Notater
Documents
48/2013

Statistics Norway
Statistisk sentralbyrå

Håkon Torfinn Karlsen

Dekomponeringsanalyse under usikkerhet

Notater 48/2013

Håkon Torfinn Karlsen

Dekomponeringsanalyse under usikkerhet

Statistisk sentralbyrå • Statistics Norway
Oslo–Kongsvinger

Notater I denne serien publiseres dokumentasjon, metodebeskrivelser, modellbeskrivelser og standarder.

© Statistisk sentralbyrå
Ved bruk av materiale fra denne publikasjonen skal Statistisk sentralbyrå oppgis som kilde.
November 2013

ISBN978-82-537-8797-8(trykt)
ISBN978-82-537-8798-5(elektronisk)
ISSN1891-5906
Emne:Energi og industri

Trykk: Statistisk sentralbyrå

Standardtegn i tabeller	**Symbol**
Tall kan ikke forekomme	.
Oppgave mangler	..
Oppgave mangler foreløpig	…
Tall kan ikke offentliggjøres	:
Null	-
Mindre enn 0,5 av den brukte enheten	0
Mindre enn 0,05 av den brukte enheten	0,0
Foreløpig tall	*
Brudd i den loddrette serien	—
Brudd i den vannrette serien	|
Desimaltegn	,

Forord

SSB publiserer årlige statistikker over utslipp til luft og energibruk. Det har lenge vært et ønske å forbedre presentasjonen med en mer inngående forklaring av de viktigste drivkreftene spesielt bak utslippsutviklingen, men også for utviklingen av energibruk.

De komplekse sammenhengene som styrer utviklingen, kan sammenfattes i et antall indikatorer eller ”drivere”, for eksempel utslipp av CO_2per enhet forbrukt energi, det generelle aktivitetsnivået i økonomien, endringer i næringsstrukturen og/eller endringer i den sektorvise energiintensiteten (”teknologiendringer”). En dekomponeringsanalyse viser hvordan indikatorene hver for seg og samlet har påvirket utslippene.

SSB har publisert flere rapporter om energiforbruk og utslipp til luft, hvor analysene er helt eller delvis er basert på dekomponering. Vi ønsker i denne rapporten å diskutere hvordan statistisk usikkerhet påvirker slike analyser. Vi vil vise at det er mulig å gi estimater på den stokastiske usikkerheten i analyseresultatet samt hvordan dette påvirker hvilke konklusjoner som kan trekkes.

Rapporten inngår i arbeidet med å forbedre SSBs analyser av energistatistikk og statistikk over utslipp til luft. Miljøverndepartementet har finansiert deler av arbeidet.

Statistisk sentralbyrå, 11. oktober 2013

Hans Henrik Scheel

Sammendrag

SSB publiserer årlige statistikker over utslipp til luft og energibruk. Det har lenge vært et ønske å forbedre presentasjonen med en mer inngående av analyse av de viktigste drivkreftene bak utviklingen. Eksempler på drivkrefter kan være det generelle aktivitetsnivået i økonomien, endringer i næringsstrukturen og endringer i den næringsvise energiintensiteten (”teknologiendringer”).

Dekomponering er en metode som er mye brukt for å tallfeste effekten av endringen i drivkreftene, modellert som indekser. Dekomponeringsanalyser anvendes av bl.a. det internasjonale energibyrået (IEA) i deres analyser av utviklingstrekk i energimarkedet. SSB har også publisert analyser av drivkreftene bak energiutvikling samt utslipp til luft som til dels er basert på dekomponeringsmetoder.

Både SSBs egne rapporter og internasjonale fagartikler foretar dekomponeringsanalyser på observerte data i tidsseriens endepunkter. Man diskuterer i liten grad om disse er representative for den langsiktige tidsutviklingen og hvordan usikre data påvirker beregningsresultatet.

I denne rapporten tas det utgangspunkt i at observerte verdier kan oppfattes som tilfeldige variable. Vi viser en framgangsmåte for å dekomponere effektene av drivkreftenes langsiktige utvikling med et kvantitativt estimat på virkningen av observasjonenes stokastiske usikkerhet.

Vi benytter energi- og produksjonstall for 1990 til 2009 for industrinæringene og tjenesteytende næringer til å beskrive uviklingen i energibruk som summen av tre effekter: Den generelle økonomiske utviklingen, målt som samlet norsk produksjonsverdi (aktivitetseffekten), samt to indekser: Energibruk per produsert krone (energiintensitetsindeksen) og næringsstrukturen, målt som enkeltnæringens andel av total produksjonsverdi (strukturindeks).

Datainnsamlingsmetoder og statistisk bearbeiding gjør at både energi- og produksjonstallene er beheftet med usikkerhet. Denne usikkerheten kan anslås hvis vi antar at endringen i observerte verdier er et uttrykk for en underliggende trend som kan estimeres. Dekomponeringsanalyse basert på trendene flytter oppmerksomheten fra to mer eller mindre tilfeldig valgte tidspunkt til effekten av den langsiktige utviklingen. For å bestemme trendene er det viktig å vurdere om noen observasjoner representerer ekstremverdier i en eller annen forstand, spesielt på eller nær endepunktene i tidsserien. Dersom man for eksempel mener at finanskrisen i 2008 ikke har gitt et varig bidrag til endring av næringsstruktur m.m., kan trenden for en 20-årsperiode fra og med 1990, estimeres ved å behandle 2009 som utligger. Alternativt kan det vurderes om analysens formål er bedre tjent med å utelate ett av ytterpunktene, det vil si å forkorte tidsrommet for analysen. Trendskifter fjernt fra ytterpunktene kan behandles ved å foreta stykkevis estimering av tidsserien.

Trendestimeringen gir i tillegg til estimert verdi også et predikert standardavvik i hvert målepunkt. De observerte verdiene antas å være tilfeldige (stokastiske), hvor den ”sanne” verdien er estimert av trendlinjen. Usikkerheten i dekomponeringsanalysen kan ikke beregnes direkte fordi kovariansen mellom variablene (indeksene) som inngår er ukjent. For å beregne effekten av den stokastiske variasjonen benyttes derfor en Monte Carlo simulering: Dekomponeringsanalysen foretas for et stort antall tilfeldige verdier av energibruk og produksjonsverdi. Fordelingen av analyseresultatene benyttes som et mål for usikkerheten av de estimerte effektene. Denne framgangsmåten krever dog at avviket mellom stokastiske verdi og estimert (”sann”) verdi (residualene) ikke er korrelert.

Innhold

Forord ... **3**

Sammendrag ... **4**

1. Innledning ... **6**

2. Dekomponering med Logarithmic Mean Divisia Index ... **7**

3. Dekomponering uten hensyn til datausikkerhet ... **8**

4. Dekomponering under usikkerhet ... **12**
4.1. Estimering – dekomponering av trender ... 13
4.2. Dekomponering med Monte Carlo simulering ... 15
4.3. Uavhengige trekninger ... 16
4.4. Valg av analyseperiode ... 17

5. Trendskifte og utliggere ... **18**

6. Konklusjoner ... **19**

Referanser ... **20**

Vedlegg ... **21**

1. Innledning

SSB publiserer årlige statistikker over utslipp til luft og energibruk. Det har lenge vært et ønske å forbedre presentasjonen med en mer inngående av analyse av de viktigste drivkreftene bak utviklingen.

De komplekse sammenhengene som styrer utviklingen i utslippene, kan sammenfattes i et antall indikatorer eller "drivere", for eksempel utslipp av CO_2per enhet forbrukt energi, det generelle aktivitetsnivået i økonomien, endringer i næringsstrukturen og/eller endringer i den sektorvise energiintensiteten ("teknologiendringer"). En dekomponeringsanalyse viser hvordan indikatorene hver for seg og samlet har påvirket utslippene.

En oversikt over norske og andre nordiske dekomponeringsanalyser kan finnes i forprosjektrapporten til Nordisk ministerråd (Bruvoll og Fæhn, 2007). Internasjonale organisasjoner, som IEA, benytter dekomponeringsmetoder i sine analyser av verdens energiforbruk. SSB har også publisert flere rapporter om energiforbruk og utslipp til luft, hvor analysene er helt eller delvis basert på dekomponering. Se for eksempel Bøeng, Isaksen, Jama og Stalund (2011).

Både SSBs egne rapporter og internasjonale fagartikler foretar dekomponeringsanalyser på observerte data i tidsseriens endepunkter. Det diskuteres i liten grad hvordan statistisk usikkerhet i datamaterialet påvirker analyseresultatet, eller om dataene er representative for den langsiktige tidsutviklingen.

