

## **NOTAT**

Dato: 16. maj 2024

Sag: VAND-24/01567

Sagsbehandler: KIPH, EHE og  
MAFR

## **Uddybning af udfordringer i SFA og MOLS**

Forsyningssekretariatet har i februar 2024 sendt ”Notat om ændring af metoden for økonomisk benchmarking” i teknisk høring. I høringmaterialet argumenterer Forsyningssekretariatet for, at Order-M er den bedste benchmarkingmodel til regulering af vandselskaberne. Vurderingen bygger på analyser, som Forsyningssekretariatet har foretaget af SFA og alternativer til SFA, hvor alternativer sammenlignes i forhold til en række evalueringskriterier. I notatet skriver Forsyningssekretariatet, at baggrunden for undersøgelserne er, at SFA ikke er robust og i praksis derfor ikke har de fordele, som Forsyningssekretariatet forventede, da SFA blev indført i reguleringen til de økonomiske rammer for 2017 og frem.

**KONKURRENCE- OG  
FORBRUGERSTYRELSEN**

**ERHVERVS MINISTERIET**

DANVA og TREFOR afgav høringssvar til den tekniske høring, og begge parter gav udtryk for, at Forsyningssekretariatet bør fortsætte med at bruge *best-of-two*. Endvidere angav DANVA og TREFOR, at der bør fortsættes med en af de stokastiske modeller, enten SFA eller MOLS, fordi disse modeller tager hensyn til støj.

I dette notat uddyber Forsyningssekretariatet problemerne med SFA, og hvorfor den ikke er robust til regulering af vandselskaber. Forsyningssekretariatet viser videre, at der med nærmeste alternativ til SFA; MOLS, opstår samme udfordringer som med SFA. Udfordringerne indebærer, at der er stor usikkerhed om validiteten af resultaterne. Det medfører, at modellerne ikke altid kan beregne selskabernes efficiensscorer, fordi modellernes underliggende antagelser ikke er opfyldte.

Forsyningssekretariatet har foretaget en empirisk robusthedsanalyse på data, som vandselskaberne har indberettet til brug for fastsættelse af de økonomiske rammer for 2023 for drikkevandsselskaber. I analysen simuleres og tilføjes støj i data for at undersøge, om SFA og MOLS reelt håndterer datastøj, som det af teorien bag modellerne fremgår, at modellerne skal. På grundlag af resultaterne af denne analyse konkluderer Forsyningssekretariatet, at hverken SFA eller MOLS er robuste i forhold til at håndtere små tilføjelser af støj til datasættet. Konklusionen baseres på, at mange iterationer giver store udsving i resultaterne. Hertil viser den empiriske analyse også, at der hyppigt opstår iterationer med fejl i de underliggende modelantagelser. Fejlene skyldes enten, at fordelingen af støj og inefficiens ikke kan adskilles, eller at forholdet mellem mængden af

støj og inefficiens er urealistisk. Herudover oplever Forsyningssekretariatet fejl i modelspecifikationerne, hvor centrale modelantagelser ikke er opfyldte, eller hvor der er høj grad af korrelation mellem variable, der skal være uafhængige.

Forsyningssekretariatet har ved anvendelse af SFA løbende oplevet udfordringer med, at små ændringer til data har ført til, at forskellige grundlæggende antagelser i SFA blev brudt, jf. robusthedsanalysen nedenfor. Forsyningssekretariatet vurderer, at der med stor sandsynlighed vil komme et år, hvor SFA ikke kan beregne efficiensscorer med det data og modelspecifikation, som anvendes.

### **SFA og MOLS er næsten den samme model**

SFA og MOLS har mange grundlæggende ligheder, og litteraturen refererer ofte til begge modeller som to forskellige varianter af SFA. Den eneste forskel mellem SFA og MOLS er, hvilken bagvedliggende teknik der bruges til at estimere modellerne. SFA estimeres ved hjælp af teknikken 'maximum likelihood'<sup>1</sup>, og MOLS estimeres ved hjælp af teknikken 'method of moments'<sup>2</sup>. I litteraturen fremstår SFA som en bedre teoretisk model sammenlignet med MOLS.<sup>3</sup> Der er dog studier, der påpeger, at MOLS i praksis kan være en bedre estimator for små datasæt.<sup>4</sup>

SFA er således teoretisk en stærkere model i forhold til MOLS, og derfor er SFA også mere udbredt end MOLS i regulering af forsyningssektorer. Der er dog regulatorer, som har erstattet SFA med MOLS, da der indgår for få selskaber i deres regulering til, at SFA giver retvisende estimater.

Bemærk, at der i litteraturen og diverse internationale rapporter er uklarhed om flere modellers navne. Forsyningssekretariatet ser ofte, at modellerne SFA, MOLS og COLS bliver forvekslet, og at diverse afarter af modellerne ikke rapporteres som afarter. Derfor gennemgås Forsyningssekretariatets definitioner af de tre modeller nedenfor. Definitionerne afspejler den videnskabelige konsensus.

