

Sammendrag

Forbedringspotensial ved måling av effektivitet i kommunal sektor

TOI rapport 1879/2022

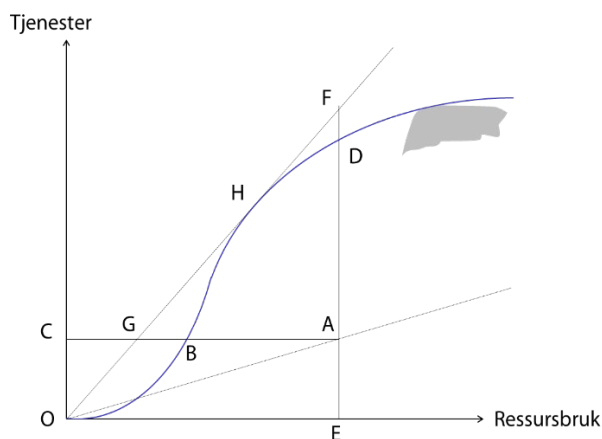
Forfattere: Kenneth Løvold Rødseth, Rasmus Bøgh Holmen, Sverre Kittelsen og Finn Førsund

Oslo 2022 89 sider

Teknisk beregningsutvalg for kommunal og fylkeskommunal økonomi (TBU) ønsket et prosjekt som vurderer mulighetene for og den praktiske anvendbarheten til effektivitetsanalyser av offentlig tjenesteyting. Transportøkonomisk institutt og Frisch-senteret fikk oppdraget. TBU ba om at det skal gjøres en vurdering av dagens metoder for å analysere effektiviteten i kommunene og effektivitetsutviklingen i kommunene som Senter for økonomisk forskning (SØF) gjør årlig for TBU. De ulike metodenes evne til å analysere effektivitetsutviklingen for de største tjenesteområdene samlet over tid og for den enkelte kommune over tid skal vurderes, samt metodenes evne til å sammenlikne effektiviteten mellom kommuner. Det skal gis anbefalinger om hvordan fremtidige effektivitetsanalyser bør gjennomføres enten ved bruk av de metodene som benyttes i dag eller ved å ta i bruk andre metoder. I sammendraget har vi fremhevet konklusjoner og anbefalinger i fet skrift.

Effektivitetsmål

Mens SØFs analyser er fokusert på *effektivitet*, er det vanlig i litteraturen også å være opptatt av *produktivitet*. Slik begrepene brukes i denne rapporten kan forskjellen mellom disse illustreres i en figur:



Figur S.1: Effektivitet og produktivitet.

Figur S.1 ser på tilfellet hvor en tjeneste produseres med en innsatsfaktor eller ressurs. Vi ser på en hypotetisk observert kommune A. Denne kommunen produserer et tjenestevolum lik avstanden OC og bruker OE innsatsfaktor eller ressurs.

Den blå kurven beskriver beste praksis ved gitt teknologi, dvs. effektiv produksjon eller *fronten*. Denne kan estimeres ved ulike metoder; her er den vist som en glatt kurve, mens den i f.eks. data omhyllingsanalyse (Data Envelopment Analysis på Engelsk, forkortet til DEA) vil være en stykkevis lineær front. Kommunens produktivitet (dvs. produkt/innsatsfaktor) er gitt av helningen på den stiplede linjen OA gjennom datapunktet A. Produktiviteten kan angis som selve forholdstallet (f.eks. elever per lærer), men kan også normaliseres

ved å se produktiviteten i forhold til en fast referanse, f.eks. størst mulig produktivitet. I figuren er den størst mulige produktiviteten tegnet som den rette linjen OGHF, siden produktiviteten i punktet H er den høyest mulige med denne teknologien. Når vi her baserer oss på teknologien kalles den normaliserte produktiviteten gjerne *teknisk produktivitet* (TP) og er forholdet mellom helningene på de to linjene. Dette kan forstås som den faktiske tjenesteproduksjonen per enhet ressursbruk i forhold til den høyst mulige tjenesteproduksjonen per enhet ressursbruk. Det kan vises at dette geometrisk kan beregnes som $TP=CG/CA$.

Ressursbesparende *teknisk effektivitet* er derimot et mål på avstanden fra datapunktet til fronten, som her kan beregnes i innsatsfaktorbesparende retning som $TE=CB/CA$. Dette er nødvendig ressursbruk i forhold til faktisk ressursbruk, gitt størrelsen på tjenesteproduksjonen. Forskjellen på de to målene skyldes at kommune A ikke er av optimal størrelse, dvs. at selv om kommunen hadde vært effektiv og redusert sin ressursbruk til avstanden CB ville den fortsatt hatt lavere produktivitet enn om den hadde vært større. Vi definerer derfor ressursbesparende skalaeffektivitet som forholdstallet mellom teknisk produktivitet og teknisk effektivitet $SE=TP/TE=CG/CB$. Alle tre mål vil ha verdier fra 1 og nedover, der f.eks. effektivitet lik 1,0 (eller 100%) tilsier at en kommune er fullt effektiv mens verdier lavere enn 1,0 tilsier at kommunen anslås til å være mer eller mindre ineffektiv.

Dersom det faktisk er konstant skalutbytte i sektoren vil teknologifronten være en rett linje som sammenfaller med linjen maksimal produktivitet. Da vil teknisk effektivitet være lik teknisk produktivitet, og dermed vil også skalaeffektiviteten være 1 over alt.

