



Forsyningsssekretariatet
Nyropsgade 30
1780 København V



Skanderborg, den 8. april 2011

Hørings svar på udkast til ”Netvolumen mål og Omkostnings- ækvivalenter”

FVD og DANVA vil gerne takke for muligheden for at afgive høringssvar på ”Netvolumen mål og Omkostningsækvivalenter”. Endvidere vil vi gerne bemærke, at Forsyningsssekretariatet i løbet af processen generelt har søgt at inddrage branchen på en fornuftig måde, bl.a. ved at man har valgt en forholdsvis åben linje med hensyn til adgang til data. FVD og DANVA påskønner dette og håber på at kunne fortsætte den åbne dialog fremover. Angående udkastet har Forsyningsssekretariatet, under de givne forudsætninger, lavet et fint analysearbejde, men den underliggende datakvalitet og manglende data på en række væsentlige parametre gør dog, at FVD og DANVA er skeptiske overfor resultaternes pålidelighed. I flere tilfælde afviger resultaterne fra hvad man kunne forvente, hvilket underbygger, at der er problemer med datakvalitet og udeladte variable.

Tidsplan

Forsyningsssekretariatet har skullet gennemføre grundlaget for den resultatorienterede benchmarking på meget kort tid. Sammenligner man processen med implementeringen af lignende reguleringssystemer indenfor el-området i vores nabolande, har der været usædvanlig kort tid til at indsamle og behandle data i reguleringen af den danske vandsektor. Den meget korte dataindsamlings- og analysefase har naturligvis konsekvenser for det grundlag, der er til rådighed for udarbejdelsen af den resultatorienterede benchmarkingmodel. Ud fra erfaringerne med lignende systemer i andre lande, ville det have været rimeligt at afsætte 2-3 år til indledende analyser i stedet for knap 1 år, som det er tilfældet i nærværende arbejde. Den meget stramme tidsplan har konsekvenser for da-

DANVA,
Vandhuset, Danmarksvej 26, 8660 Skanderborg, Telefon: 70 21 00 55, Email: danva@danva.dk

FVD,
Solrød Center 22 c, 2680 Solrød Strand, Telefon: 56 14 42 42
Email: fvd@fvd.dk

takvaliteten, forståelsen af data og hvilke parametre, der indgår i modellen. Dette er ikke tilfredsstillende for branchen, og FVD og DANVA opfordrer derfor til, at der fremadrettet vil blive arbejdet målrettet på at forbedre grundlaget for den resultatorienterede benchmarking. FVD og DANVA opfordrer endvidere Forsyningssekretariatet til at være meget lempelige i anvendelsen af den resultatorienterede benchmarking indtil der foreligger et bedre sammenligningsgrundlag.

Gennemgang af materiale

FVD og DANVA har gennemgået udkastet, ligesom Professor Peter Bogetoft, CBS har gennemgået og kommenteret udkastet for DANVA. FVD og DANVA vil dog gerne understrege, at der ikke er foretaget en grundig gennemgang af de enkelte beregninger, men at der kun er kigget på udvalgte områder. Der er flere områder, hvor der potentielt kunne laves forbedringer, men hvor FVD og DANVA ikke har haft ressourcer til indenfor høringsperioden at analysere dette, da der i givet fald vil være tale om et omfattende analysearbejde.

Datakvalitet

Kvaliteten af de data, der ligger bag beregningen af omkostningsækvivalenterne er, efter FVD og DANVA's opfattelse, desværre ikke særlig god. Der er en række forklaringer på, hvorfor vi mener det er svært at stole på data:

- Der blev aldrig lavet et pilotstudie forud for den egentlige dataindsamling.
- Definitionerne af både de tekniske og de økonomiske parametre har været utilstrækkelig.
- Kontoplanerne i selskaberne har ikke været nøjagtige nok i forhold til det ønskede detaljeringniveau, hvorfor der ligger et betydeligt antal skøn bag data.
- Der blev aldrig udarbejdet en vejledende konteringsinstruks.
- Forsyningssekretariatet efterspurgte data fra 2009, hvor mange endnu ikke var selskabsudskilte med deraf følgende problemer.

Et pilotstudie blandt en mindre gruppe af selskaber forud for den egentlige dataindsamling havde muliggjort indledende analyser af tendenser i data. Herved kunne man have fået værdifuld information om, hvilke parametre, der burde indsamles data om. Et andet og måske endnu vigtigere udbytte af et pilotstudie havde været, at mange af de problemer der sidenhen viste sig med eksempelvis indberetning af husstande, udregning af volumen for ledninger, definition af renselanlægsvariabler og så videre i vid udstrækning kunne have været undgået i forbindelse med den egentlige dataindsamling. Dermed kunne man i højere grad have imødegået et af de store problemer i forbindelse med dataindsamlingen, nemlig de utilstrækkelige definitioner.

Selskabernes kontoplaner har ikke været indrettet til at opsamle data på det efterspurgte detaljeringniveau. Dette har ført til, at mange selskaber har været nødt til at fordele en meget stor del af deres omkostninger ud fra skøn. Da der samtidig ikke har været særligt klare definitioner af, hvad hver post i indberetningen skulle indeholde har det vanskeliggjort indberetningen for selskaberne. Det kunne bl.a. have været hensigtsmæssigt med en overordnet vejledende konteringsinstruks, der specificerede, hvilke omkostninger der skulle henføres til hvilke cost drivers på et mere detaljeret niveau.