### **Fremgangsmåde for SFA, MOLS og COLS**

Estimeringen af SFA, MOLS og COLS foregår overordnet i tre trin. I første trin estimeres en omkostningsfunktion, i andet trin forskydes omkostningsfunktionen, og i sidste trin estimeres inefficiens, jf. Figur 1.

---

<sup>1</sup> 'Maximum likelihood estimation' er en statistisk metode, der bruges til at finde parameterverdier, der maksimerer sandsynligheden for at observere de faktiske data.

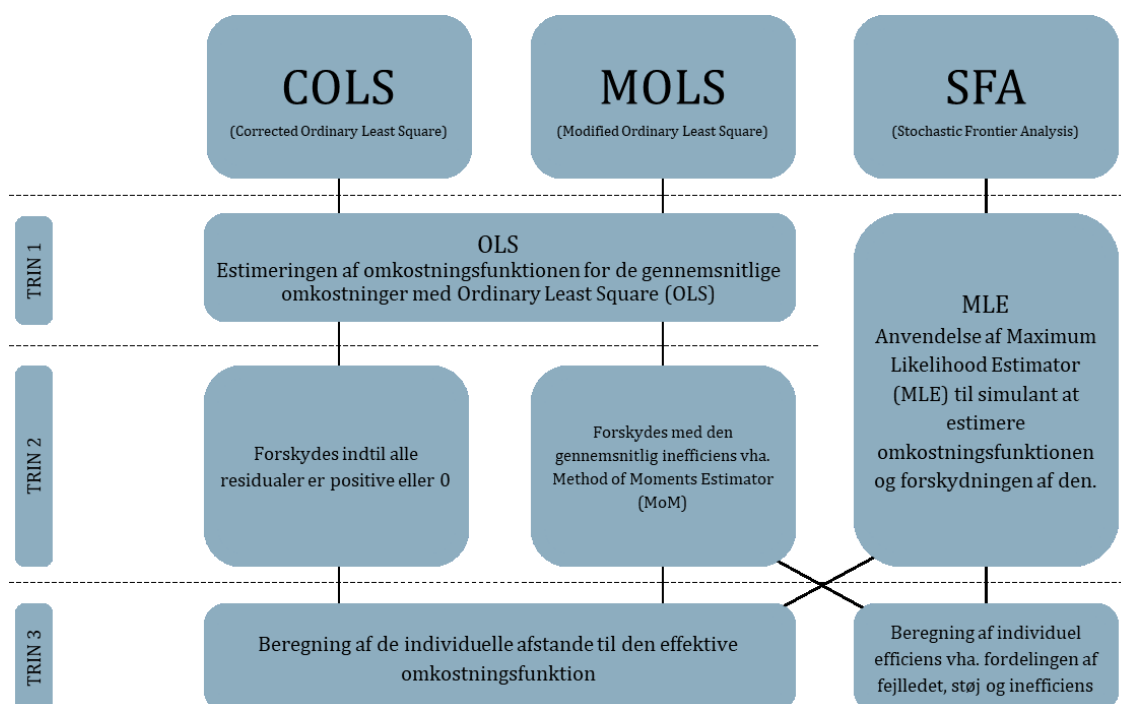
<sup>2</sup> 'Method of moments' er en statistisk metode til at finde en række nøgletal som fx middelværdi, varians og skævhed for en fordeling.

<sup>3</sup> Jf. Battese & Coelli (1988).

<sup>4</sup> Jf. Behr & Tente (2008).

Overordnet kan MOLS og COLS betragtes som teoretisk forsimplede udgaver af SFA. Det skyldes, at SFA estimerer alle relevante parametre på samme tid, hvorimod MOLS og COLS gør det i to opdeltte trin. Da MOLS og COLS er tæt relateret til SFA, har de dog fortsat mange af de samme fordele og ulemper som SFA. Forsimplingen kan dog medføre, at det er nemmere at anvende MOLS og COLS fremfor SFA i specifikke tilfælde.<sup>5</sup>

Figur 1 – Fremgangsmåden for SFA, MOLS og COLS



### Trin 1:

I trin 1 estimerer alle tre modeller en omkostningsfunktion. COLS og MOLS bruger klassiske regressionsmetoder (OLS), imens SFA bruger en mere avanceret maximum likelihood estimator (MLE). Omkostningsfunktionerne forklarer selskabernes gennemsnitlige omkostninger ud fra en række faktorer. I Forsyningssekretariatets benchmarking-setup bliver omkostningerne forklaret med OPEX- og CAPEX-netvolumenmålene samt selskabernes alder og tæthed i en translog-omkostningsfunktion.<sup>6</sup>

Bemærk, at omkostningsfunktionen i SFA estimeres simultant med trin 2, hvilket gør, at trin 1 og trin 2 for SFA påvirker hinanden.