Metoden som SØF bruker til å estimere fronten er ikke-parametrisk og kalles DEA.

Metoden baseres på å sammenlikne observerte data for produkter og innsatsfaktorer for kommunale tjenestesektorer mot en referansefront som bygger på effektive kommuner.

I avsnitt 2.1 vises de matematiske formuleringene av DEA-metoden som lineære programmeringsproblemer (LP-problemer), og det drøftes spesielt hvordan en kan håndtere ulike variabeltyper i en DEA-analyse. I øvrige metodavsnitt beskrives metoder for å analysere utviklingen over tid, bruk av produksjonsindeksen til å måle effektivitet og produktivitet, samt enkelte alternative metoder. Deretter gir kapittel 3 en gjennomgang av datagrunnlaget og kapittel 4 sammenligner empiriske resultater ved bruk av SØFs metodikk med ulike alternativer. I sammendraget drøfter vi datagrunnlaget og metodenes styrker og svakheter sammen med de empiriske analysene.

De tre sektorer som det beregnes effektivitet for

SØF tar for seg tre sektorer; grunnskole, barnehage og pleie- og omsorgstjenester. Disse sektorene står for brorparten av kostnadene. I skolesektoren benyttes data for de enkelte skoler, og det innføres det en skolebidragsindikator som bygger på karakterer. Kontekstuelle variabler¹ er foreldres utdanning, inntekt og innvandrerbakgrunn.

Det innføres også en miljøindikator som bygger på bl.a. tilfeller av mobbing.

Datagrunnlaget

Rapporten er basert på de data SØF har benyttet i sine analyser og hentes i hovedsak fra KOSTRA og data fra SSB. Ideelt skal data i en effektivitetsanalyse være komplette, dvs. dekke all ressursbruk og all tjenesteproduksjon for en klart avgrenset sektor. Det må heller ikke være overlapp eller dobbelttelling som gjør at samme informasjon opptrer flere ganger. Variable bør dessuten måle de viktige aspektene uten målefeil. Hvilke data som praktisk

¹ Variabler som en kommune ikke kan påvirke.

sett bør inngå i effektivitetsanalyser av kommunal sektor er likevel gjestand for en rekke kompromisser knyttet til tilgjengelighet, datakvalitet, sammenlignbarhet over tid og på tvers av kommuner, samt viktigheten av en eventuell utelatt variabel. Ofte vil også antall variable i en analyse begrenses av antallet observasjoner (frihetsgrader) hvis den skal kunne gi meningsfylt informasjon.

Vi gir i avsnitt 3 en gjennomgang av datagrunnlaget i de ulike sektorene. En rekke variabler er allerede tatt i bruk eller vurdert for SØFs analyser, men vi viser også eksempler på mulige nye variabler i sektorer som allerede analyseres samt andre sektorer hvor det vil være mulig å gjennomføre effektivitetsanalyser. **Vi anbefaler at det gjennomføres årlige effektivitetsanalyser også i flere delsektorer enn skole, pleie og omsorg og barnehager. I sektoravgrensningen er det viktig å ta hensyn til eventuelle samdriftsfordeler og vanskeligheter med å skille data mellom sektorer, f.eks. mellom SFO og skole. Vi anbefaler også at det gjennomføres videre forskning på hvordan en best kan ta hensyn til forskjeller i kostnader mellom by og land og betydningen av interkommunalt samarbeid.**

Innsatsfaktorer

På innsatsfaktorsiden kan det være særlige problemer med måling av kapital på en konsistent måte. Bruk av verdimål heller enn kvantum skjuler problemene med ulik prissetting over tid og på tvers av kommuner; tomtepriser er f.eks. svært ulike i ulike distrikter. Også bruk av lønnskostnader kan skjule hetereogenitet i arbeidskraften; er det høye lønnskostnader fordi en har høyt kvalifiserte og derfor høyt lønnede medarbeidere, eller fordi en bruker mye arbeidskraft?

Hvis en skal aggregere ulike innsatsfaktorer er kostnader det beste målet fordi det veier sammen med priser som skal reflektere den samfunnsøkonomiske verdien av hver enkelt faktortype. Over tid er det likevel riktig å deflatere med gode prisindekser for hver komponent separat. I avsnitt 4.1 har vi sammenliknet resultatene i modeller med ulike kapitalmål og arbeidskraftsmål, og finner rimelig samsvar mellom de ulike modellene.

Vi finner ingen begrunnelse i materialet vi har mottatt for valget å justere av driftskostnader med delkostnadsnøklene i inntektssystemet for kommunene utover at det tar hensyn til bosettingsmønsteret. Det er vanskelig å vurdere kvaliteten i disse korreksjonsfaktorene. Det er rimelig å tro at bosettingsmønsteret vil ha betydning for eller samvariere med kostnadene, men vi drøfter i avsnitt 2.1 om informasjonen om bosettingsmønsteret er utnyttet på den rette måten i effektivitetsanalysene (jf. også neste avsnitt om kontekstuelle variable i tjenesteproduksjonen).

Etter vår vurdering har SØF fremskaffet gode data for de viktigste innsatsfaktorene og i hovedsak behandlet disse på en rimelig måte. Likevel er vi kritiske til å korrigere driftsutgifter med mål for bosettingsmønsteret og anbefaler at en i stedet benytter sentralitetsmål i en samvariasjonsanalyse.

