

Bilag 7

SFA-modellen



Bilag 7

Konkurrence- og Forbrugerstyrelsen Forsyningssekretariatet

Carl Jacobsens Vej 35
2500 Valby
Tlf.: +45 41 71 50 00
E-mail: kfst@kfst.dk

Online ISBN 978-87-7029-650-2

Bilag 7 er udarbejdet af
Forsyningssekretariatet.

November 2016

Indhold

Kapitel 1	
Indledning.....	4
Kapitel 2	
Parametrisk metode	5
Kapitel 3	
Estimering i SFA.....	6
3.1 Estimering af fronten.....	6
3.2 Estimering af selskabernes effektivitet.....	7

Kapitel 1

Indledning

Stochastic Frontier Analysis (SFA) er en udbredt benchmarkingmodel, der modsat DEA-modellen, er i stand til at skelne mellem ineffektivitet og generel støj i data. SFA er en parametriske metode, der kan anvende forskellige typer af regressioner til at sammenligne en række selskaber med hinanden og bestemme effektiviteten af de forskellige selskaber.

Selskaberne sammenlignes, som i DEA, på baggrund af et sæt af inputs og outputs. Inputs er selskabernes faktiske omkostninger, og outputs er et udtryk for det, selskaberne producerer.

I SFA er alle repræsentative selskaber med til at bestemme en front, som hver enkelt forsyning efterfølgende kan sammenlignes med. Ved repræsentativ menes, at selskabet skal have en struktur, der er sammenlignelig med størstedelen af branchen. Jo længere fra fronten et selskab ligger, jo mere ineffektiv er selskabet.

SFA-modellen kan grundlæggende antage to former. Modellen kan være inputstyret eller outputstyret. Den inputstyrede model danner en regression, hvor input bliver forklaret ud fra en række outputs. Den outputstyrede model danner en regression, hvor output bliver forklaret ud fra en række inputs. I benchmarking af vandsektoren anvendes en inputstyret model, da det antages, at selskaberne ikke har mulighed for at øge deres output, men at de derimod kan mindske deres input, det vil sige reducere deres omkostninger, for at opnå en højere effektivitet.

I det følgende beskrives først, hvad det betyder, at SFA er en parametriske model, derefter hvordan modellens parametre estimeres, og til sidst hvordan effektiviteten for det enkelte selskab findes. Der er taget udgangspunkt i P. Bogetoft and L. Otto, *Benchmarking with DEA, SFA, and R*, International Series in Operations Research & Management Science 157 (2011).

Kapitel 2

Parametrisk metode

SFA er en variation af klassiske statistiske metoder og er dermed parametrisk. En parametrisk metode adskiller sig fra en ikke-parametrisk metode, som for eksempel DEA, ved at antage, at data har en bestemt sandsynlighedsfunktion, og at der er et bestemt antal parametre. Når den parametriske model udvikles, skal en række antagelser være opfyldt. Dette gør, at en sådan model ofte kan være svær at opstille. Til gengæld er den mere præcis end en ikke-parametrisk model, når antagelserne er korrekte. Hvis der kommer ny information til modellen, vil man skulle begynde forfra, da antallet af parametre ændrer sig, hvilket netop går imod en af antagelserne. Dette medfører, at ny information har risiko for at ændre på modellen.

En af antagelserne i den parametriske metode SFA er, at fronten dannes på baggrund af en produktionsfunktion. Denne produktionsfunktion kan have forskellige former, og det er dermed vigtigt, at den rigtige produktionsfunktion vælges. En ofte anvendt produktionsfunktion er Cobb-Douglas funktionen:

$$y = \beta_0 x_1^{\beta_1} x_2^{\beta_2} \dots x_m^{\beta_m},$$

hvor $\beta_0, \dots, \beta_m$ er ukendte parametre. y kaldes den afhængige variable, og $x_1, \dots, x_m$ er de uafhængige eller forklarende variable. Ved hjælp af det observerede data kan de ukendte parametre estimeres, og funktionen kan derefter anvendes til at danne en front i SFA.