<sup>5</sup> Jf. Ander & Hesse (2014)

<sup>6</sup> Selskabernes alder og tæthed medtages som såkaldte z-variable, hvilket betyder, at de ikke indgår i interaktioner i translog-omkostningsfunktionen.

**Trin 2:**

I trin 2 forskyder modellerne omkostningsfunktionen fra trin 1 nedad for at tage højde for, at der er inefficiens i sektoren. Omkostningsfunktionen i trin 1 kan betragtes som en omkostningsfunktion for de gennemsnitlige omkostninger (herefter gennemsnitlig omkostningsfunktion), hvorimod den forskudte omkostningsfunktion i trin 2 kan betragtes som den effektive omkostningsfunktion.

For at forskyde den gennemsnitlige omkostningsfunktion skal der foretages antagelser om, hvor stor en del af den gennemsnitlig omkostningsfunktion, der stammer fra inefficiens. COLS antager, at inefficiensen kan forklares ved det laveste residual (mest ekstreme negative residual) blandt selskaberne. MOLS og SFA antager, at inefficiens for den gennemsnitlige omkostningsfunktion skal findes ved at opdele residualerne i støj og inefficiens. Til brug for det antages det, at støj-delen af residualerne er normalfordelt, og at inefficiens-delen er halvnormalfordelt.

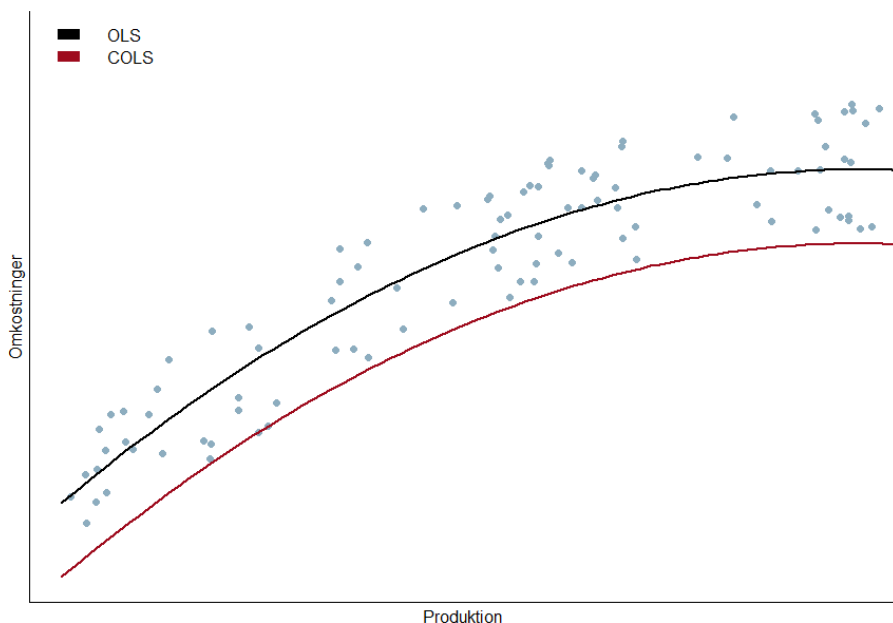
*Trin 2 i COLS:*

I COLS antages det som sagt, at den gennemsnitlige omkostningsfunktion skal forskydes ned med det laveste residual (mest ekstreme negative residual) blandt selskaberne, jf. Figur 2. Figuren viser, at den gennemsnitlige omkostningsfunktion fra OLS (sort linje) forskydes ned indtil den indkapsler alle selskaberne. Den forskudte omkostningsfunktion (rød linje) betegnes nu som den effektive omkostningsfunktion.

Bemærk, at COLS antager, at der ikke er støj i modellen, men at hele afstanden fra et selskab til den effektive omkostningsfunktion kan betragtes som inefficiens.

Forsyningssekretariatet vurderer ikke, at COLS har nogle fordele i forhold til Order-M, da COLS ikke kan opdele residualerne i støj og inefficiens. Denne version undersøges derfor ikke nærmere.

Figur 2 – Forskydning af omkostningsfunktionen fra OLS til COLS



### Trin 2 i MOLS

For at finde den effektive omkostningsfunktion ved anvendelse af MOLS, skal den gennemsnitlige omkostningsfunktion som tidligere beskrevet forskydes med den forventede gennemsnitlige inefficiens blandt selskaberne. For at beregne den forventede gennemsnitlige inefficiens i sektoren anvender MOLS 'method of moments'.

For at beregne den gennemsnitlige inefficiens opsættes antagelser om selskabernes fordeling af inefficiens og støj, som er identisk med antagelserne i SFA. I Forsyningssekretariatets benchmarking-setup antages det, at inefficiensen er halvnormfordelt, og at støjen er normalfordelt.

Beregningen af den gennemsnitlige inefficiens uddybes i Boks 1.