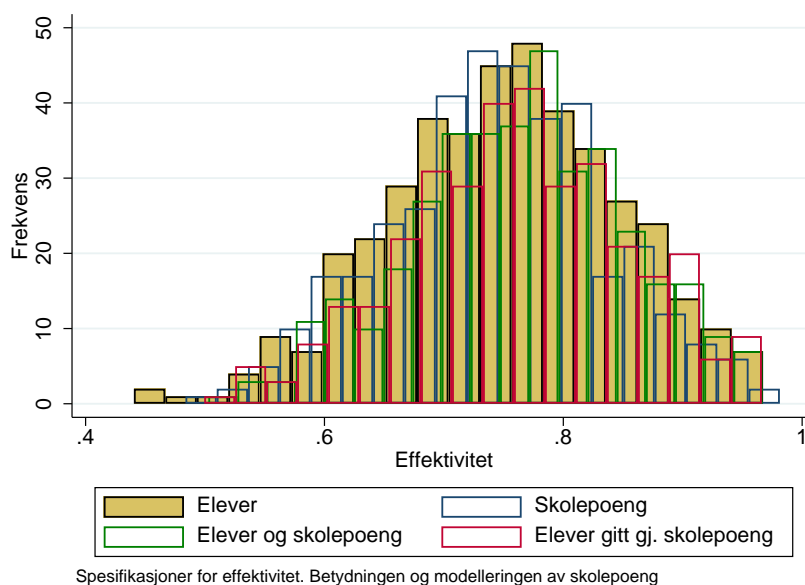
Tjenesteproduksjonen

Måling av tjenesteproduksjonen byr på store problemer. Det finnes gode data på viktige volummål som f.eks. antall elever på hvert trinn i skolen, antall sykehjemspasienter osv. Derimot kan det være store forskjeller i sammensetningen av disse. Det gjelder f.eks. hvor krevende de er å yte tjenester til (vanskelighet) og kvaliteten på tjenestene som ytes. Bruk av mikrodata til å beskrive og kvantifisere slike forskjeller er en lovende utvikling, slik f.eks. skolebidragsindikatoren fanger opp kommunenes bidrag til karakterer for elevene på ulike trinn. Danske myndigheter er her et godt forbilde. **Utvikling av slike indikatorer også på**

andre kommunale sektorer er en forskingsoppgave som TBU og KDD bør oppmuntre til, og ta i bruk i effektivitetsanalysene etterhvert som de foreligger.

Mens det er viktig å kartlegge kvaliteten til tjenestene har bruken av slike indikatorer i selve produksjonsmodellen i DEA-analysen en rekke fallgruver og vi foreslår derfor i stede å ta hensyn til kvalitetsaspektene i analyser av samvariasjonen med volumbaserte effektivitetsmål. Først og fremst er det metodisk ikke rimelig å behandle slike indekser som produkter/outputs i en effektivitetsanalyse; det er f.eks. ikke det samme å øke antall elever og øke det gjennomsnittlige karakternivået. Bruken av delkonstnadsindekser for å korrigere for geografiske og topografiske forskjeller mellom kommunene er av samme type. SØF korrigerer driftskostnadene for slike forskjeller, men i prinsippet gjenspeiler disse forskjeller i vanskeligheten av å yte tjenester heller enn feilkilder i den målte faktorbruken. I tillegg dekker ikke slike kvalitetsindikatorer alle aspekter av sammensetning og kvalitet. Det er f.eks. mange andre aspekter av skolehverdagen enn karakterer og skolemiljø som er viktige for elevenes utvikling. Skal en ta hensyn til mange sammensetnings- og kvalitetsaspekter blir det etter hvert en komplisert og lite gjennomsluktig modell.

I avsnitt 2.1.3 gjennomgås bruken av skolebidragsindikatoren i detalj. I figur S.2 hentet fra den empiriske analysen i avsnitt 4.1 har vi illustrert effektivitetsscorene for grunnskolen estimert ved DEA ved tre ulike håndteringer av kvalitetsaspektet sett opp mot en spesifisering der kun elever per hovedtrinn er benyttet som produkter.



Figur S.2: Frekvensfordelinger av effektivitetsscorer i grunnskolen over norske kommuner i 2019 beregnet ved innsatsfaktororientert og bootstrappet DEA og variabelt skalautbytte for 2018 og 2019.

Faktorinnsats: Bruttodriftsutgifter mot bruttodriftsutgifter fratrukket arbeidsgivergift og avskrivninger

Produkter: a) og d) Elever på hvert hovedtrinn, b) samlede skolepoeng på hvert hovedtrinn og

c) Elever og samlede skolepoeng på hvert hovedtrinn

Kontekstuelle variabler: a) til c) Ingen, d) gjennomsnittlig skolepoeng på hvert hovedtrinn

Figuren gir ikke indikasjon på store avvik i resultatene for ulike måter å behandle skolebidrag/skolepoeng. Generelt vil mindre avvik i kvalitet også gi små avvik i effektivitetsmål i de ulike metodene. Derimot vil en analyse basert på rene volumtall i produksjonsmodellen gi mere gjennomsluktige og metodisk konsistente resultater, samtidig som koeffisientene i den estimerte ettermodellen vil gi informasjon om betydningen av de ulike kontekstuelle variablene for ressursbruk og effektivitet. **Vi anbefaler at variabler som beskriver sammensetning og kvalitet i stedet blir behandlet som kontekstuelle variabler, som**