Yamakava og Peters (2008) diskuterer datausikkerhet i forbindelse med Input-Output analyser og strukturell dekomponering. De anvender ulike teknikker for å fjerne støy i inngangsdataene, samt Monte Carlo simuleringer[1] for å estimere usikkerheten i beregnede data.

I denne rapporten viser vi hvordan deres framgangsmåte kan overføres til indeksdekomponering som er en annen hovedgruppe innenfor dekomponeringsmetoder. Vi tar utgangspunkt i at observerte verdier kan oppfattes som uttrykk for en langsiktig trend og stokastisk usikkerhet. Trender og usikkerhet estimeres og vi beregnereffekten av drivkreftenes langsiktige utvikling samt et kvantitativt estimat på virkningen av observasjonenes stokastiske usikkerhet.

I kapittel 2 gis en kortfattet gjennomgang av den valgte indeksdekomponeringsmetode, LMDI. Kapittel 3 beskriver datakildene og metoden anvendt på norsk energiforbruk fra 1990 til 2009 i industri og tjenesteytende næringer, uten hensyn til datausikkerhet. Kapittel 4 viser hvordan man kan foreta en dekomponeringsanalyse på variablenes underliggende trender og hvordan datausikkerhet kan estimeres og inngå i analysen. Til slutt vises hvordan man kan ta hensyn til at det oppstår mulige trendskifter eller utliggere nær tidsseriens ytterpunkter.

[1] Monte Carlo metoden innebærer at man foretar et stort antall beregninger basert på tilfeldig trukne (stokastiske) data.

2. Dekomponering med Logarithmic Mean Divisia Index

Det finnes en rekke dekomponeringsmetoder. I SSB har man først og fremst benyttet en tilpasset variant av Laspeyres indeksmetode (for eksempel i Bruvoll, Flugsrud og Medin, 2000 og Hass et.al., 2005). I det siste tiåret har imidlertid LMDI (Logarithmic Mean Divisia Index 1) vunnet terreng og er brukt i Bøeng, Isaksen, Jama og Stalund (2011)[2]. LMDI er aktuell for videre arbeid med dekomponeringsanalyser i SSB, først og fremst fordi metoden har egenskaper som løser noen beregningstekniske problemer i analyser basert på for eksempel Laspeyres indeksmetode (se nedenfor).

Indeksdekomponering generelt, og LMDI spesielt, tar utgangspunkt i at en aggregert størrelse, for eksempel samfunnets energibruk ved tidspunkt t, kan beskrives som summen av energibruken i ulike næringer. Næringens energibruk kan modelleres som et produkt av et antall forklaringsvariable (indekser):

$$E(t) = \sum_i I_i^t S_i^t Q^t$$

Her er I_i er energibruk per produsert krone (produksjonsverdi) i næring i, S_i er denne næringens andel av total produksjonsverdi: I tillegg til indeksene per næring, inngår en ”aktivitets”variabel (Q), her representert ved total produksjonsverdi som et mål på den samlede økonomiske aktiviteten.

Dekomponeringsanalysen viser hvor mye av endringen i den samlede energibruk som skyldes endringer i hhv. I, S og Q: $\Delta E = \Delta E_I + \Delta E_S + \Delta E_Q$ hvor

ΔE_I = Intensitetseffekten: Endring i den samlede energibruken som følge av at energibruk per produsert krone har endret seg i en eller flere næringer.

ΔE_S = Struktureffekten: Endring som følge av at det har skjedd en forskyvning i næringenes andel av den totale produksjonsverdien (Strukturelle endringer i økonomien).

ΔE_Q = Aktivitetseffekten: Viser hvor mye av endringen i energibruken som kan tilskrives endringer i samlet produksjonsverdi (økonomisk aktivitetsnivå).

Energiintensitetseffekten ΔE_I representerer summen av effektene fra alle endringer i økonomien som ikke skyldes produksjonsverdi eller strukturendringer. Den kalles derfor også for ”teknologieffekten”.

Med LMDI-metoden, representeres effektene[3] i $\Delta E = \Delta E_I + \Delta E_S + \Delta E_Q$ ved

$$\Delta E_I = \sum_i w_i \ln\left(\frac{I_i^T}{I_i^0}\right) \qquad \Delta E_S = \sum_i w_i \ln\left(\frac{S_i^T}{S_i^0}\right) \qquad \Delta E_Q = \sum_i w_i \ln\left(\frac{Q^T}{Q^0}\right)$$

hvor $w_i = \dfrac{E_i^T - E_i^0}{\ln(E_i^T) - \ln(E_i^0)}$

[2] Se også Ang (2004)

[3] LMDI-metoden sikrer at den samlede observerte endringen i energibruk er summen av de beregnede effekter, dvs. at $\Delta E = \Delta\hat{E}_I + \Delta\hat{E}_S + \Delta\hat{E}_Q$. Andre mye anvendte metoder, for eksempel Laspeyres metode, vil typisk resultere i et uforklart residual: $\Delta E = \Delta\hat{E}_I + \Delta\hat{E}_S + \Delta\hat{E}_Q + \text{residual}$.

og

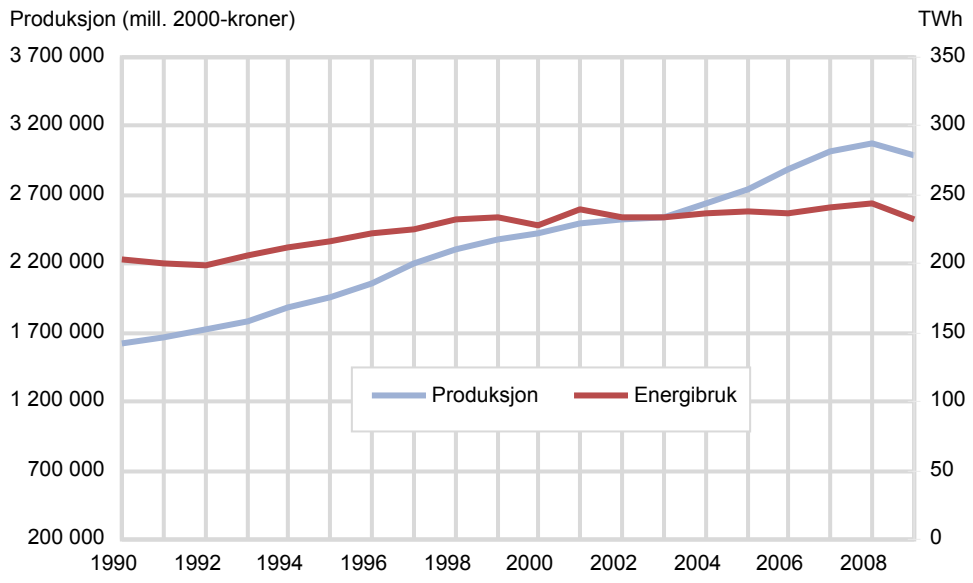
$I_i^t = \frac{E_i^t}{Q_i^t}$	Q Q_i	Samlet produksjonsverdi (aktivitetsmål) Produksjonsverdien i næring i
$S_i^t = \frac{Q_i^t}{Q^t}$	I_i	Energiintensitetsindeksen: Energibruken per produsert krone i næring j
$Q^t \geq \sum_{i=1}^{n} Q_i^t$	S_i	Strukturindeksen: Andelen av produksjonsverdien i næring j

3. Dekomponering uten hensyn til datausikkerhet

SSB-rapporten ”Energiindikatorer for Norge 1990-2009” (Bøeng, Isaksen, Jama og Stalund, 2011) inneholder dekomponeringsanalyser av blant annet energiforbruk i industri og i tjenesteytende næringer. Dette omfatter alle aktive virksomheter innenfor henholdsvis næringene 10-33 og 77-96 (unntatt transportnæringene) etter Standard for næringsgruppering (SN07). Energibruken omfatter ikke energi brukt som råstoff. Vi benytter det samme datagrunnlaget for å illustrere hvordan effekten av datausikkerhet kan estimeres i dekomponeringsanalysen. I motsetning til Bøeng et.al. brukes imidlertid samlet norsk produksjonsverdi som aktivitetsmål, og ikke kun summen av produksjonen i de næringene som inngår i analysen. At tidsperioden inkluderer 2009 gjør det mulig å vurdere om økonomiske ”sjokk” som finanskrisen får betydning for metode og resultater.