#### Boks 1 - Beregning af den gennemsnitlige inefficiens

Antag, at en standard lineær omkostningsfunktion i en OLS fra trin 1 er givet i ligning (1).  $x_i$  angiver selskab "i"s omkostninger, og  $f(y_i, \beta)$  angiver omkostningsfunktionen givet selskabets output  $y_i$  og OLS-koefficienterne  $\beta$ .  $\epsilon_i$  angiver residualen, som er den del af selskabets omkostninger, som omkostningsfunktionen ikke kan forklare.

$$x_i = f(y_i, \beta) + \epsilon_i \quad (1)$$

Residualt består af henholdsvis støj og inefficiens, jf. ligning (2). Bemærk, at støjen og inefficiensen er ukendt, så kun det samlede residual kendes.

$$\epsilon_i = v_i + u_i \quad (2)$$

Den gennemsnitlige inefficiens  $\bar{u}$  fra residualt på tværs af selskaberne skal udskilles. Her udnyttes det, at fordelingen af henholdsvis støj og inefficiens antages at være forskellig. Det antages, at støjen er normalfordelt, og at inefficiensen er halvnormalfordelt. Intuitivt kan det nu ses, at hvis residualt følger en normalfordeling, består hovedparten af støj. Hvis derimod residualt følger en halvnormalfordeling, består hovedparten af inefficiens.

Den gennemsnitlige inefficiens  $\bar{u}$  kan beregnes ud fra standardafvigelsen på inefficiensen ( $\sigma_u$ ) i en halvnormalfordeling, jf. ligning (3), hvor  $\pi \approx 3,14$ . Det betyder, at  $\sigma_u$  skal estimeres, før den gennemsnitlige inefficiens kan beregnes.

$$\bar{u} = \sigma_u \sqrt{\frac{2}{\pi}} \quad (3)$$

$\sigma_u$  kan beregnes ved hjælp af 'method-of-moments' og efterfølgende indsættes i formelen for  $\bar{u}$ .

Til det skal bruges det tredje moment for at beregne  $\sigma_u$ . Det tredje moment fortolkes som en fordelings skævhed. Bemærk, at en normalfordeling ikke har nogen skævhed, hvorimod en halvnormalfordeling har en skævhed tæt på 1.

Det teoretiske tredje moment er givet i ligning (4), og det empiriske tredje moment er givet ved ligning (5), hvor  $K$  angiver antallet af selskaber.<sup>7</sup> Bemærk, at der ledes efter standardafvigelsen for inefficiensen ( $\sigma_u$ ), og at denne er indeholdt i det teoretiske moment.

$$M_3 = \sqrt{\frac{2}{\pi}} \cdot \left(\frac{4}{\pi} - 1\right) \sigma_u^3 \quad (4)$$

$$\hat{M}_3 = \frac{\sum_{i=1}^K (\epsilon_i - \bar{\epsilon})^3}{K} \quad (5)$$

<sup>7</sup> Jf. Kuosmanen (2012)

$\sigma_u$  isoleres ved at sætte det teoretiske og empiriske moment lig hinanden. På den måde udnyttes det, at det empiriske moment kan estimeres, hvilket det teoretiske ikke kan. Det resulterer i ligning (6). Standardafvigelsen for inefficiensen er nu beregnet og kan indsættes i ligning (3).

$$\sigma_u = \sqrt[3]{\frac{\hat{M}_3}{\sqrt{\frac{2}{\pi}} \cdot \left(\frac{4}{\pi} - 1\right)}} \quad (6)$$

#### *Trin 2 i SFA:*

I SFA forklares inefficiensen i den gennemsnitlige omkostningsfunktion som branchens gennemsnitlige inefficiens, ligesom det var tilfældet for MOLS. SFA estimerer imidlertid den gennemsnitlige inefficiens samtidig med trin 1 i én samlet estimering. Det vil sige, at SFA anvender MLE (maximum likelihood estimator) til at estimere alle parametrene for den gennemsnitlige omkostningsfunktion samtidig med, at den estimerer de relevante standardafvigelser for henholdsvis støj og inefficiens.

Ligesom i MOLS antages det, at støjen er normalfordelt og inefficiensen er halvnormalfordelt, hvilket også følger antagelserne i de fleste SFA-analyser i den akademiske litteratur.<sup>8</sup>

MLE forsøger gennem en iterative proces at maksimere sandsynligheden for at observere de faktiske data under antagelsen om støj. På den måde forsøger MLE at korrigere omkostningsfunktionen for den observerede støj samtidig med, at den estimerer de sande parametre i omkostningsfunktionen.

#### **Trin 3:**

I det sidste trin bestemmes 1) selskabernes efficiensscore og/eller 2) selskabernes individuelle afstand til den effektive omkostningsfunktion.