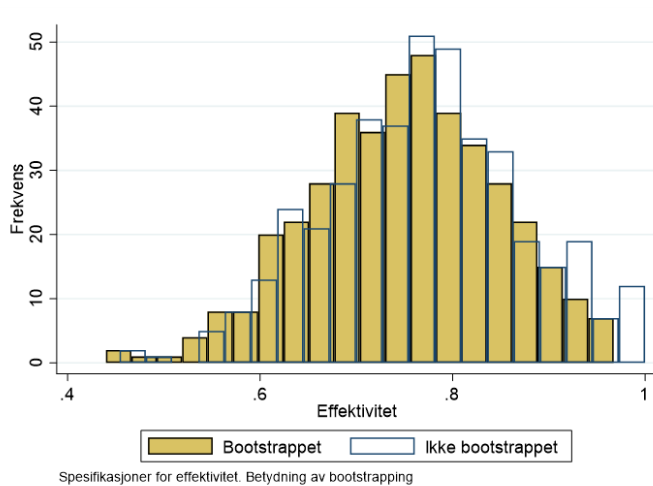
i en DEA-analyse tilsier at det legges i en trinn-to analyse som viser samvariasjonen mellom disse kjennetegnene og effektiviteten basert på en ren volum-modell.

I kapittel 2.1.2 gir vi også en oversikt over metodeutvidelser av DEA-metoden som tillater å inkludere variabler som ikke er på volummål. TBU kan vurdere uttesting av slike modeller som del av det videre arbeidet med effektivitetsanalysene.

Bootstrapping

Effektivitet og produktivitet målt ved DEA-metoden vil være lett å beregne uansett antall variabler og antall observasjoner, men estimatene vil være mer usikre desto færre observasjoner som foreligger. Dette problemet kan avhjelpest ved å bruke *bootstrapping* til å tallfeste usikkerheten i form av konfidensintervall eller standardfeil (Simar og Wilson, 1998). I stolpediagrammet figur S.3 ser vi betydningen av effektivitetsfordelingen med og uten bootstrapping. Bootstrapping vil generelt gi en lavere effektivitet. Vi ser at dette gjelder spesielt for de høyeste effektivitetstall.

Boostrapping hindrer at resultater overfortolkes gjennom å tro at estimatene er mere presise enn de er. Også rangeringen av kommuner kan være usikker og bruk av bootstrap gir mulighet for å vise i hvilken grad en kommune er statistisk signifikant mer effektiv enn en annen. **Vi anbefaler at en fortsatt bruker bootstrapping til å beskrive usikkerheten til effektivitetsestimaterne i DEA-metoden.**



Figur S.3: Frekvensfordelinger av effektivitetsscorer i grunnskolen over norske kommuner i 2019 beregnet ved innsatsfaktororientert og bootstrappet og ikke bootstrappet DEA med variabelt skalautbytte for 2018 og 2019. Faktorinnsats: Bruttodriftsutgifter fratrukket arbeidsgivergift og avskrivninger. Produkter: Elever på hvert hovedtrinn. Kontekstuelle variabler: Ingen.

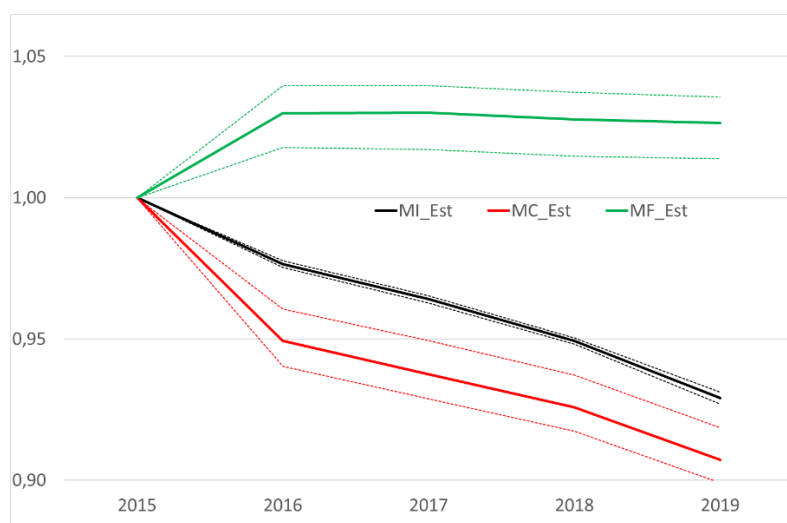
Malmquist-indeksen for produktivitetsendringer

SØF bruker det som kan kalles en vindu-analyse for å beskrive effektivitetsutviklingen over tid. Denne er basert på utvikling av effektivitetstall fra et år til det neste i hvert par av år. Referansteknologien endres med overlapp av årene hver annen periode. I SØFs opplegg vet en ikke om en endring over et lengre tidsrom skyldes endring i variabelspesifikasjonen, i dimensjonaliteten til modellen, i bruk av bootstrap for å korrigere effektivitetsestimaterne, eller i faktisk endring av kommunenes effektivitet mellom tidspunktene. Dermed er ikke disse effektivitetsmål sammenliknbare over tid og metoden er ikke standard i litteraturen.

