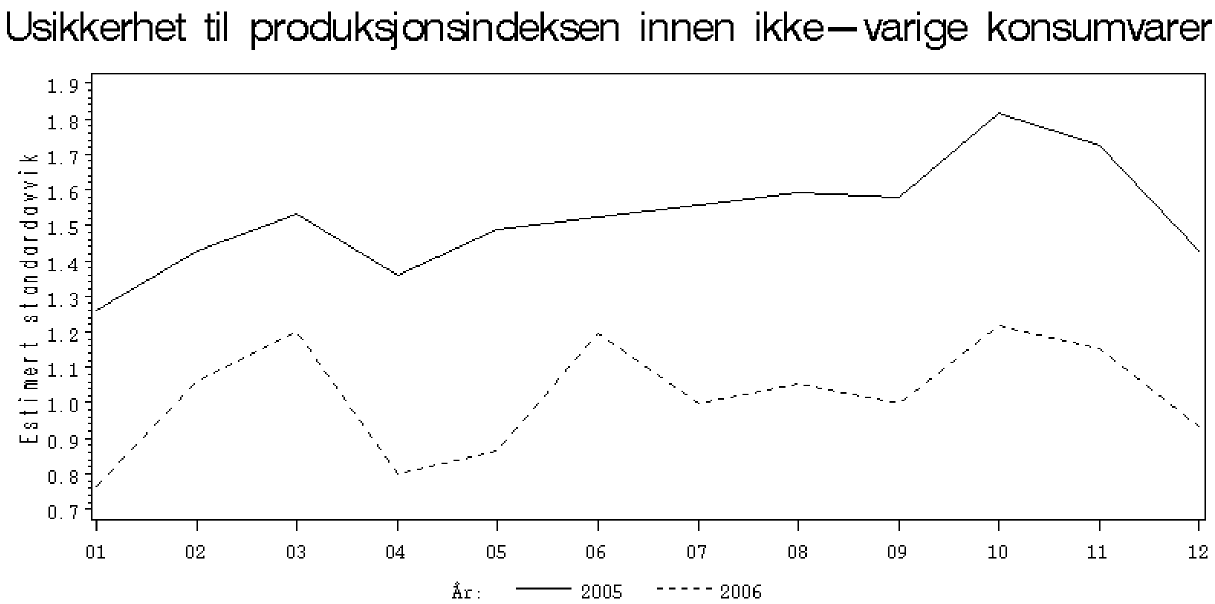
Figur 10 *Figuren viser usikkerheten, dvs. estimert standardavvik, til produksjonsindeksen innen varige konsumvarer. For hver måned (horisontal akse) er usikkerheten i 2005 og 2006 gitt ved henholdsvis den heltrukne og den stiplede linjen.*

Tabell 6 Usikkerhet til produksjonsindeksen innen ikke-varige konsumvarer

	Produksjons-indeksen	Usikkerhet (estimert standardavvik)	Variasjonskoeffisient (usikkerhet i % av indeksen)	95% konfidensintervall	
				Nedre grense	Øvre grense
2005 01	93.51	1.26	1.35	91.04	95.98
2005 02	97.48	1.43	1.47	94.68	100.28
2005 03	96.60	1.53	1.59	93.60	99.60
2005 04	98.99	1.36	1.37	96.32	101.65
2005 05	93.71	1.49	1.59	90.80	96.63
2005 06	104.75	1.52	1.46	101.76	107.74
2005 07	71.10	1.56	2.19	68.04	74.15
2005 08	95.92	1.59	1.66	92.80	99.04
2005 09	105.67	1.58	1.49	102.58	108.76
2005 10	105.75	1.81	1.72	102.19	109.31
2005 11	108.17	1.72	1.59	104.79	111.55
2005 12	90.16	1.43	1.58	87.36	92.96
2006 01	99.91	0.76	0.77	98.41	101.41
2006 02	98.08	1.06	1.08	96.00	100.16
2006 03	107.14	1.20	1.12	104.78	109.49
2006 04	80.91	0.80	0.99	79.34	82.47
2006 05	98.39	0.86	0.88	96.70	100.09
2006 06	104.85	1.20	1.14	102.51	107.20
2006 07	71.21	1.00	1.40	69.25	73.17
2006 08	100.11	1.05	1.05	98.04	102.17
2006 09	101.85	1.00	0.98	99.89	103.81
2006 10	111.48	1.22	1.09	109.10	113.86
2006 11	108.99	1.15	1.06	106.74	111.25
2006 12	90.01	0.93	1.03	88.18	91.83