##### 1) Selskabernes efficiensscorer

For MOLS og SFA kan selskabernes individuelle efficiensscorer beregnes på baggrund af fordelingen af støj og inefficiens. Der findes flere forskellige estimatorer til dette, som giver cirka samme resultater. Beskrivelsen undlades i dette notat, da det ikke er relevant for problematikken i næste kapitel om robusthedsanalysen.<sup>9</sup>

<sup>8</sup> Jf. Bogetoft & Otto (2010)

<sup>9</sup> Til at beregne individuelle efficiensscorer kan man fx anvende JMLS estimator, jf. Jowdrow et. al (1982), eller point estimator, jf. Battese & Coelli (1988)

## 2) Selskabernes individuelle afstand til den effektive omkostningsfunktion

For alle modellerne kan selskabernes effektive omkostningsniveau angives ved at beregne selskabernes afstand fra deres faktiske omkostninger til den effektive omkostningsfunktion.<sup>10</sup> Denne afstand kan bruges til at bestemme det effektive omkostningsniveau og dermed et individuelt effektiviseringskrav. Afstanden tager dog ikke højde for, at det enkelte selskab kan opleve støj. Afstanden kan derfor ikke betragtes som en efficiensscore i modeller, der antager, at der er støj.

### Robusthedsanalyse

Forsyningssekretariatet har undersøgt robustheden af MOLS og SFA i forbindelse med regulering af vandselskaber ved at undersøge, hvad der sker i modellerne, når der tilføjes støj til data.<sup>11</sup> I analysen anvendes data fra benchmarkingen til brug for de økonomiske rammer for 2023, hvor der er fjernet ét selskab som outlier. Selskabet, der er fjernet, påvirker således ikke estimerne, men vil fortsat få tildelt en efficiensscore.<sup>12</sup>

Robusthedsanalysen består af 100 iterationer. I hver iteration estimeres henholdsvis en ny SFA og MOLS på baggrund af Forsyningssekretariatets nuværende SFA-specifikation. Det vil sige, at der bruges en translogomkostningsfunktion med de eksisterende netvolumenmål, alder og tæthed. Den eneste forskel mellem iterationerne er, at der tilføjes nyt simuleret støj til selskabernes omkostninger. Det vil sige, at Forsyningssekretariatet bruger de simulerede omkostninger fra ligning (7) og (8) nedenfor i hver iteration, hvor  $\tilde{x}_{i,t}$  angiver de simulerede omkostninger for selskab  $i$  og iteration  $t$ .  $\epsilon_{i,t}$  er simuleret støj fra en normalfordeling med middelværdien 1 og en standardafvigelse på 0,05. Bemærk, at den absolutte middelværdi af den simulerede støj er cirka 4 pct., hvilket kan tolkes til, at der i gennemsnit lægges/fratrækkes 4 pct. støj til hvert selskabs faktiske omkostninger.

$$\tilde{x}_{i,t} = x_{i,t} \cdot \epsilon_{i,t} \quad (7)$$

$$\epsilon_{i,t} \sim N(1; 0,05) \quad (8)$$

Både SFA og MOLS bør teoretisk identificere den simulerede støj som netop støj. Teoretisk bør efficiensscorene i hver iteration derfor være

---

<sup>10</sup> Det effektive omkostningsniveau angiver, hvor mange omkostninger et selskab maksimalt bør have afholdt for at producere selskabets mængde af output, hvis det er efficient, og hvis der ikke er støj for selskabet.

<sup>11</sup> Bemærk, at Forsyningssekretariatet undlader COLS, da det allerede er vist, at den ikke teoretisk har nogle fordele over Order-M.

<sup>12</sup> KALUNDBORG OVERFLADEVAND A/S er fjernet. Bemærk, at Forsyningssekretariatet har været nødt til at bruge samme data og outlier som for benchmarkingen til de økonomiske rammer for 2023, da SFA ellers ikke ville konvergere til et resultat. Det indebærer, at Forsyningssekretariatet bruger det såkaldte fastfrosset data til at estimere modellen, og at der er én outlier. Dette adskiller sig resten af Forsyningssekretariatets analyse om ny benchmarkingmodel, hvor der bruges to outliers og det nyeste opdateret data fremfor det fastfrosset.