Innen effektivitetsanalysen er det vanlig å benytte den såkalte Malmquistindeksen (Malmquist, 1953; Caves mfl. 1982) til å måle produktivitet utviklingen og så dekomponere denne til ett mål for *frontendring*, ett for *effektivitetsutviklingen* og ett mål for *endring i skalaeffektivitet*. Fordelene med Malmquistanalyse drøftes nærmere i avsnitt 2.2 mens avsnitt 4.2 viser hvordan dette kan slå ut empirisk for skolesektoren.

Som en ser av figur S.1 må referansen for produktivitet ha konstant skalautbytte (være en rett linje). Over tid vil både fronten og referansen kunne endre seg. Ved beregning av produktivitet *endring* er det derfor også nødvendig å bruke en fast referanse, slik f.eks. prisindekser bruker faste vekter. Den faste referansen kan f.eks. være fronten i første periode, fronten i siste periode, eller en referansefront som utnytter observasjonene i alle år.

En illustrasjon av resultater for Malmquist-indeksen og dets hovedkomponenter vises i figur S.4 hvor vi har beregnet indeksen med en referansefront basert på observasjoner i alle år:



Figur S.4: Estimert Malmquistindeks MI og dets komponenter frontskift MF og opphenting MC; $MI=MC \times MF$. Estimer og 95% konfidensintervall for snittet over kommuner ved bruk av bootstrapping.

«Bootstrapping»² brukes for å beregne konfidensintervall for tallene for indeksen. Vi ser at Malmquist-indeksen MI synker med omtrent 7%, men at konfidensintervallet er ganske smalt. Multiplikativ dekomponering av indeksen i to deler viser at frontindeks MF og opphentingsindeksen MC begge har vesentlig bredere konfidensintervall enn MI indeksen. Dette viser usikkerheten ved estimatene

Malmquistindeksen gir resultater som er lett tolkbare og basert på samme definisjoner over tid. Indeksen og dekomponeringen gir også informasjon om produktivitet utviklingen, i tillegg til effektivitetsutviklingen. Produktivitet er et mål på hvor mye samfunnet får ut av ressursene. Effektivitet er kun et mål på tjenesteproduksjon i forhold til ressursbruk ut fra mulighetene på hvert tidspunkt og ved sin størrelse.

Utfordringen ved Malmquistindeksen er kun at en må ha en enkel modellspesifikasjon hvor det er mulig å skaffe data tilbake i tid og må gjennomføre en større modellkjøring over alle relevante år på nytt. En enkel modell er likevel enklere å forstå. En enkel modell tilbake i

² Bootstrapping brukes av SØF for første gang i TBU rapporten Borge et al. (2020). Meget kort sagt er metoden å analysere sensitiviteten til estimerte effektivitetstall ved å trekke nye sampel for enhetene mange ganger, f.eks. 2000, og så utføre egne effektivitetsberegninger på hvert enkelt datasett. Konfidensintervall på avviket mellom det originale effektivitetstallet til en enhet og det estimerte ved observasjonene som er trukket kan da lages.

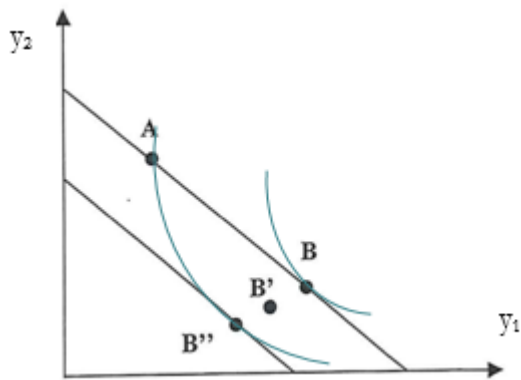
tid forhindrer dessuten ikke at rikere modellspesifikasjoner kan brukes i analysen for det siste året eller de siste årene. **Vi anbefaler at TBU tar i bruk Malmquistindeksen for analyse av effektivitetssutviklingen over tid basert på en enkel modellspesifikasjon.**

Produksjonsindeksen og uavhengighet av prioritering

Produksjonsindeksen (Borge m.fl., 2001) er ment som et mål på kommunenes samlede tjenestetilbud. Den er hovedsakelig utviklet med sikte på å måle produksjon og ikke effektivitet, men TBU har bedt om en vurdering av bruk av indeksen til effektivitetsmåling.

Avsnitt 2.3 beskriver indeksens oppbygging og egenskaper.

Overordnet kan indeksen beskrives som bestående av et sett av kriterier med tilhørende vekt, hvor vektene i hovedsak er budsjettandeler. Det beregnes delindekser for barnehage, grunnskole, primærhelsetjeneste, pleie og omsorg, barnevern, sosialkontortjenester og kultur, samt den overordnede produksjonsindeksen som omfatter alle disse tjenestene.