I industrinæringene under ett gikk energibruken ned med 3,1 TWh mellom 1990 og 2009. Produksjonsverdien økte med 258 milliarder kroner. For tjenestenæringene under ett økte energibruken med 4,5 TWh og produksjonsverdien med 870 milliarder kroner. Figur 2 viser utviklingen grafisk. Figur 1 viser den samlede produksjonsverdien og energibruken for Norge i samme periode.

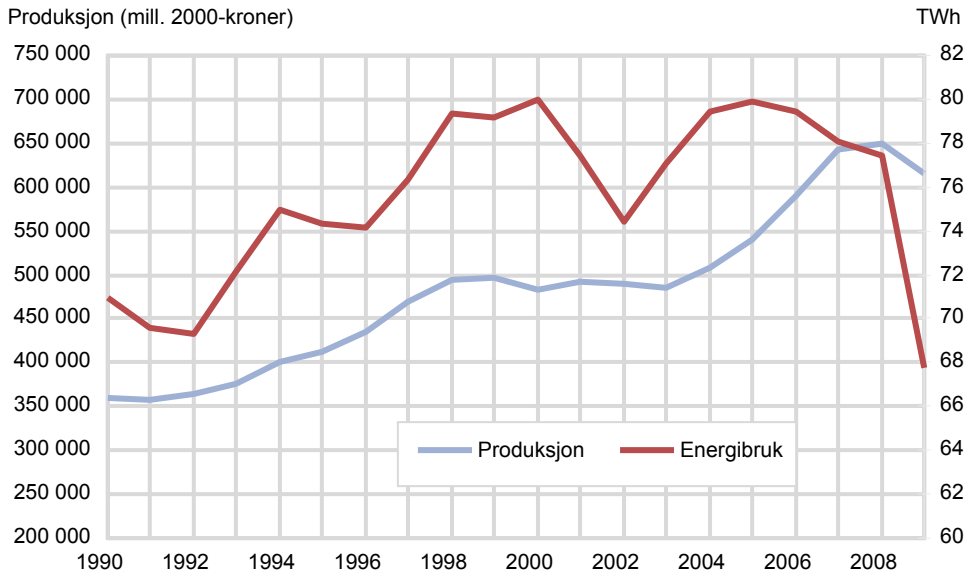
Figur 1. Norsk energibruk og produksjonsverdi. 1990 til 2009



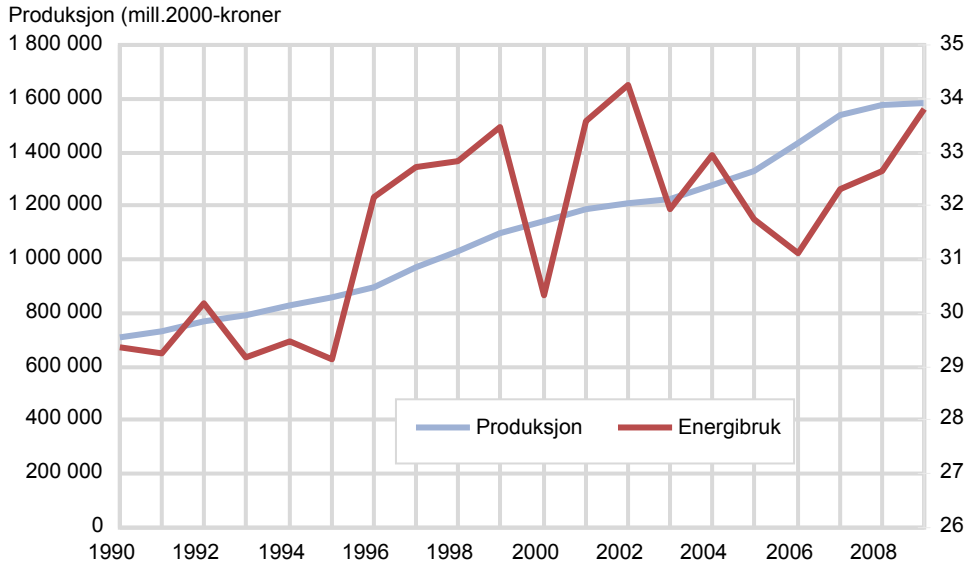
Kilde: Energiregnskapet og nasjonalregnskapet.

Figur 2. Uvikling av energibruken og produksjonsverdi i industri og tjenesteyting. 1990 til 2009

Industrinæringene i alt



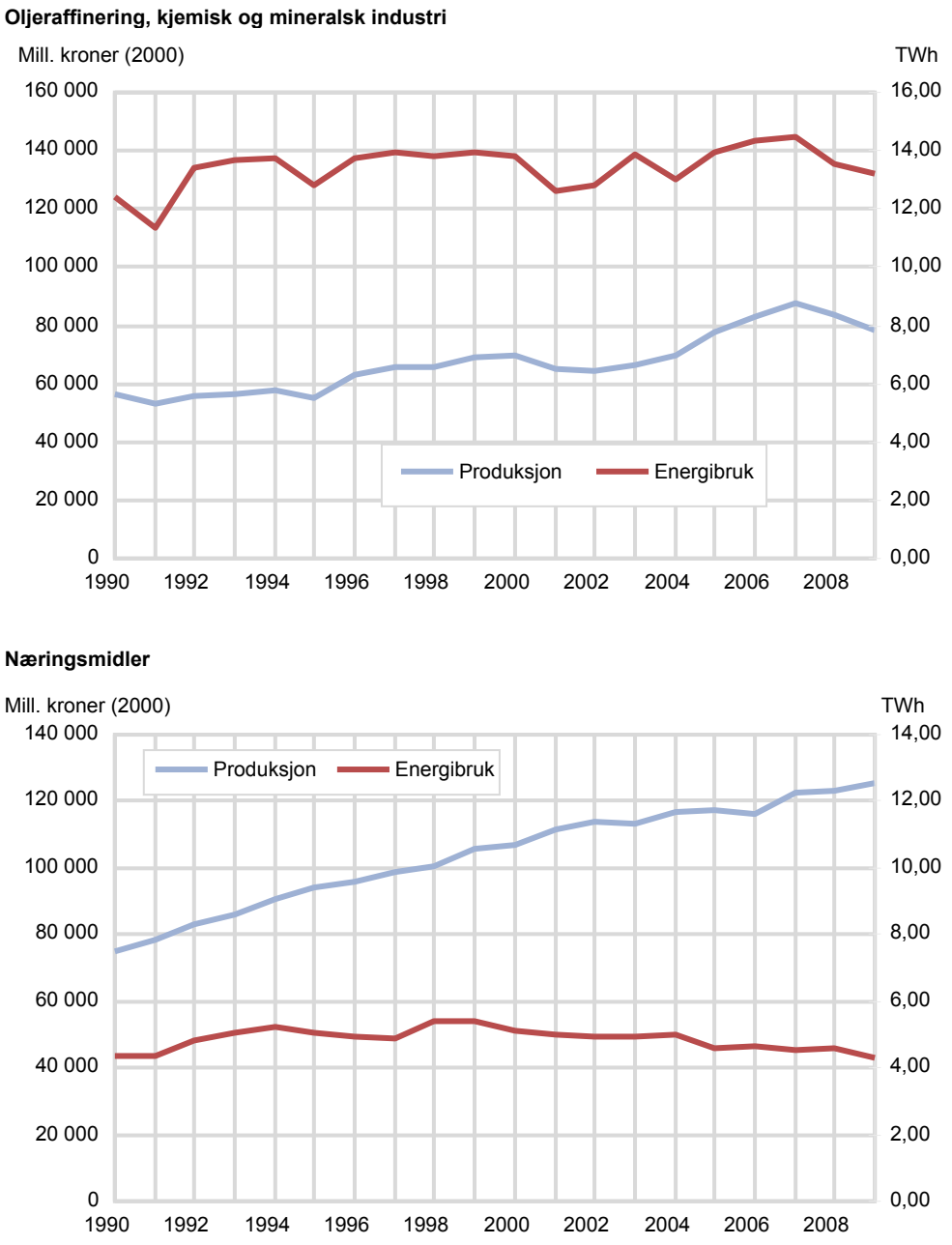
Tjenesteytende næringer i alt



Kilde: Bøeng, Isaksen, Jama og Stalund (2011).

Industrinæringene er samlet i 8 grupper: hvorav to er vist i figur 3. For oljeraffinering, kjemisk og mineralsk industri ser man virkningen av finanskrisen i 2009, mens dette ikke er like tydelig for næringsmiddelindustrien. Figuren viser også at samvariasjonen mellom energiforbruk og produksjonsvolum er ulik for de to industrinæringene.