Figur 11 *Figur som viser konfidensintervallet til volumendringen* $I_{t,m}$ *innen ikke-varige konsumvarer. Produksjonsindeksen er markert med • og intervallet er markert med |. Langs den horisontale aksen er datoen gitt på formen ååmm.*



Figur 12 *Figuren viser usikkerheten, dvs. estimert standardavvik, til produksjonsindeksen innen ikke-varige konsumvarer. For hver måned (horisontal akse) er usikkerheten i 2005 og 2006 gitt ved henholdsvis den heltrukne og den stiplede linjen.*

6. Oppsummering

I dette notatet har vi laget et usikkerhetsmål for produksjonsindeksen. Vi har valgt å si at produksjonsindeksen er en estimator for den størrelsen vi ville fått hvis vi har fulltelling, og målt usikkerheten ut fra dette. Dvs. vi har laget et usikkerhetsmål som sier noe om hvor nær vi kan forvente at produksjonsindeksen er det tallet vi ville fått med fulltelling.

Vi har laget det vi kaller et modellbasert usikkerhetsmål. Det betyr at vi behandler utvalget som gitt, mens vi tenker oss at bedriftenes produksjonsverdi og timeverkbruk er stokastiske variable. Videre har vi valgt å se bort fra mulige feilkilder som frafall og målefeil.

Vi har benyttet usikkerhetsmålet til å beregne usikkerheten i produksjonsindeksen for 2005 og 2006. Resultatet av dette tyder på at hvis aggregeringsnivået omfatter mange nok sektorer, så er det liten usikkerhet i produksjonsindeksen.

Referanser

Bjørnstad, J.F. (1986): *Asymptotisk teori*, instetutt for matematiske realfag, Universitetet i Tromsø.

Chevalier, M. (2003): *Chain Fisher Volume Index Methodology*, Technical serues no. 42, Statistics Canada, Income and Expenditure Accounts Devisinon.

Statistisk sentralbyrå (2003): *Produksjonsindeks for industri og bergverksdrift, olje- og gassutvinning og kraftforsyning 1996-2002*, NOS D 243.

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

Sørensen, E. (1998): *Produksjonsindeks for industrien*, Notater 98/44, Statistisk sentralbyrå.

Zhang, L-C. (2006): *Prisindeksberegninger*, Notater 2006/74, Statistisk sentralbyrå.

Vedlegg

Modellvalg

Basert på data fra 2005 og 2006 har vi kommet fram til at både produksjonsverdien $\sum_j p_{i,j,t-1} q_{i,j,t,m}$ og brukte timeverk $t_{i,t,m}$ stort sett lar seg beskrive veldig godt med ratemodellen, når forrigge års verdier er forklaringsvariabel. Dvs., for bedrifter i produksjonssektorene har vi modellen

$$\mathrm{E}\left[\sum_j p_{i,j,t-1} q_{i,j,t,m}\right] = \beta_g \cdot \sum_j p_{i,j,t-1} q_{i,j,t-1} / 12$$

og

$$\mathrm{V}\left(\sum_j p_{i,j,t-1} q_{i,j,t,m}\right) = \sigma_g^2 \cdot \sum_j p_{i,j,t-1} q_{i,j,t-1} / 12 ,$$

der *g* deler inn bedriftene i 6 grupper, avhenging av hvilken sektor bedriften tilhører.

Gruppe:	Produksjonssektorer som tilhører næring:
1	10, 13 og 14 (Bergverksdrift)
2	15 og 16 (Nærings- og nytelsesmiddelindustri)
3	17-19 (Tekstil- og bekledningsindustri)
4	20 (Trelast- og trevareindustri)
5	21-35 (Innsats- og investeringsvareproduserende industri)
6	36 og 37 (Møbler og annen industri)

For bedrifter i timeverkssektorene har vi

$$\mathrm{E}\left[t_{i,t,m}\right] = \beta_g \cdot t_{i,t-1} / 12$$

og

$$\mathrm{V}\left(t_{i,t,m}\right) = \sigma_g^2 \cdot t_{i,t-1} / 12 ,$$

der *g* deler inn bedriftene i 3 grupper.