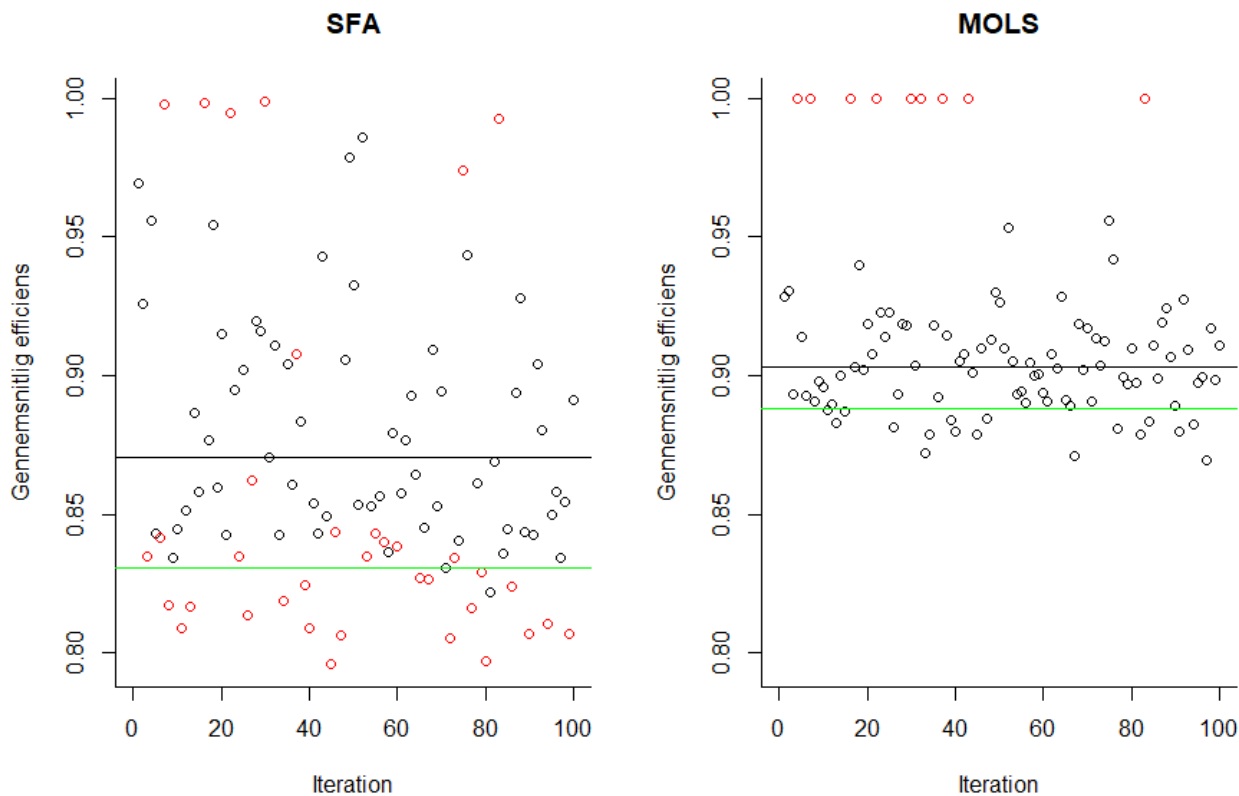
upåvirket af denne simulering. I praksis forventes dog, at resultaterne ændrer sig en smule, da modellerne ikke præcist kan identificere, om den simulerede støj er sand støj eller inefficiens/efficiens.

Resultaterne viser imidlertid, at både SFA og MOLS er meget følsomme overfor den simulerede støj, jf. Figur 3. Figuren viser branchens gennemsnitlige efficiensscorer i henholdsvis SFA og MOLS for 100 iterationer, hvor der er tilføjet støj til selskabernes omkostninger. De sorte horisontale linjer indikerer de gennemsnitlige efficiensscorer over alle iterationerne, og de grønne horisontale linjer indikerer efficiensscorerne i en model, hvor der ikke er simuleret eller tilføjet ekstra støj til selskabernes omkostninger.

Figuren illustrerer, at hverken SFA eller MOLS er robuste modeller, når der anvendes data fra vandselskaberne og Forsyningssekretariatets modelspecifikationer. I SFA svinger branchens gennemsnitlige efficiens mellem cirka 0,8 og 1. Dette er et meget højt spænd i forhold til, at der kun er simuleret relativt lidt støj i modellen (gennemsnitligt 4 pct.). I MOLS svinger resultaterne ligeledes meget, men dog i mindre grad end i SFA. Bemærk, at begge modeller teoretisk bør være upåvirket af den simulerede støj, da de netop bør kunne identificere den og dermed tage hensyn til den.

Resultaterne viser samtidig, at modellernes underliggende antagelser ikke er opfyldt i flere iterationer. Disse iterationer er angivet med de røde punkter i figuren. Efficiensscorer på 1 i SFA og MOLS skal betragtes som, at modellerne ikke har kunnet estimere nogen resultater.

Figur 3 – Efficiensscorer i SFA og MOLS med simuleret støj



Forsyningssekretariatet oplever hovedsageligt tre underliggende fejl i estimationerne:

1. For det første har modellernes residualer i nogle iterationer en negativ skewness. Det betyder, at residualerne ikke retvisende kan opdeles med antagelserne om, at inefficiensen er halvnormalfordelt, og støjen er normalfordelt.
2. For det andet er forholdet mellem støj og inefficiens i flere iterationer urealistisk. Det forekommer, når modellerne tilskriver det fulde residual enten støj eller inefficiens, hvilket ikke i praksis er sandsynligt.
3. Slutteligt er der også iterationer, hvor estimatoren melder fejl om modelspecifikationer. Dette kan forekomme, når antagelsen om den multivariate normalfordeling i data ikke er opfyldt, når der er for lidt variation i data, eller når uafhængige variable er stærkt korrelerede med hinanden.