Figur 3. Uvikling av energibruken og produksjonsverdi i utvalgte industrigrupper. 1990 til 2009

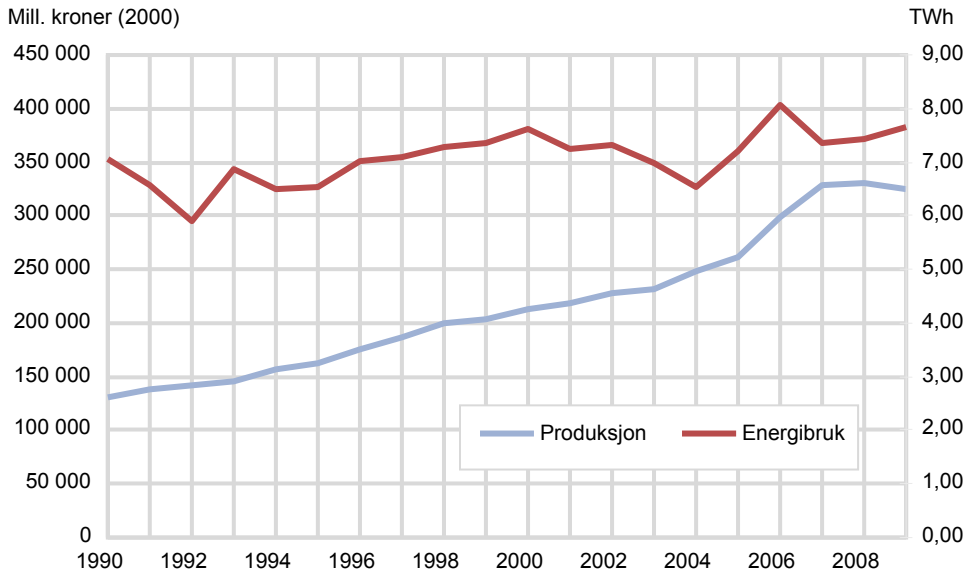


Kilde: Bøeng, Isaksen, Jama og Stalund (2011).

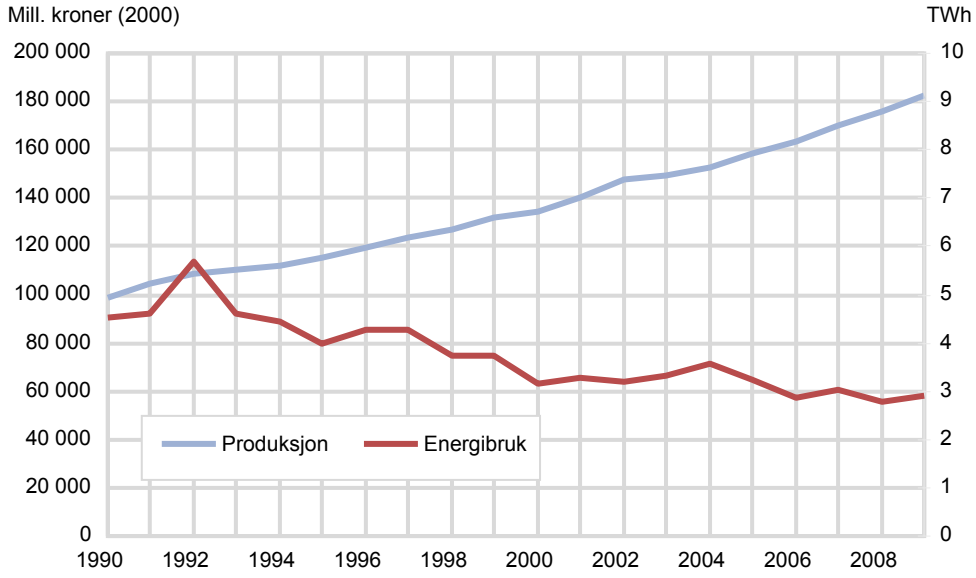
De tjenesteytende næringene er inndelt i 11 grupper. Utviklingen av energibruk og produksjonsverdi er vist i figur 4 for to av disse. Produksjonsverdien øker for begge gruppene, med en utflating fra 2008 til 2009 for varehandel m.m. Siden utflatingen ikke gjenfinnes i helse- og sosialtjenester, som i stor grad er offentlig finansiert, er det mulig at dette er en virkning av finanskrisen i 2008. Energibruken er omtrent konstant i perioden for varehandel m.m., mens den nesten halveres i helse- og sosialtjenester. I de øvrige gruppene er det eksempler på både økende og synkende energiforbruk i perioden, mens produksjonen øker for samtlige.

Figur 4. Uvikling av energibruken og produksjonsverdi i utvalgte tjenesteytende næringer. 1990 til 2009

Varehandel, reparasjon av kjøretøyer mv



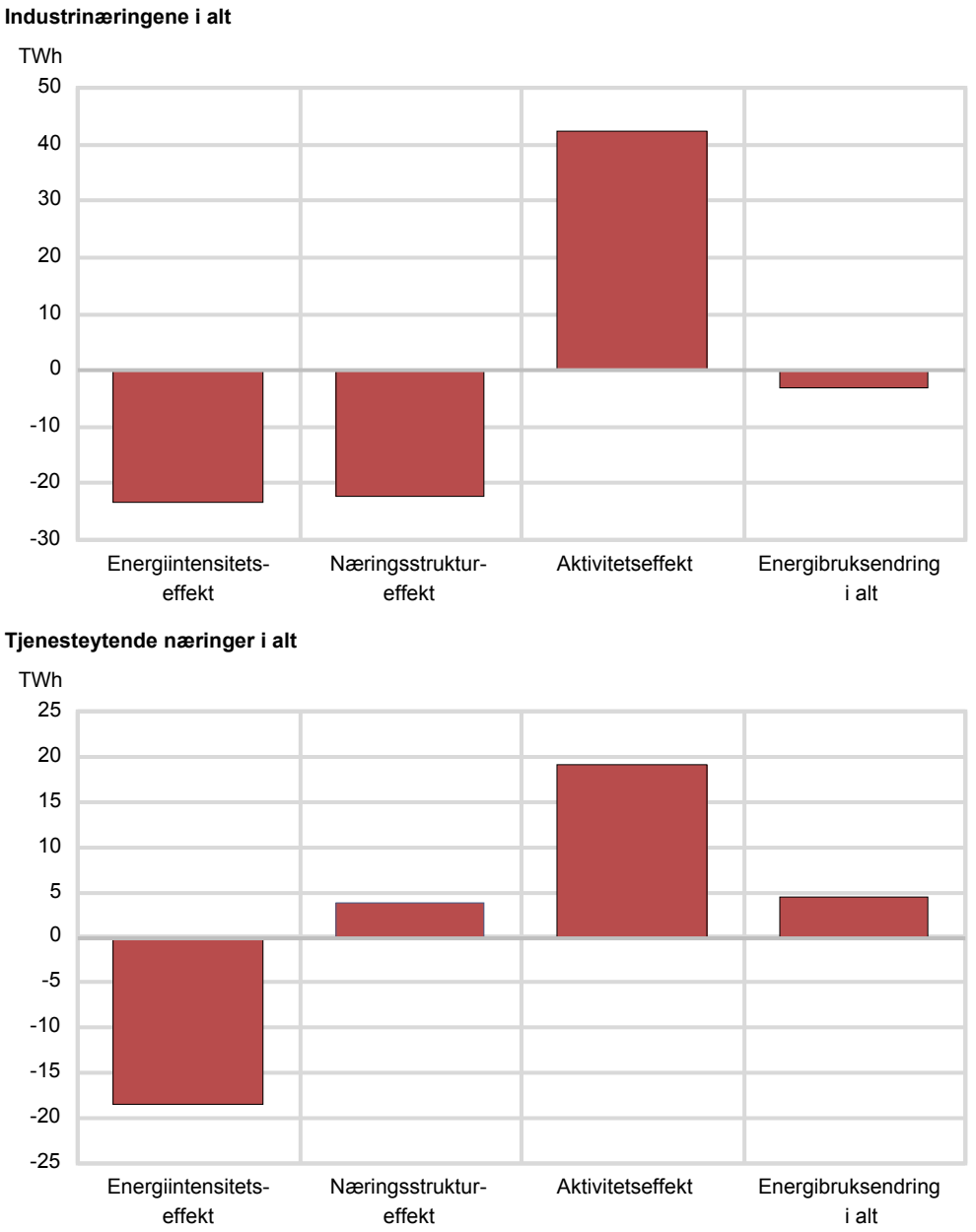
Helse- og sosialtjenester



Kilde: Bøeng, Isaksen, Jama og Stalund (2011)

Figur 5 viser resultatet av dekomponeringsanalysen. I begge næringsgruppene har den viktigste driveren for energibruken vært økt aktivitet, målt ved endringen i samlet norsk produksjonsverdi. Strukturendringer (for eksempel at enkeltnæringers produksjonsandel av samlet produksjon har økt) har relativt sett betydd mer for den samlede energibruksendring i industrien enn i tjenesteyting, mens andre, ikke eksplisitt modellerte faktorer som endrer forhold mellom energibruk og produksjonsverdi (energiintensitetseffekten), har betydd mest i de tjenesteytende næringer. Legg for øvrig merke til at det er ulike skalaer i figuren for de to næringsgruppene. Figuren illustrerer dermed først og fremst at indeksene har ulikt påvirkning på energibruken i de to næringsgruppene.