Gruppe:	Timeverkssektorer som sorterer under:
7	Innsatsvarer (E1)
8	Investeringsvarer (E2)
9	Konsumvarer (E3 og E4)

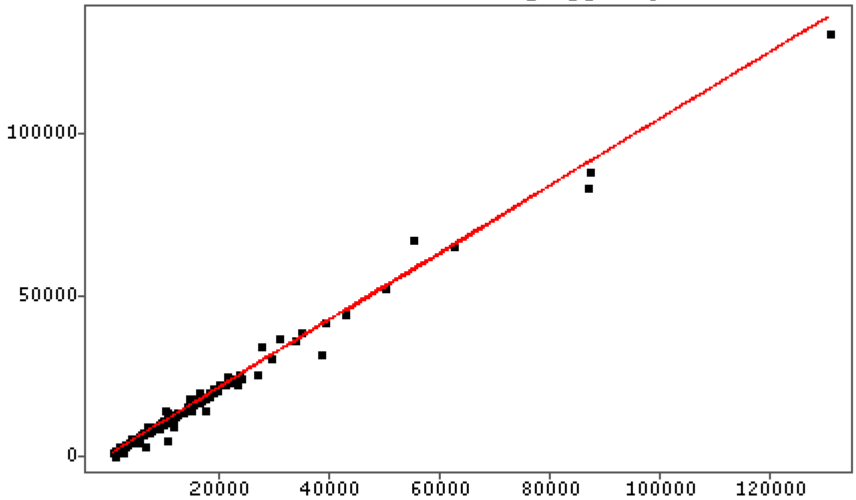
Hvor bra produksjonsverdien og brukte timeverk lar seg beskrive av ratemodellen, varierer fra gruppe til gruppe, og fra måned til måned. Hvis vi bruker R^2 til som mål modelltilpassingen, får vi at gruppe 9 (timeverkssektorer som sorterer under konsumvarer) har aller best tilpassing.[3] For denne gruppen

[3] R^2 er en størrelse som ligger mellom 0 og 1, og er et mål på hvor stor andel av den totale variasjonen i dataene som forklares av modellen. Jo nærmere R^2 er 1, jo mer forklarer modellen av den totale variasjonen.

ligger R^2 mellom 0.91 og 0.989, noe vi må kunne si er veldig bra. Minst bra tilpassing har vi for gruppe 3, dvs. produksjonssektorer i næring 17-19. For denne gruppen varierer R^2 mellom 0.61 og 0.90. For de resterende gruppene ligger R^2 mellom 0.66 og 0.98.

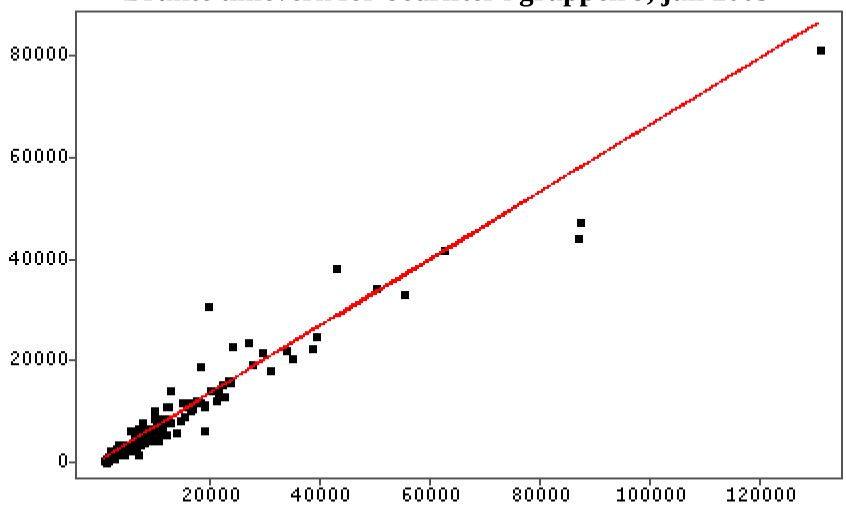
Sammenligner vi modelltilpassingen for en og samme gruppe fra måned til måned, finner vi at det stort sett er juli som har minst R^2. Dvs. juli er den måneden hvor modelltilpassingen er minst bra når vi bruker R^2 som mål.