Både SFA og MOLS bygger, som tidligere beskrevet, på OLS. Analysens resultater viser, at denne OLS er robust.<sup>13</sup> Problemet opstår, når modellerne forsøger at opdele residualerne i inefficiens og støj. Dette element

<sup>13</sup> Forsyningssekretariatet har undladt at vise resultater for dette i notatet.

er ikke robust, jf. Figur 4. Figuren viser den såkaldte lambda for hver iteration i SFA og MOLS. Lambda angiver, hvor stor en del af residualen, der kommer fra henholdsvis støj og inefficiens. Et højt lambda angiver, at der er meget inefficiens relativt til støj, og et lavt lambda angiver, at der er meget støj relativt til inefficiens. De grønne horisontale linjer angiver lambda fra de oprindelige modeller uden simuleret støj. De røde punkter i SFA indikerer lambda-værdier, der er større end 15, og som for illustrationens skyld er repræsenteret ved et rødt punkt øverst i figuren.<sup>14</sup> De røde punkter i MOLS indikerer lambda-værdier, der ikke kan beregnes, hvorfor de sættes til 0.

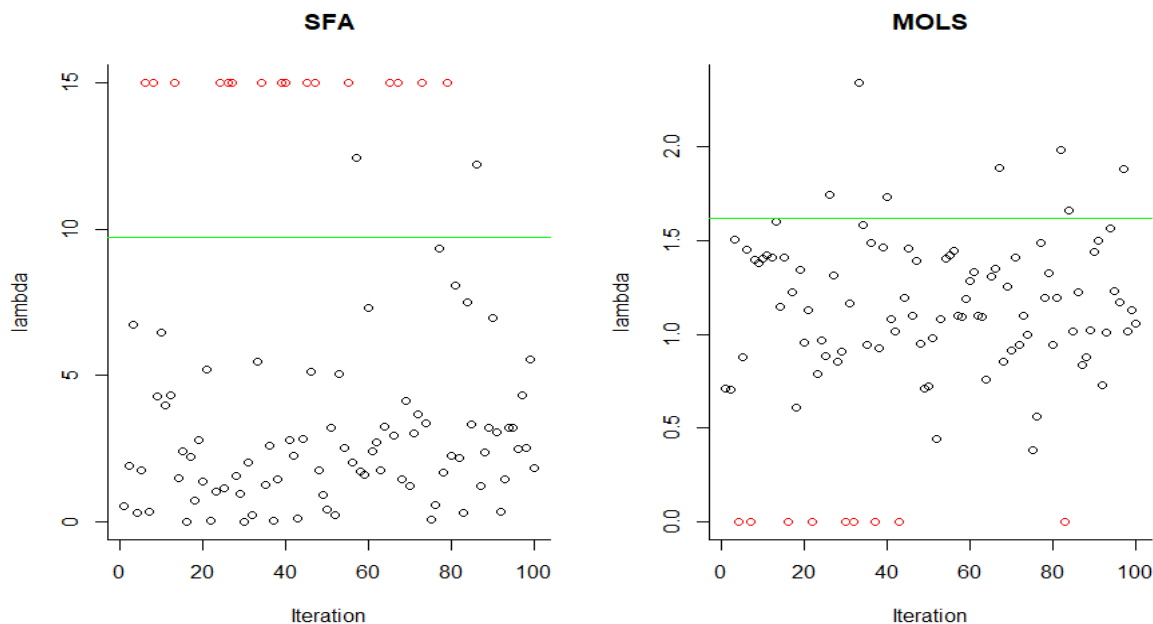
Figuren viser, at lambda falder, når støj tilføjes til modellen. Dette er forventet, da det netop medfører, at der kommer mere støj i forhold til inefficiens. Det er dog ikke forventet, at lambda i nogle iterationer stiger, hvilket er et tegn på, at modellerne ikke korrekt identificerer den simulerede støj som støj. Figuren viser samtidig, at der er stor spredning i lambda. Det betyder igen, at modellerne ikke konsistent kan identificere, at den tilføjede støj er støj.

Slutteligt viser figuren, at lambda er væsentlig højere i SFA end i MOLS. Det betyder, at SFA tilskriver en større andel af residualerne til inefficiens, end MOLS gør. Dette svarer også til resultaterne for Figur 3, hvor SFA giver lavere efficiensscorer end MOLS.

---

<sup>14</sup> Bemærk, at flere af disse iterationer har en lambdaværdi på mere end 1.000, hvilket er et tegn på, at modellens ikke kan beregne retvisende resultater.

Figur 4 – Fordeling af lambdaværdien ved 100 iterationer af SFA og MOLS



MOLS er mindre følsom for simuleret støj end SFA, fordi den effektive omkostningsfunktion i MOLS estimeres i to trin, hvorimod den i SFA estimeres i ét trin. Det betyder, at MOLS estimerer en retvisende og robust gennemsnitlig omkostningsfunktion i trin 1, men at MOLS har svært ved efterfølgende at forskyde den gennemsnitlige omkostningsfunktion til en effektiv omkostningsfunktion i trin 2. I SFA sker estimationerne simultant, hvilket betyder, at problemerne med at fordele inefficiens og støj også påvirker estimationen af den gennemsnitlige omkostningsfunktion, hvorfor alle resultaterne svinger mere.