Figur 5. Dekomponering av energibruken i industrien og tjenesteytende næringer. 1990 til 2009



4. Dekomponering under usikkerhet

Dekomponeringsanalysen beskrevet i forrige kapittel viser blant annet at dersom det ikke hadde skjedd noen endringer i energiintensitet eller næringsstruktur, ville den samlede energibruken i de tjenesteytende næringer ha økt med 19 TWh eller 65 prosent fra 1990 til 2009. Samtlige indekser og aktivitetsnivå er imidlertid usikre størrelser som er framkommet gjennom datainnsamling og statistisk bearbeiding. Dette innebærer at også analyseresultatet vil være omfattet av statistisk usikkerhet.

Alternativt til å forutsette at dataene for 1990 og 2009 er nøyaktig angitt, kan vi anta at forventningsverdien for energibruk og produksjonsverdi følger en kjent utvikling, men hvor dataene også viser målestøy. Dataene kan dermed oppfattes som en tidsserie av tilfeldige variable med en underliggende funksjonell form, og ukjent stokastisk fordeling[4]. Figurene 1-4 antyder for eksempel en underliggende

[4] Generelt vil både fordelingstypen og variansen være ukjent. I denne rapporten vil vi anta at vi står overfor normalfordelinger.

lineær trend, med et mulig unntak for 2008-2009. Estimeres trenden for hver variabel kan vi foreta dekomponeringsanalysen på de estimerte verdier, det vil si at vi finner et mål for den langsiktige effekten av drivkreftene.

Indeksdekomponeringen innebærer beregning av variablenes kvotienter og deres logaritmer. Energibruk og produksjonsverdi for hver næring (E_i og Q_i) er korrelert[5]. Tilsvarende kan man ikke utelukke at $\Delta E_I, \Delta E_S$ og ΔE_Q er korrelert. Men, hvis vi kjenner forventningsverdi og varians for henholdsvis t= 0 og t= T, kan vi benytte en Monte Carlo simulering til å estimere fordelingen til resultatene. Det vil si at vi foretar beregningene på en stor mengde tilfeldige verdier og spredningen av resultatet brukes som uttrykk for den tilfeldige usikkerheten i dekomponeringsanalysen.

Denne tilnærmingen har to viktige forutsetninger:
1. At forventningsverdi og spredning kan estimeres
2. At verdiene som trekkes, er uavhengige av hverandre. Mer presist, at det ikke er noen systematisk samvariasjon (korrelasjon) mellom residualene til hhv. energi og produksjon på et hvert tidspunkt: $Cov(E_i^t, Q_i^t) = 0$

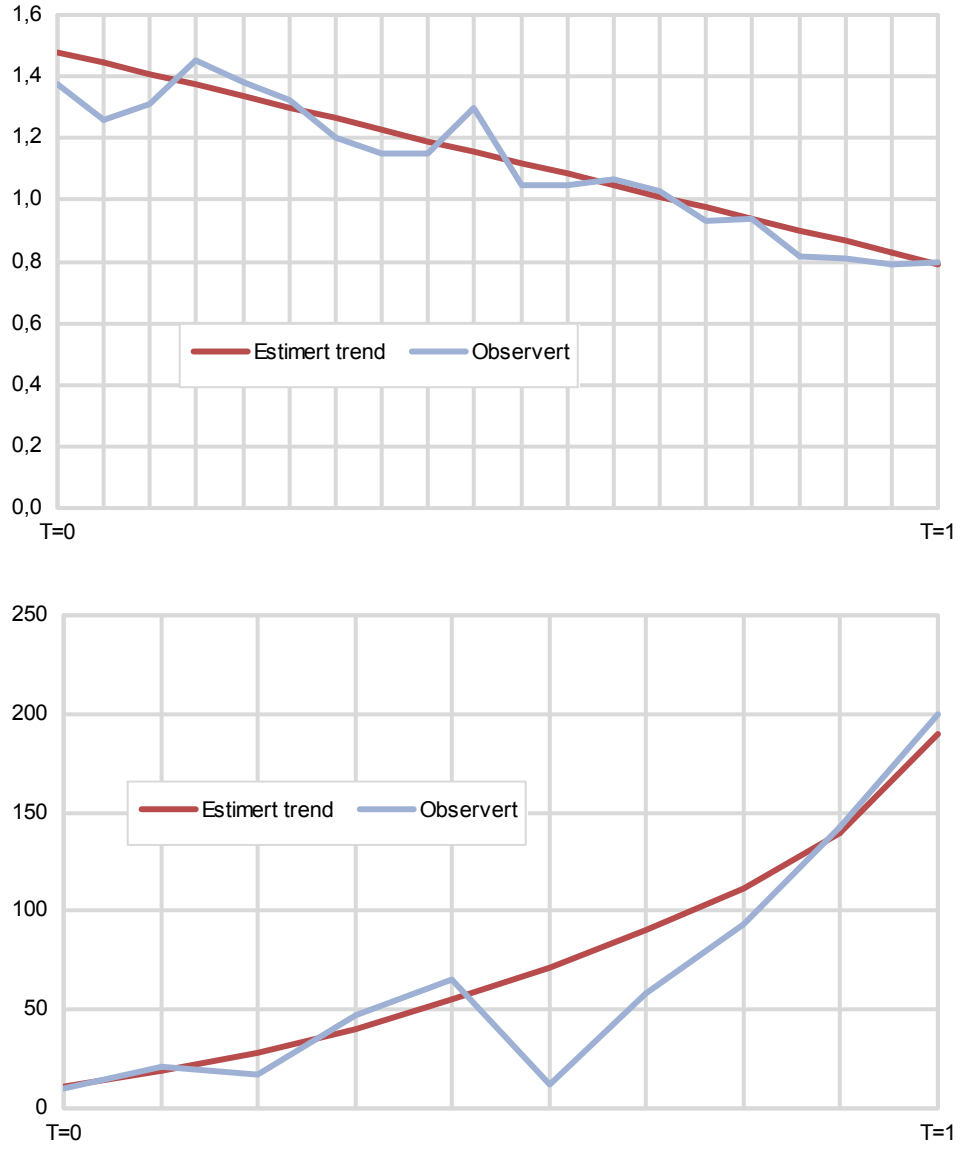
4.1. Estimering – dekomponering av trender

Som en første tilnærming antar vi at både energibruk og produksjonsverdi per næring, som er de underliggende variablene, er lineære funksjoner av tiden, og observerte avvik fra denne (regresjons)linjen representerer stokastisk måleusikkerhet, dvs. at $y(t) = \hat{\alpha} + \hat{\beta}t + \varepsilon$ hvor $\varepsilon = N(0, \hat{s})$ og $\hat{s}$ er det predikerte standardavviket til den estimerte verdien ($\hat{y}$). Mer presist antar vi at $\hat{y}(2009) - \hat{y}(1999)$ er den beste tilnærmingen til endringen i y mellom disse tidspunktene, og at y er normalfordelt i begge punktene, dvs. at $y = N(\hat{y}, \hat{s})$ .Vi ser bort fra eventuell autokorrelasjon, dvs. at ε= ε(t) og eksistensen av utliggere.