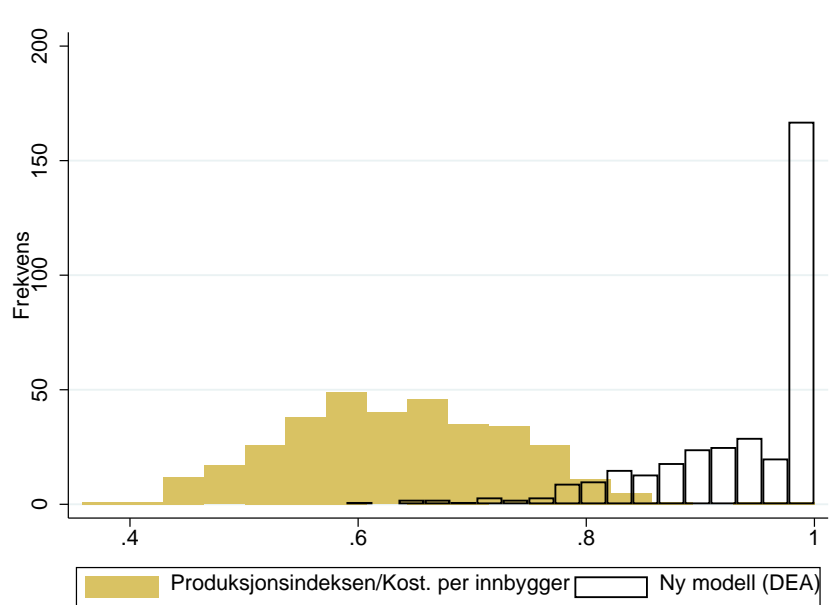


Figur S.5: Nytteoptimering under budsjettbetingelse.

Figur S.5 viser budsjettlinjer for kostnader for kommunene A, B, B' og B''. SØF innfører en forutsetning om at punkter på den samme kostnadslinje har den samme produksjonen. Kommune A har altså den samme produksjonen som kommune B. Men hvis de to kommunene prioriterer tjenestene forskjellig, f.eks. hvis aldersfordelingen av innbyggerne er forskjellig, så gir det ikke god mening at produksjonen er lik. De to blå kurvene («indifferenskurver») representerer nivåer på preferansene kommunene har. Med de samme preferanser med nytte som øker utover i diagrammet, så har A en klart lavere nytte enn B for samme kostnad. En optimering av nyten når preferansene tas i betraktning er ikke forenlig med samme kostnad. Vi innser at utledning av alternative vekt basert på en sosial nyttefunksjon kan vise seg både vanskelig og svært ressurskrevende. Vi vil i stedet anbefale metodikk til effektivitetsmåling som tillater at vi ikke må benytte priser eller kostnader i beregningen av effektivitet i produksjonsindekssammenheng.

Den enkleste bruken av produksjonsindeksen til å måle produktivitet er å dele produksjonsindeksen på korrigerte brutto driftskostnader pr innbygger. For å normalisere dette som et effektivitetsmål må en velge enheten med høyest produktivitet som benchmark for andre enheter, og beregne effektivitet som forholdet mellom hver kommunes produktivitet og den høyeste produktiviteten målt i utvalget. Dette kalles Corrected ordinary Least Squares (COLS) i effektivitetslitteraturen. I avsnitt 2.3.2 beskrives et alternativt opplegg til måling av effektivitet ved bruk av DEA-metoden på produksjonsindeksens delindekser. Forskjellen mellom SØFs produksjonsindeks og en DEA modell når det gjelder effektivitet vises i figur S.6. DEA modellen har konstant skalautbytte og vekt for produkter som

bestemmes av modellen og ikke som produksjonsindeksen som bruker gitte budsjett-andeler. Begge modeller har kostnadene per innbygger som eneste innsatsfaktor:



Figur 5.6: En sammenlikning mellom kommuneeffektivitet målt ved DEA og SØFs produksjonsindeks som er estimert med Corrected Least Square-estimatoren (COLS).

Vi ser at COLS-metoden gir effektivitetsscore som i større grad er normalfordelte – noe som kan sees i sammenheng med at produksjonsindeksen er innrettet mot å måle avvik fra det gjennomsnittlige produksjonsnivået.

En DEA-analyse basert på produksjonsindeksen har også klare svakheter ved at den i stor grad er basert på variabler med vidt forskjellig målestokk. Det er ordinale data (f.eks. skolemiljø), korrigerede negative data (f.eks. skolebidrag), og forholdstall. Vi har tidligere drøftet at standard DEA ikke er tilpasset denne typen data, men at det kreves justeringer av DEA-modellen. **En anbefaling, i alle fall i forbindelse med DEA, vil være å i stor grad benytte tradisjonelle volummål i definisjonen av frontmodellen, og heller å korrigere for mer komplekse variabler i en ettermodell. Vi viser til kapittel 2.1.2 for ytterligere detaljer rundt dette.**