Den mest anvendte metode til at estimere parametre i en parametrisk model er maximum likelihood. Ved hjælp af denne metode findes værdien af $\hat{\beta}$, så sandsynligheden for at observere det observerede data bliver maksimeret.

Hvis et selskabs observerede data varierer fra modellens forventede værdi for selskabet, siges det, at data afviger fra modellen. Der kan være tre årsager til, at det observerede data kan afvige fra modellen. Det kan skyldes støj i modellen, altså at modellen ikke er præcis og dermed ikke fuldt forklarende. Dette er en ordinær regression. En anden mulighed er, at afvigelsen kan forklares ved, at selskabet er ineffektivt. Dette minder om DEA-tilgangen og kaldes for en deterministisk metode. Den sidste årsag er en blanding af de to forrige. Nemlig at afvigelser kan skyldes både støj i modellen og, at det enkelte selskab er ineffektivt. Dette er den tilgang, SFA anvender. Matematisk kan dette skrives som følgende:

$$y = f(x; \beta) \exp(v) \exp(-u) \quad (1),$$

hvor $f(x; \beta)$ er produktionsfunktionen, som for eksempel kunne være Cobb-Douglas funktionen, v er støj i modellen og $u \geq 0$ er ineffektivitet. Støjen kan altså både øge og mindske den afhængige variabel, hvor ineffektiviteten kun kan mindske den. Intuitivt betyder dette, at støjen kan give enten en højere eller lavere værdi til det enkelte selskab, hvorimod en ineffektivitet altid vil give en lavere værdi.

Kapitel 3

Estimering i SFA

3.1 Estimering af fronten

Fronten i SFA dannes ved hjælp af en produktionsfunktion samt støj og ineffektivitet jf. ligning (1). Der er dermed en række ubekendte variable, som skal estimeres, før fronten kan fastsættes. I ligning (1) er det kun x , der er kendte, da disse er de observerede værdier, og de resterende variable er ubekendte. Det vil sige, at vi skal estimere β , v og u for at finde fronten. For at kunne estimere disse parametre er der brug for en række antagelser om fordelingerne på v og u .

De mest udbredte antagelser om støjen og ineffektiviteten i SFA er:

$$v^k \sim N(0, \sigma_v^2), \quad u^k \sim N_+(0, \sigma_u^2), \quad k = 1, \dots, K$$

Dette betyder, at støjen for selskab k er normalfordelt med middelværdi 0 samt variansen σ_v^2 og, at ineffektiviteten for selskab k er halvnormalfordelt over intervallet $[0, \infty]$ med variansen σ_u^2 .

Disse to antagelser er ikke tilstrækkelige til at estimere de to parametre. Det er dog tilstrækkeligt til at estimere den samlede afstand til fronten, altså $\epsilon = v - u$. Vi ønsker nu at finde fordelingen for ϵ .

Antag at variansen for støjen er større end variansen på ineffektiviteten, altså at $\sigma_v^2 > \sigma_u^2$. Da vil variansen på ϵ påvirkes mere af støjen end af ineffektiviteterne. Fordelingen af den samlede afstand til fronten vil derfor nærmere ligne fordelingen af støjen end fordelingen af ineffektiviteten. Dette kan udnyttes til at estimere fordeling af ϵ .

Sammenhængen mellem de tre fordelinger kan se forskellig ud, jf. figur 3.1. λ angiver forskellen på de to varianser, og σ^2 er variansen på den samlede afstand til fronten, ϵ . Matematisk kan vi derfor skrive, at $\lambda = \sqrt{\frac{\sigma_u^2}{\sigma_v^2}}$ og $\sigma^2 = \sigma_v^2 + \sigma_u^2$. Det ses i figuren, at jo større λ er, desto mere ligner fordelingen af de samlede fejl, ϵ , fordelingen for ineffektiviteten, u .