Dersom det er grunn til å anta at det har skjedd betydelige strukturelle endringer i perioden, må man vurdere om det er grunnlag for å foreta stykkevis estimering, det vil si å ha ulike verdier av α og β i deler av tidsintervallet. Man bør også vurdere om trenden representeres bedre ved andre funksjonsformer, om det kan være andre fordelinger av feilleddet enn normalfordeling. I figur 6 er dette vist for to hypotetiske sett av variable og aktuelle funksjonsformer.

[5] Det følger av direkte av modellen: E = I * S * Q.

Figur 6. Prinsippskisse for estimat av stokastisk usikkerhet: $y = \alpha + \beta t + \varepsilon$ og $y = \alpha + \beta t^2 + \varepsilon$



Ved å benytte de estimerte verdiene som framkommer med en lineær regresjon, ville energibruken i tjenestenæringene økt med 3,8 i stedet for de observerte 4,5 TWh og produksjonsverdien med 915 mot de observerte 870 milliarder kroner.[6]

Når dekomponeringsanalysen foretas på estimerte verdier får vi et resultat som i figur 7. I tjenestenæringene gir den estimerte aktivitetseffekten gir et litt større bidrag til økt energibruk. Energiintensiteten estimeres lavere enn observert. For industrien er det forskjell mellom estimert og observert effekt av alle tre variablene.

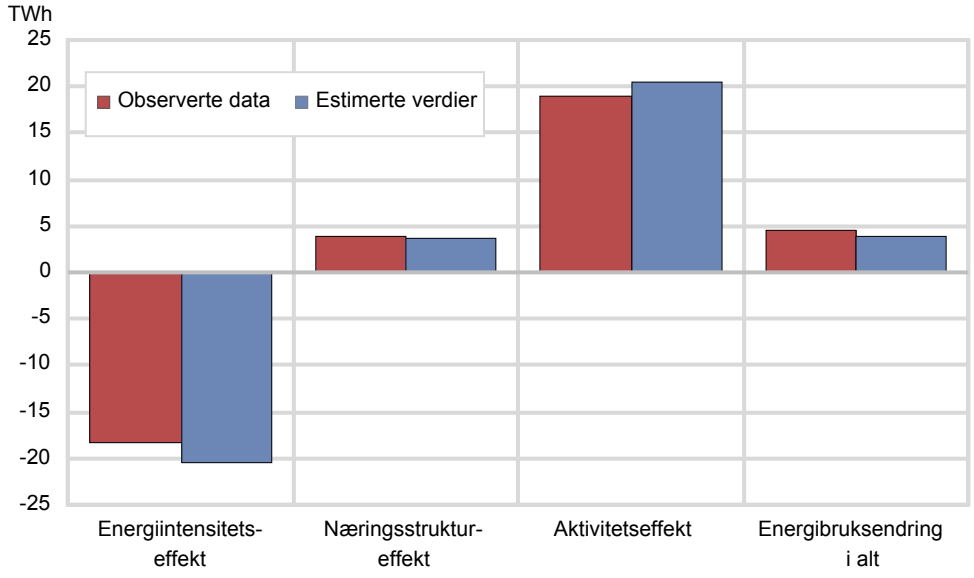
[6] Den enkle regresjonsestimeringen av tjenestenæringene forklarer variasjonen i produksjonsverdi betydelig bedre enn regresjonen av energiforbruket.

For de aggregerte verdiene av energibruk og produksjon er R^2 er hhv. 0,46 og 0,99.

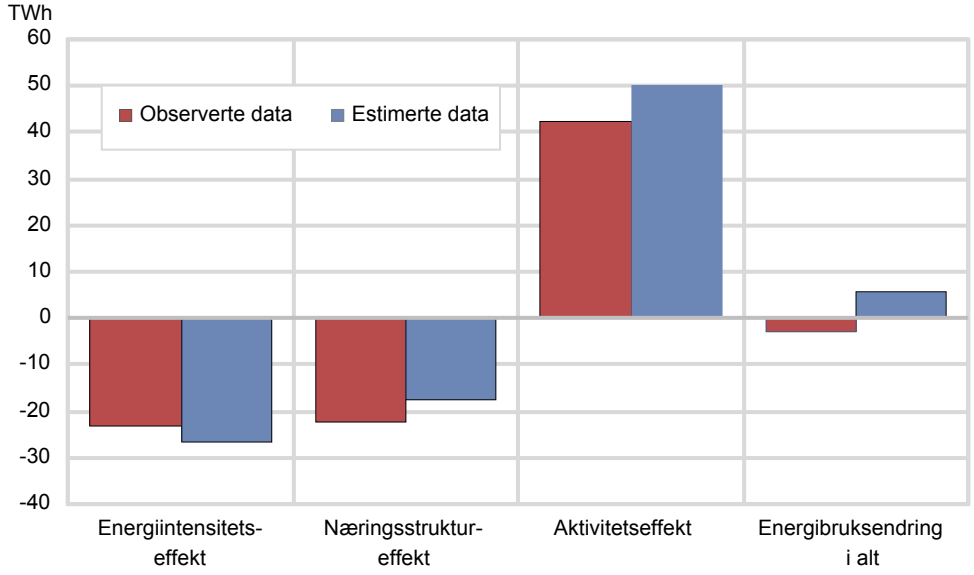
I enkeltnæringene, som benyttes i dekomponeringsanalysen er det forskjeller mellom hvor gode estimater regresjonsanalysen gir: I varehandel er R^2 (energibruk) = 0,42 og R^2 (produksjon) = 0,95 og Tilsvarende tall for finansiell tjenesteyting er hhv. 0,66 og 0,94

Figur 7. Dekomponering av energibruken med observerte og estimerte data. 1990 til 2009

Tjenesteytende næringer i alt



Industrinæringene i alt



4.2. Dekomponering med Monte Carlo simulering

Regresjonen gir en estimert verdi for henholdsvis energiforbruk og produksjon for (del)næringene for hvert tidspunkt, og en usikkerhet gitt ved predikert standard-avvik[7]. Framgangsmåten for å benytte Monte Carlo simulering for å anslå usikkerheten i estimatene blir dermed:

a) Ved t = 0 og t = T: Beregning av forventningsverdi og forventingsverdiens varians for energi og produksjon for hver næring (E_i, og Q_i) samt for den samlede produksjon (Q).

b) Foreta en vurdering av den statistiske fordelingen på disse tidspunktene. (her brukes for enkelhets skyld normalfordeling).

[7] Må ikke forveksles med standardavviket til regresjonskoeffisientene: α og β

c) Resultatet brukes til stokastisk trekning av energi og produksjonstall for hver næring og for total produksjonsverdi på begge tidspunktene.[8]

d) Dekomponeringsanalysen foretas for hver trekning.

Punkt c) og d) gjennomføres mange ganger.

e) Til sammen gir dekomponeringsresultatene en sett av resultater med et gjennomsnitt og varians. Dette brukes som forventningsverdi og anslag for usikkerhet for den totale dekomponeringsanalysen.

Tabell 1. Dekomponering med Monte Carlo simulering. 1990-2009

	Energi-intensitets-effekt	Nærings-struktur-effekt	Aktivitets-effekt	Energibruks-endring i alt
Tjenesteytende næringer				
Observerte data	-18.42	3.87	19.00	4.45
Estimerte verdier	-20.52	3.77	20.53	3.78
Monte Carlo simulering (n=1000)	-20.61	3.97	20.50	3.87
Standardfeil (95% konfidensintervall)	0.14	0.09	0.05	0.11
Standardfeil i prosent	1 %	2 %	0 %	3 %
Industrinæringene				
Observerte data	-23.32	-22.25	42.42	-3.14
Estimerte verdier	-26.49	-17.72	50.03	5.82
Monte Carlo simulering (n=1000)	-26.78	-17.43	49.95	5.75
Standardfeil (95% konfidensintervall)	0.36	0.28	0.12	0.22
Standardfeil i prosent	1 %	2 %	0 %	4 %

Tabell 1 viser at beregningene med estimerte trendlinjer, med og uten usikkerhet, er tydelig forskjellig fra dekomponering av observerte data.

Beregningene med Monte Carlo simulering av datausikkerheten rundt estimerte verdier, gir en lavere økning i den samlede energiforbruket enn analysen av observasjonene. Med et 95 % konfidensintervaller forskjellen mellom de to framgangsmåtene signifikant. Det er også små forskjeller mellom dekomponeringsanalysen basert på trender (estimerte verdier) alene, og med Monte Carlo simuleringene. Dette diskuteres i neste kapittel.