### Forsyningssekretariatet har løbende udviklet SFA-modellen

Forsyningssekretariatet har siden indførelsen af SFA til de økonomiske rammer for 2017 oplevet tilsvarende problemer med modellen, som er beskrevet i dette notat. Forsyningssekretariatet er løbende blevet opmærksomme på omfanget af problemerne og er derfor også blevet klogere på dem. Forsyningssekretariatet har derfor løbende haft fokus på at forbedre SFA-modellen, hvilket er sket af flere omgange.

Forbedringerne af SFA indebærer blandt andet, at antallet af variable er reduceret for at undgå problemer med multikollinearitet, at den underliggende omkostningsfunktion er udskiftet fra en Cobb-Douglas til translog, og at effekten af selskabernes alder og tæthed i højere grad bliver estimeret korrekt. Herudover har Forsyningssekretariatet løbende analyseret, hvordan SFA kan forbedres. Dette indebærer blandt andet, at Forsyningssekretariatet har undersøgt, hvilke statistiske antagelser om fordelingerne

af henholdsvis støj og inefficiens der passer bedst til reguleringen af vandselskaberne.

En stor del af dette løbende udviklingsarbejde er foretaget med konsultation fra Arne Henningsen, som er lektor på Københavns Universitet og ekspert i SFA. Rådgivningen fra Arne Henningsen ledte til Forsyningssekretariatets seneste tekniske SFA-papir.<sup>15</sup>

SFA bruges ofte til analyseformål, hvilket giver god mulighed for at vælge den modelspecifikation, der passer bedst, i det konkrete år som analysen laves for. Det betyder dog ikke, at den valgte modelspecifikation beskriver den sande underliggende struktur for den analyserede sektor. Det medfører, at en ny analyse på et opdateret datagrundlag ville resultere i en anden modelspecifikation. Dette refereres i litteraturen til ”overfitting”.

Da SFA løbende skal bruges til årlig benchmarking, har Forsyningssekretariatet ikke samme mulighed for at ignorere problemet med ”overfitting”, som man har til analyseformål. Forsyningssekretariatet skal vurdere og vælge den modelspecifikation, som bedst beskriver den sande underliggende struktur for vandselskaberne. Det er ikke sandsynligt, at den underliggende struktur for vandselskaberne ændrer sig markant fra år til år. Forsyningssekretariatet kan derfor ikke troværdigt ændre modelspecifikationen fra år til år.

En løbende SFA-model skal være konsistent i forhold til de underliggende antagelser om fx omkostningsfunktionens form. Hvis omkostningsfunktionens form ændrer sig mellem årene, er det et tegn på, at den underliggende antagelse ikke er korrekt.<sup>16</sup> Netop dette løbende arbejde med den samme SFA-model har gjort, at flere regulatorer har problemer med SFA.<sup>17</sup>

Forsyningssekretariatet vurderer, at SFA-modellen er udarbejdet og udviklet til den bedst mulige specifikation til brug for regulering af vandselskaberne. Hertil vurderer Forsyningssekretariatet, at på grund af de fortsatte udfordringer, som er beskrevet ovenfor, bør der arbejdes med mulighederne for at anvende alternative benchmarkingmodeller.

## Opsamling

På baggrund af Forsyningssekretariatets løbende udvikling af SFA og ovenstående robusthedsanalyse vurderer Forsyningssekretariatet, at der er betydelige udfordringer ved at bruge SFA og MOLS til regulering af vandselskaberne. Forsyningssekretariatet vurderer derfor, at SFA bør udfases i benchmarkingen af vandselskaber, og at MOLS ikke er et velegnet

---

<sup>15</sup> <https://www.kfst.dk/media/zzwie0bf/dokumentationsnotat-for-forsyningssekretariatets-sfa-model.pdf>

<sup>16</sup> Dette kaldes normalt for overfitting. Hvis en model overfittes i ét år for at give en god estimation, resulterer det typisk i dårlige estimater næste år.

<sup>17</sup> Blandt andet er både Forsyningstilsynet og E-control (østrigsk el-regulator) gået væk fra SFA.

alternativ i reguleringen. Forsyningssekretariatet vurderer, at det i stedet for er relevant at anvende en mere robust model, som fx DEA-Order-M.

Teoretisk har SFA og MOLS den fordel over DEA-Order-M, at de kan opdele residualerne i støj og inefficiens. Det er dog netop i forsøget på at dele residuallet op, at der i praksis opstår problemer med SFA og MOLS, jf. ovenstående analyse. Modellerne kan således i praksis ved anvendelse på data fra vandselskaberne ikke estimerer fordelingen mellem støj og inefficiens troværdigt, og SFA og MOLS har derfor i praksis ikke den fordel sammenlignet med DEA-Order-M.