Brukte timeverk for bedrifter i gruppe 9, januar 2005

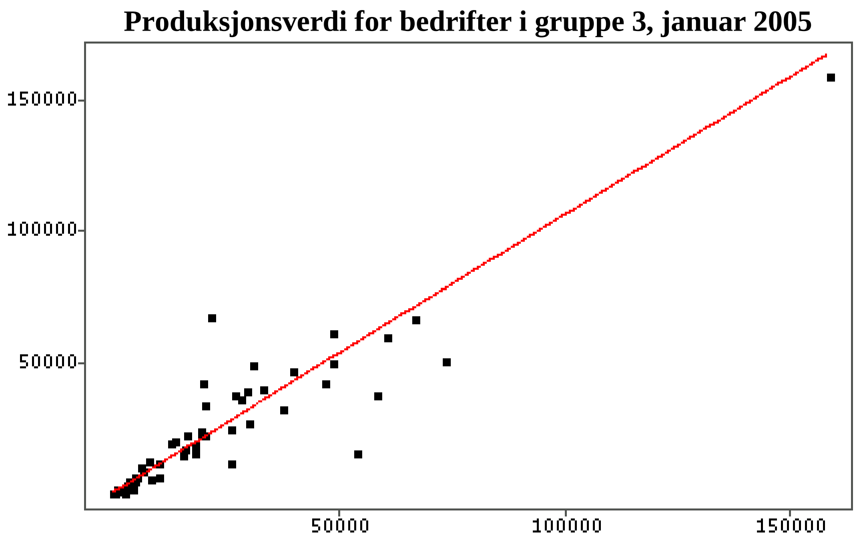


Figur A.1 *Plott som viser sammenhengen mellom* $t_{i,t,m}$ *(vertikal akse) og* $t_{i,t-1}/12$ *(horisontal akse), i måned m = januar og år t=2005. Den heltrukne linjen er regresjonslinjen vi får med ratemodellen.*

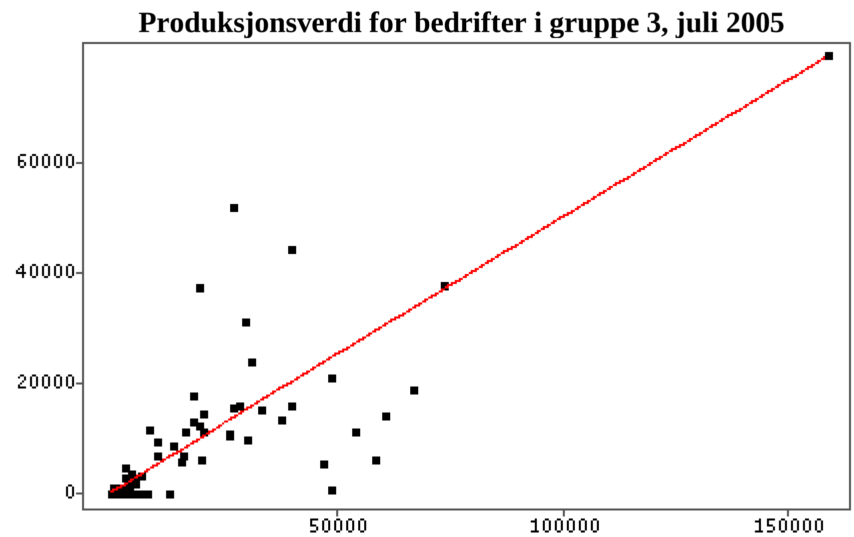
Brukte timeverk for bedrifter i gruppen 9, juli 2005



Figur A.2 *Plott som viser sammenhengen mellom* $t_{i,t,m}$ *(vertikal akse) og* $t_{i,t-1}/12$ *(horisontal akse), i måned m = juli og år t=2005. Den heltrukne linjen er regresjonslinjen vi får med ratemodellen.*



Figur A.3 *Plott som viser sammenhengen mellom* $\sum_j p_{i,j,t-1}q_{i,j,t,m}$ *(vertikal akse) og* $\sum_j p_{i,j,t-1}q_{i,j,t-1}/12$ *(horisontal akse), i måned m = januar og år t=2005. Den heltrukne linjen er regresjonslinjen vi får med ratemodellen.*



Figur A.4 *Plott som viser sammenhengen mellom* $\sum_j p_{i,j,t-1}q_{i,j,t,m}$ *(vertikal akse) og* $\sum_j p_{i,j,t-1}q_{i,j,t-1}/12$ *(horisontal akse), i måned m = juli og år t=2005. Den heltrukne linjen er regresjonslinjen vi får med ratemodellen.*