4.3. Uavhengige trekninger

En viktig forutsetning for å kunne benytte Monte Carlo-teknikken er at avviket mellom estimert regresjonslinje og observerte data, ikke viser noen systematisk sammenheng. Med andre ord at residualene (estimert minus observerte verdi) ikke er korrelert.

For å teste denne forutsetningen har vi sjekket om det er en lineær sammenheng mellom residualene:

$\text{residual}(E_i) = \alpha_E + \beta_E \,\text{residual}(Q_i)$ hvor $H_0 : \beta_E = 0$

$\text{residual}(E_i) = \alpha_{EQ} + \beta_{EQ}\text{residual}(Q)$ hvor $H_0 : \beta_{EQ} = 0$

$\text{residual}(Q_i) = \alpha_Q + \beta_Q\text{residual}(Q)$ hvor $H_0 : \beta_Q = 0$

Med et 95 % konfidensintervall, betyr det at vi bør forkaste hypotesen om uavhengighet dersom sannsynligheten for den "kritiske t-verdien" er < 0,025, eller hvis modellen viser høy forklaringsevne gjennom R^2. Dette er tilfellet for de næringene som vist i tabell 2.

[8]For næringer med små forventningsverdier for energi eller produksjon t = 0 eller t = T samt og store standardavvik, kan trekningen gi negative verdier. Disse må fjernes/utelukkes siden LMDI ikke takler negative tall og trekningen vil derfor få en viss grad av skjevhet. I datasettet som benyttes her, har dette problemet ikke oppstått. Hvis antallet trekninger som må fjernes er betydelig, kan det påvirke residualenes uavhengighet og man bør vurdere om annet enn normalfordeling er mer egnet.

Tabell 2. Mulige ikke-uavhengige residualer

	Energi næring vs aktivitet næring		Energi næring vs total aktivitet		Aktivitet næring vs total aktivitet	
	$P(\beta_E > t)$	R^2	$P(\beta_{EQ} > t)$	R^2	$P(\beta_Q > t)$	R^2
Tjenesteytende næringer						
Hotell- og restaurantvirksomhet					0.000	0.670
Tjenester tilknyttet transport	0				0.002	0.427
Post og telekommunikasjon	0.000	0.54			0.024	0.25
Finansiell tjenesteyting	0.000	0.64				
Forretningsmessig tjenesteyting	0.002	0.43			0.007	0.34
Kommunal administrasjon	0.001	0.49				
Industri						
Annen industri	0.000	0.51			0.000	0.62
Oljeraffinering, kjem. og mineralsk industri	0.023	0.26			0.001	0.48
Trelast- og trevareindustri					0.005	0.36
Verksted skipsbygging og oljeplattformer						
Næringsmidler	0.000	0.57				
Treforedling	0.000	0.52				
Kjemiske råvarer					0.009	0.33
Metallindustri	0.000	0.63			0.004	0.37

For 6 av 11 tjenesteytende næringer er det residualer som trolig ikke er uavhengige. I industrien gjelder det alle næringene unntatt verkstedsindustrien. Derimot er residualene for næringens energibruk vs total produksjonsverdi uavhengige for alle næringer.

Imidlertid er forklaringsevnen til denne modellen ikke stor. Den største andelen av forklart variasjon, 67 %, finner vi for sammenhengen mellom produksjonsverdi i hotell- og restaurantvirksomhet og samlet norsk produksjonsverdi.

At det er en svak sammenheng mellom residualene er trolig årsaken til at forventningsverdien for effektene i dekomponeringen med Monte Carlo simulering avviker noe fra effektene beregnet med estimerte verdier (forventningsverdier) for variablene E og Q. Dessuten vekter LMDI utregningen med et logaritmisk middel av energibruken. Logaritmen av en normalfordeling er ikke normalfordelt.

Avhengighetene må tolkes dit hen at den enkle regresjonsanalysen ikke er en fullstendig modell av samvariasjonene mellom energibruk og produksjonsverdig i enkeltnæringene. På grunn av den lave andelen forklart variasjon kan likevel Monte Carlo teknikken være egnet til å angi usystematisk usikkerhet i dekomponeringen for disse datasettene.

Merk at det er forskjell på om residualene viser avhengighet og om det er samvariasjon mellom variablene selv. Dette framkommer tydelig ved å sammenligne figur 3 med tabell 2: Energibruk og produksjon for oljeraffinering m.m. følger hverandres opp- og nedgang, mens det altså ikke er noen sammenheng mellom residualene i det enkelte år.

4.4. Valg av analyseperiode

Dekomponeringsanalysen og Monte Carlo simuleringen benytter estimert verdi og varians for energi og produksjon i ytterpunktene av tidsserien (1990 og 2009). I tabell 3 sammenliknes de estimerte verdier og usikkerheter på flere tidspunkt for varehandelsnæringen (3A) og for forretningsmessig tjenesteyting (3B). Regresjonen er fortsatt foretatt for hele perioden 1990-2009

Tabell 3A: Varehandel – Estimert energibruk (TWh) og produksjonsverdi (mill. 2000-kr)

	Energibruk			Produksjonsverdi		
År	Estimat	Std.feil 95% KI	Std.feil / estimat	Estimat	Std.feil 95% KI	Std.feil / estimat
1990	6,56	0,73	11 %	113 529	27 116	24 %
1991	6,62	0,72	11 %	124 333	26 805	22 %
1992	6,67	0,71	11 %	135 137	26 525	20 %
2007	7,49	0,71	9 %	297 202	26 525	9 %
2008	7,54	0,72	10 %	308 006	26 805	9 %
2009	7,60	0,73	10 %	318 811	27 116	9 %

Tabell 3B. Forretningsmessig tjenesteyting – estimert energibruk (TWh) og produksjonsverdi (mill. 2000-kr)

	Energibruk			Produksjonsverdi		
År	Estimat	Std.feil 95% KI	Std.feil / estimat	Estimat	Std.feil 95% KI	Std.feil / estimat
1990	0,94	0.59	62 %	143 278	30 441	21 %
1991	1,06	0.58	55 %	159 931	30 091	19 %
1992	1,17	0.57	49 %	176 583	29 777	17 %
2007	2,89	0.57	20 %	426 373	29 777	7 %
2008	3,00	0.58	19 %	443 026	30 091	7 %
2009	3,12	0.59	19 %	459 679	30 441	7 %

For varehandelsnæringen vil ikke en forkorting av tidsserien for analysen bety mye for estimatene, mens forskjellene på konfidensintervallene til energibruk i forretningsmessig tjenesteyting er vesentlige.

Hvorvidt dette er ønskelig å forkorte tidsrommet for analysen må sees i sammenheng med hvilken usikkerhet i estimerte verdier (trendene) som er akseptabel for analysen som helhet, og formålet for analysen (hvis det er viktig å få med 1990 - 1991 må man akseptere en større usikkerhet i svaret).

5. Trendskifte og utliggere

Framgangsmåten som beskrevet i forrige kapittel forutsetter at variablene for energi og produksjon kan estimeres for hele perioden, eventuelt oppdelt i delintervaller som estimeres for seg. Av figur 1 ser vi at en markant nedgang fra 2008 til 2009 for den samlede utviklingen av energibruk og produksjonsverdi i industrinæringene. Denne gjenfinnes i for eksempel oljeraffinering, kjemisk og mineralsk industri (figur 3). Siden tallene for 2009 i praksis ikke forelå før i 2010, ”vet vi” at nedgangen trolig skyldes finanskrisen i 2008.

Hvis man mener finanskrisen ikke har gitt et varig bidrag til endring av næringsstruktur m.m. kan en analyse av den langsiktige trenden for en 20-årsperiode fra og med 1990, gjøres ved å behandle 2009 som utligger og estimere regresjonslinjen på grunnlag av 1990 til 2008. Analysen gir da ikke svar på hva de ulike effektene har betydd for perioden fra eksakt 1990 til eksakt 2009. Tabell 5 viser dekomponeringsresultatene av alternativene, og det framgår tydelig at behandlingen av 2009 er helt avgjørende for resultatet. Til sammenlikning vises også dekomponering av 1990-2008, hvor også estimeringen er foretatt for 1990-2008.