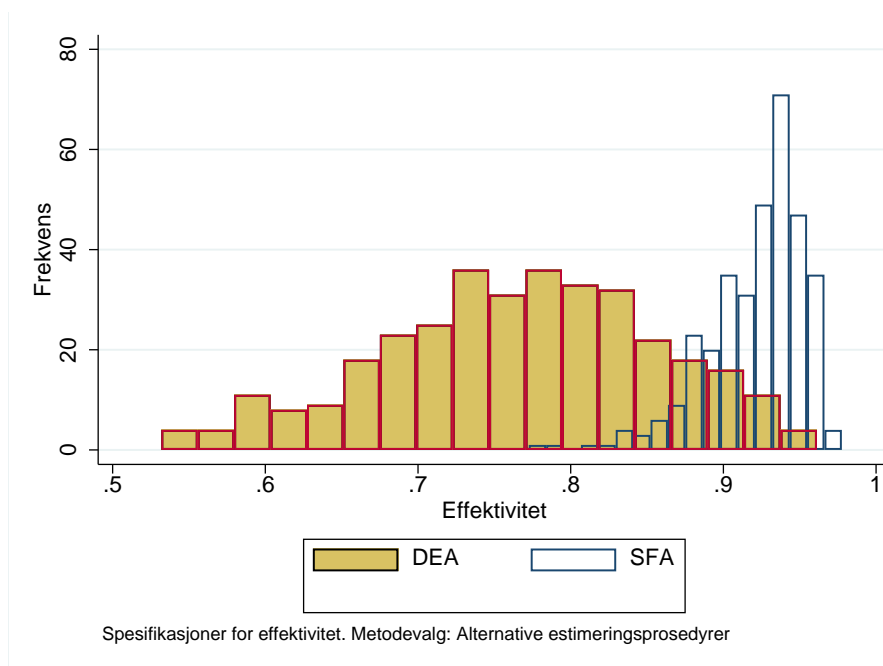
Benchmarking

Benchmarking brukes gjerne som betegnelse på nøkkelindikatorer for tjenester. Slike mål er partielle størrelser. For en kommune kan det være driftsutgifter til barnehager per innbygger, elever per lærer, pleietrengende per pleier og liknende. Men slike partielle mål kan være misvisende når det gjelder å måle samlet effektivitet for en kommune som har flere tjenester og innsatsfaktorer. Metoden tar heller ikke inn over seg avveininger i ressursbruk (f.eks. avveininger mellom ulike produkter i tjenesteproduksjonen), noe SØFs metodikk er i stand til. Benchmarking basert på partielle indikatorer kan heller ikke ta hensyn til utnyttelse av skalaøkonomi og samdriftsfordeler eller ulemper, noe DEA-metoden er velegnet til. Veies de ulike partielle forholdstallene sammen vil det ha samme egenskaper som en effektivitetsindeks basert på produksjonsindeksen som beskrevet i forrige avsnitt. **Mens partielle forholdstall kan gi viktig sektorinformasjon anbefaler vi ikke bruk av disse i en effektivitetsanalyse.**

Sammenlikning av frontmetoder

I tillegg til standard DEA modell er to andre modeller i bruk som er basert på stokastikk, SFA (Stochastic Frontier Analysis) og StoNED modellen (Stochastic Nonparametric Envelopment of Data). StoNED er en videreutvikling av DEA som gjør det mulig å ta simultant hensyn til kontekstuelle variabler. En standard DEA modell som SØF bruker har ikke stokastiske variabler, dvs. variabler med målefeil eller tilfeldige utslag. Dette er en svakhet da stokastisk variasjon i data som oftest vil være relevant for estimering av effektivitet. En modell som er mye brukt er SFA. Men en slik modell krever at man må formulere en parametrisert (matematisk) frontfunksjon. En funksjon som er brukbar kan være vanskelig å konstruere spesielt for tjenesteytende sektorer som kommuner.

Metodene er nærmere forklart i i avsnitt 2.5. En sammenlikning av effektivitet når DEA og SFA brukes til estimering av grunnskole-effektivitet med SØFs variabelspesifikasjon vises i figur S.7:



Figur S.7: Frekvensfordelinger av effektivitetsscorer i grunnskolen over norske kommuner i 2019 beregnet ved innsatsfaktororientert og a) bootstrappet DEA og b) SFA og 2018 og 2019 med variabelt skalautbytte. Faktorinnsats: Bruttodriftsutgifter fratrukket arbeidsgivergift og avskrivninger. Produkter: Elever og samlede skolepoeng på hvert hovedtrinn og skolemiljø. Kontekstuelle variabler: Ingen

SFA gir her en klart sterkere konsentrasjon av kommuner med høyere effektivitet enn DEA. SFA tolker derfor en betydelig del av variasjonen mellom kommunene som tilfeldigheter og målefeil heller enn som forskjeller i effektivitet. Et problem med SFA er imidlertid at det på grunn av de sterke forutsetningene ikke alltid er mulig å finne en løsning. DEA-modellen er velkjent i litteraturen, lett å beregne, og konseptuelt enkel å forholde seg til. DEA-metoden vil også ha mere presis informasjon om effektiviteten og potensialet for forbedring for den enkelte kommunen. **Vi anbefaler fortsatt bruk av DEA (eller dets nære utvidelse StoNED) som hovedmodell i analysene, men ser gjerne at en også bruker SFA til robusthetsanalyser, spesielt fokusert på om rangeringen av kommuner er ulik mellom metodene.**

Paneldataanalyse

Det er fokusert på metodikk for estimering av effektivitet basert på tverrsnittsdata. SFA og StoNED (og varianter av DEA) kan derimot utvides til å ta hensyn til at kommunedata har en paneldatastruktur, dvs. at den samme kommunen observeres gjentatte ganger over tid.

En viktig fordel med paneldata over tverrsnittsdata er at det er mulig å identifisere uobserverbar heterogenitet; dvs. kjennetegn ved kommunene som påvirker deres produksjon men som ikke enkelt kan kvantifiseres ved hjelp av relevante variabler i modellspesifikasjonen. Dette er blitt utnyttet i litteraturen om effektivitetsmåling, hvor det er blitt utviklet modeller som skiller mellom kommuneheterogenitet (som ikke er en del av effektivitetsleddet, men som heller tilskrives variasjon i uforklarte kontekstuelle variabler) og effektivitet. Dette betinger at en er i stand til å skaffe data for de samme variablene for noen år tilbake i tid, noe som igjen antakelig innebærer å velge en enkel variabelspesifikasjon. **Vi anbefaler at paneldataanalyser vurderes nærmere framover.**

Konklusjoner

SØF har opp gjennom årene levert et solid arbeid i sine analyser av effektivitet i kommunesektoren, i all hovedsak basert på det som har vært standard innenfor effektivitetslitteraturen. Likevel har det i analysene vært foretatt tilpasninger som fremstår som noe mer ad hoc og derfor mindre gjennomskubare og etterprøvbare. SØFs analyser avviker fra standard metodikk innen effektivitetsanalyser på en del punkter.