Figur 3.1 Fordelingen på likelihood funktionen afhængig af σ_v^2 og σ_u^2

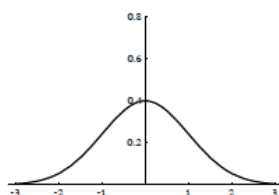


Fig. 7.5 $\sigma_v^2 = 1, \sigma_u^2 = 0, \sigma^2 = 1, \lambda = 0$

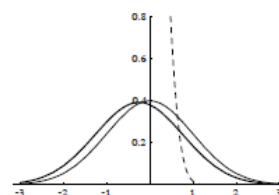


Fig. 7.5a $\sigma_v^2 = 1, \sigma_u^2 = .1, \sigma^2 = 1.1, \lambda = .3$

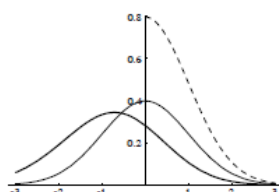


Fig. 7.5b $\sigma_v^2 = 1, \sigma_u^2 = 1, \sigma^2 = 2, \lambda = 1$

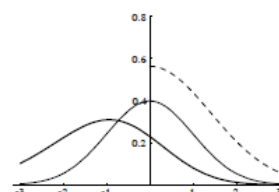


Fig. 7.5c $\sigma_v^2 = 1, \sigma_u^2 = 2, \sigma^2 = 3, \lambda = 1.4$

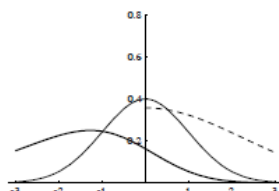


Fig. 7.5d $\sigma_v^2 = 1, \sigma_u^2 = 5, \sigma^2 = 6, \lambda = 2.2$

Lines in the figures:
 — Normal error, v
 - - - Efficiency error, u
 — Total error, $\epsilon = v - u$

Kilde: P. Bogetoft and L. Otto(2011)

For at estimere σ^2 og λ kunne man for eksempel estimere en model ud fra en ordinær regression, udregne residualerne, svarende til ϵ , og plote deres fordeling. Herefter kunne man vælge den af ovenstående figurer, som passer bedst til plottet. Dette er selvfølgelig ikke en optimal løsning, men er grundlæggende idéen bag, hvordan maximum likelihood estimerer σ^2 og λ i SFA.¹

3.2 Estimering af selskabernes effektivitet

Selskaberne får udregnet deres effektivitet ved hjælp af estimatorne af σ^2 og λ , der blev brugt i ovenstående afsnit til at finde fronten. Selskabernes effektivitet er givet ved:

$$E_k = \exp(-\hat{u}_k),$$

hvor $\hat{u}_k$ er estimeret af u for selskab k .

¹ For en nærmere beskrivelse af maximum likelihood princippet se P. Bogetoft and L. Otto, *Benchmarking with DEA, SFA, and R*, International Series in Operations Research & Management Science 157 (2011).

Selvom vi kender $\epsilon_k = v_k - u_k$ er det ikke simpelt at estimere u_k . Problemet er, at der kun er én ligning med to ubekendte. ϵ_k giver dog stadig brugbar information om u_k .

En af de antagelser SFA laver er, at middelværdien af støjen er lig med 0, $E[v] = 0$. Dette medfører, at det bedste gæt på støjens værdi er 0. Forventningen til den samlede afstand til fronten er dermed $E[\epsilon_k] = E[v] - u_k = 0 - u_k = -u_k$. Da dette kun er en forventning af ϵ_k , kan vi ikke sige, at denne sammenhæng altid er gældende. Til gengæld kan vi antage, at den med stor sandsynlighed gælder. Der er dermed en direkte negativ sammenhæng mellem ϵ_k og u_k . Hvis ϵ_k er lav, må u_k med stor sandsynlighed være høj og modsat. Dette kan udnyttes til at udregne middelværdi af u_k betinget af ϵ_k , $EV(u_k|\epsilon_k)$.

Da vi fra forrige afsnit kender fordelingen af ϵ_k , kan vi udregne den betingede middelværdi ved hjælp af Bayes' formel, hvilken vi ikke vil gennemgå her.² Resultatet af Bayes' formel giver ineffektiviteten for de enkelte selskaber.

² For en nærmere beskrivelse af Bayes' formel se P. Bogetoft and L. Otto, *Benchmarking with DEA, SFA, and R*, International Series in Operations Research & Management Science 157 (2011).