Tabell 5. Energibruk i industrien – alternativ behandling av siste år i tidsserien. 95% KI

	Energiintensitets-effekt		Nærings-struktureffekt		Aktivitetseffekt		Sum effekter	
		Std.feil		Std.feil		Std.feil		Std.feil
1990-2009, 2009 ordinært punkt	-26,82	0,34	-17,35	0,28	50,02	0,12	5,86	0,22
1990-2009, 2009 utligger	-26,51	0,26	-15,50	0,20	51,49	0,11	9,48	0,17
1990-2008 observerte data	-24,81		-16,19		47,50		6,50	
1990-2008 estimerte verdier	-25,33		-15,20		49,41		8,88	
1990-2008 med usikkerhet	-25,64	0,26	-14,92	0,21	49,41	0,11	8,85	0,18

6. Konklusjoner

Tradisjonelt har dekomponeringsanalyser benyttet observerte data på to tidspunkt. I dette ligger ingen vurdering av feilmarginer for målefeil i dataene eller om energiforbruket ved tidspunktene er representative for den langsiktige utviklingen.

Vi har her vist at en måte å foreta en dekomponeringsanalyse av de langsiktige trender i underliggende variable. Framgangsmåten framtvinger et bevisst forhold til om det er vesentlige endringer i løpet av tidsserien som ikke er et resultat av tilfeldige svinginger i observerte data. Monte Carlo simuleringen kan videre benyttes til å gi et uttrykk for den usystematiske usikkerheten i effekten av driverne på den langsiktige utviklingen.

Analysen dekomponerer energibruksendringer mellom to tidspunkt (målepunkter). Hva som har skjedd i mellomliggende år benyttes til å anslå den langsiktige utviklingen i variablene som inngår. Om en enkel lineær regresjon er en brukbar tilnærming må sees i sammenheng med hvert enkelt datasett. Det innebærer dog at det fortsatt er ønskelig med så lange tidsserier som mulig. Enkel regresjon er neppe egnet for svært korte tidsserier. Da bør andre estimeringsmetoder vurderes.

Dekomponering av estimerte trender, med eller uten anslag av usikkerhet, innebærer at det er mer presist å snakke om relative endringer forårsaket av de ulike effektene enn absolutte endringer som ikke kan gjenfinnes i grunnlagstallene (observasjonene). Et utsagn som bør unngås, er for eksempel ”energiforbruket har sunket fra nnTWh til mm TWh og zzTWh av nedgangen kan tilskrives økt energieffektivitet”. Mer riktig vil være å si at ”energiforbruket har gått ned med yy prosent fra 1990 til 2009 og zz prosent av dette skyldes mer energieffektiv produksjon”.

Det er spesielt viktig å kartlegge om det har skjedd større endringer nær periodens start eller slutt fra en ellers jevn utvikling[9]. Dersom det av den grunn kan ha oppstått mulige trendskifter nær tidsseriens ytterpunkter, må man presisere nøye om ytterpunktene er behandlet som utliggere. Alternativt kan det vurderes om analysens formål er bedre tjent med å utelate ytterpunktene. I vårt eksempel kan vi utelate 2009 ved å foreta analysen for 1990-2008.

I eksempeldataene er det små forskjeller i analyseresultatene mellom dekomponering av observasjonene og trender med eller uten usikkerhet. Dette kan bero på tilfeldigheter. Analysen uten hensyn til usikkerhet bør alltid utføres og sammenliknes med resultatene fra trendanalysen og Monte Carlo teknikken. Store avvik – eller til og med omvendt fortegn kan være en indikasjon på at dataene bør revideres eller at tidsutviklingen ikke er lineær og bør estimeres med en annen funksjonell form. For eksempel med et 2. grads polynom eller som en eksponensiell utvikling

[9] Dette kommer i tillegg til vurderingen om det bør foretas en stykkevis estimering på grunn av ”trendskifter” inne i tidsserien.

Referanser

Ang, 2004: B.W.Ang: Decomposition analysis for policymaking in energy: Which is the preferred method? Energy Policy 32 (2004).

Bruvoll, Flugsrud og Medin, 2000:Annegrete Bruvoll, Ketil Flugsrud og Hege Medin: Økonomisk vekst treng ikkje gi dårlegare miljø. Økonomiske analyser 9/99. Statistisk sentralbyrå

Bruvoll og Fæhn, 2007: Annegrete Bruvoll og Taran Fæhn: Driving forces behind environmental pressure in Nordic countries – comparative decomposition analyses. Nordisk ministerråd 2007

Bøeng, Isaksen, Jama og Stalund 2011: Ann Christin Bøeng, Elisabeth Isaksen, Sadiya M. Jama og Marita Stalund: Energiindikatorer for Norge 1990-2009. SSB rapporter 31/2011. Statistisk sentralbyrå 2011

Hass et.al. 2005: Julie L. Hass, Knut Ø. Sørensen, Tone Smith, Kristine Erlandsen Kolshus, Ole Magnus Jakobsen:
Norwegian greenhouse gas emissions - Pilot studies for the development of Norwegian Economic and Environmental Accounts (NOREEA) 2005

Yamakava og Peters, 2008: AsukaaYamakawa, Glen Peters:
Environmental Input-Output Analysis, Structural Decomposition Analysis and Uncertainty.Meeting paper: Input - Output & Environment, Seville (Spain), July 9 - 11, 2008

Vedlegg

Tjenesteytende næringer. Energi og produksjonsverdi. 1990 og 2009

	Energibruk (Twh)		Produksjonsverdi (mill. 2000-kr)	
	1990	2009	1990	2009
Varehandel, reparasjon av kjøretøyer mv.	7,1	7,6	130 066	325 536
Hotell- og restaurantvirksomhet	1,5	2,0	28 023	46 847
Tjenester tilknyttet transport	1,0	1,3	28 798	72 358
Post og telekommunikasjon	1,9	1,1	15 413	88 404
Finansiell tjenesteyting	1,4	0,8	47 584	109 980
Forretningsmessig tjenesteyting	1,5	3,4	173 502	455 851
Undervisning	3,6	2,9	54 509	87 567
Helse- og sosialtjenester	4,5	2,9	98 544	182 447
Andre sosiale og personlige tjenester	3,0	6,4	39 994	84 474
Statlig administrasjon og forsvar	2,9	3,1	67 359	86 889
Kommunal administrasjon	1,1	2,3	28 084	41 580
Tjenesteytende næringer og varehandel i alt	29,4	33,8	711 876	1 581 933

Kilde: Bøeng, Isaksen, Jama og Stalund 2011

Industri. Energi og produksjonsverdi. 1990 og 2009

	Energibruk (Twh)		Produksjonsverdi (mill. 2000-kr)	
	1990	2009	1990	2009
Annen industri	1.1	1.1	44 888	55 237
Oljeraffinering, kjem. og mineralsk industri	12.4	13.2	56 339	78 168
Trelast- og trevareindustri	2.4	1.6	16 870	17 547
Verksted skipsbygging og oljeplattformer	2.9	2.7	92 705	248 380
Næringsmidler	4.4	4.3	75 197	125 570
Treforedling	10.3	8.4	18 493	17 267
Kjemiske råvarer	10.5	11.8	19 574	27 698
Metallindustri	27.0	24.5	34 848	46 560
Industri i alt	70.9	67.8	358 914	616 426

Kilde: Bøeng, Isaksen, Jama og Stalund 2011

Norges samlede energibruk og produksjonsverdi, 1990-2009

	Energibruk (TWh)	Produksjonsverdi (mill. 2000-kr)
1990	204	1 620 510
1991	200	1 667 144
1992	199	1 724 481
1993	206	1 780 625
1994	212	1 881 271
1995	216	1 955 342
1996	222	2 058 773
1997	225	2 203 561
1998	233	2 305 013
1999	234	2 375 904
2000	228	2 422 501
2001	239	2 497 602
2002	234	2 525 542
2003	234	2 542 705
2004	237	2 643 517
2005	237	2 747 327
2006	236	2 883 983
2007	241	3 020 653
2008	243	3 076 800
2009	232	2 989 524

Kilde: SSB, nasjonalregnskapet

B

Returadresse:
Statistisk sentralbyrå
NO-2225 Kongsvinger

Avsender:
Statistisk sentralbyrå

Postadresse:
Postboks 8131 Dep
NO-0033 Oslo

Besøksadresse:
Kongens gate 6, Oslo
Oterveien 23, Kongsvinger

E-post: ssb@ssb.no
Internett: www.ssb.no
Telefon: 62 88 50 00

ISBN 978-82-537-8797-8 (trykt)
ISBN 978-82-537-8798-5 (elektronisk)
ISSN 1891-5906

ISBN 978-82-537-8797-8
9 788253 787978

Statistisk sentralbyrå
Statistics Norway

Statistisk sentralbyrå

48/2013

Dekomponeringsanalyse under usikkerhet

Design: Siri Boquist