Dette gjelder bl.a. behandling av kontekstuelle variabler og endring over tid. Vi anbefaler mer bruk av standardmetoder. DEA er primært egnet for volummål, men SØF blander data med ulike målestokker. Dette fører til at effektivitetsberegninger ikke blir korrekte og vanskelige å tolke. Behandling av kvalitet og sammensetning er et vanskelig punkt. Mange aktuelle variabler har spesielle måleenheter som ikke egner seg for DEA som er basert på volumdata og ikke data basert på forholdstall eller diskrete kategorier. Også korreksjoner av driftskostnadene kan spille inn på effektivitetsmålingen. Vi anbefaler bruk av disse som kontekstuelle variabler i en tilleggs- eller ettermodell som kortsiktig løsning, men anbefaler videre forskning på temaet.

Vi finner empiriske forskjeller mellom teknisk effektivitet og produktivitet. Det viser at SØF kan ha nytte av å utvide modellopplegget til å dekomponere teknisk effektivitet og skalaeffektivitet. Det sistnevnte kan gi nyttig informasjon om hvordan kommunestørrelse påvirker produktiviteten.

Spesielt for analyser av effektivitet over tid ser vi ikke SØFs vinduanalyse som hensiktsmessig da effektivitetstallene ikke er sammenliknbare når modellen skifter over tid. Våre beregninger viser at dette har til dels betydelige innvirkninger på resultatene. Kjeding av effektivitetsscore viser da en sammensatt effekt av modell- og effektivitetsendringer. Malmquist-indeksen for produktivitetsendring over diskret tid er standarden i litteraturen og vi anbefaler at denne tas i bruk, da med en enkel variabelspesifikasjon som lar seg bruke for noen år tilbake. Malmquist-indeksen gir også informasjon om produktivitetsutviklingen i tillegg til effektivitetsutviklingen.

SØFs opplegg for å måle effektivitet basert på produksjonsindeksen gir ikke et korrekt mål på effektivitet fordi opplegget gir en indeks for produktivitet, ikke effektivitet, slik begrepene vanligvis brukes. Man trenger en referanseproduktivitet for å beregne effektivitet. Produksjonsindeksen har også flere uønskede egenskaper, som bl.a. konstant bytteforhold mellom produkter. Vi foreslår en mulig alternativ produksjonsindeks som tilfredsstillende vanlige krav til effektivitetsmåling, men vil ikke anbefale dette som hovedtilnærming nå. Det virker hensiktsmessig å ta en grundig gjennomgang av variablene i

produksjonsindeksen og i hvilken grad dagens opplegg er vellykket i å konvertere dem til en felles målestokk. En anbefaling i forbindelse med DEA-metoden vil være å i stor grad benytte tradisjonelle volummål i definisjonen av modellen for produktfunksjonen, og heller korrigere for mer komplekse variabler i en ettermodell.

DEA-metoden er enkel å forstå og solid etablert som en hovedmetode i effektivitetsmåling internasjonalt. Vi anbefaler at denne fortsatt benyttes som hovedmetode, og at usikkerheten fortsatt beskrives ved bruk av bootstrapping. Benchmarking i betydningen partielle forholdstall er ikke spesielt egnet til effektivitetsanalyse. En utvidelse av metodeopplegget er å anvende stokastiske metoder (SFA; StoNED) og/eller paneldatametodikk til å beriket analysene og belyse robustheten til resultatene. Våre beregninger viser at resultatene fra stokastiske modeller avviker sterkt fra den deterministiske DEA-modellen. Paneldata kan gi ny innsikt, bl.a. ved å gjøre det mulig å skille mellom effektivitet og andre forskjeller mellom kommunene som har betydning for ressursbruken.

Løpende statistikk og analyser av effektivitets- og produktivitetsutviklingen i kommunesektoren har stor samfunnsmessig verdi. Vi anbefaler at arbeidet fortsetter, men med større vekt på standard metoder.

Vi har gjennomgått en del kilder til data for ressursbruk og tjenesteproduksjon i kommunesektoren, og dette bør kunne gi grunnlag for rikere modeller i flere delsektorer. Likevel, enkle modeller er oftest mest gjennomsluttede og forståelige.

Etterprøvnbarhet og reproducerbarhet er viktig for resultatenes legitimitet, i tillegg til å gjøre det enklere med evalueringer som denne. Det bør derfor etableres ordninger for deponering eller offentliggjøring (f.eks. på web) av datagrunnlaget både i rå form og i bearbejdede form slik at det kan for fremtiden er enklere å gjøre reproduksjonsstudier eller alternative antakelser og metoder.

I tillegg er det ønskelig å stimulere til videre forskning, f.eks. på hvordan kvalitetsaspekter kan bli tatt hensyn til, særlig i ikke-parametriske eller semi-parametriske metoder. Det er dessuten ønskelig å stimulere forskning om hvordan en skal behandle interkommunalt samarbeid, betydningen av bosettingsmønster og ikke minst utvikling av flere kvalitetsmål basert på mikrodata.