Et af de store vandselskaber, TRE-FOR, har fordelt øvrige omkostninger ved hjælp af POLKA-værdierne for de enkelte områder, hvilket de i forbindelse med indberetningen har oplyst Forsyningssekretariatet om. Havde TRE-FOR i stedet fordelt omkostningerne ud fra fordelingen af de direkte og indirekte henførte omkostninger, havde det givet et væsentligt anderledes resultat, som det kan ses af nedenstående tabel.

Tabel 1 viser forskel i procent på om øvrige omkostninger er fordelt ud fra POLKA-værdier eller ud fra fordelingen af direkte og indirekte omkostninger.

Cost driver	Forskel i %
Boring	37%
Råvandsledning	-48%
Vandværk	30%
Trykforøgere	25%
Rentvandsledning	-23%
Stik	-15%
Kunder	-4%
Driftsomkostninger fra costdrivere i alt	0%

Eksemplet illustrerer, at det er meget vigtigt for kvaliteten af data, at andelen af skøn minimeres og at der er præcise beskrivelser af hvilke kriterier skøn skal baseres på, da man ellers får en meget forskellig fordeling af omkostningerne selskaberne imellem. I den forbindelse er det meget bekymrende, at andre selskaber har angivet, at de har fordelt øvrige omkostninger ud fra den relative fordeling af de direkte og indirekte henførte omkostninger. Alle selskaber har altså ikke benyttet samme princip for fordelingen af omkostninger, da der har været tvivl om hvordan fordelingen skulle foretages.

Selskaberne har skullet indberette 2009 data. Dette har givet betydelige problemer for mange selskaber, da nogle selskaber er blevet selskabsgjorte midt i året, andre før 2009 og andre igen ultimo 2009. Det havde klart været at foretrække, at data var base-

ret på regnskaber, der alle var udarbejdet efter årsregnskabsloven. Data baseret på 2009 regnskaber øger usikkerheden i indberetningerne betydeligt.

Drift og vedligehold kontra investeringer er i Forsyningssekretariatets vejledning til indberetning af cost driver data til ækvivalentberegningerne defineret således at "*udbedringer der forlænger levetiden med mere end 10% af levetiden betragtes som investeringer*". Denne definition er langt fra praksis hos selskaberne, heller ikke hos de selskaber, der udarbejdede revisorgodkendte årsregnskaber i 2009. Alligevel har Forsyningssekretariatet valgt at holde de indberettede cost driver data op mod selskabets årsregnskab, således at der skulle opnås balance i forhold til driftsomkostningerne i selskabets officielle årsregnskab. Da der er forskellige definitioner på investeringer i de to opgørelser, giver det ikke mening at kræve, at der skal være balance. Fremgangsmåden har været med til at skabe uklarhed om datagrundlaget.

I udkastet til beregning af omkostningsækvivalenter har Forsyningssekretariatet ikke inkluderet et afsnit, der beskriver og vurderer kvaliteten af data. Det er således uklart hvordan Forsyningssekretariatet vurderer data, samt præcist hvilken kvalitet data har haft. Har både store og små selskaber indberettet i lige stort omfang, har de tidligere kommunale selskaber kunnet fremskaffe tilfredsstillende regnskabsdata, er indberetningerne komplette, hvilke cost drivere, har givet problemer, hvor stor en del af omkostningerne er blevet fordelt ud fra skøn og hvor mange er direkte henførte til de relevante poster for at nævne nogle relevante informationer man savner i udkastet.

Netvolumenmål og datakvalitet

Hele tilgangen med at beregne omkostningsækvivalenter for de forskellige cost drivers er helt afhængig af, at den fordeling af omkostningerne, som selskaberne har lavet på de enkelte cost drivers, er meningsfuld. Da der er tvivl om den initiale fordeling af omkostningerne, er det nødvendigt med en forsigtig anvendelse af den resulterende benchmarking model.

At der rent faktisk er problemer med datakvaliteten understøttes af, at mange af modellerne ikke kan forklare en betydelig del af omkostningsvariationen. Endvidere er der i nogle af modellerne problemer med koefficienter, som er meget signifikante, men som vender forkert, hvilket kan skyldes problemer med fordeling af omkostninger i selskaberne. Konsekvensen er, at omkostningsækvivalenterne ikke synes at give overbevisende mening, og at man skal være varsom med at fortolke dem. Mange selskaber bekræfter dette, da de oplever, at der er meget stor forskel på de beregnede værdier for de enkelte cost drivere og de faktiske omkostninger. I mange tilfælde udligner det sig

noget hen over de forskellige cost drivere, men de enkelte omkostningsækvivalenter virker ikke overbevisende.

De manglende fits vil bl.a. betyde, at selskaber, som totalt har et udmærket fit, men hvor de enkelte dimensioner ligger skævt, kan komme i problemer fremadrettet, hvis ikke alle aktiviteter udvides proportionalt. Hvis man fx ligger godt på ledninger men dårligt på stik, fordi man har lagt mange omkostninger på stik i forhold til andre selskaber, så kan man få problemer, hvis man fremadrettet primært ekspanderer på stik.

Usikkerhed omkring validiteten af omkostningsækvivalenterne kan gøre det nødvendigt at overveje om et netvolumenmål er den rigtige løsning eller om det er nødvendigt at sætte cost driverne direkte ind i en DEA-model. Såfremt et netvolumenmål fremadrettet skal være den rigtige metode, kræver det mere troværdige omkostningsækvivalenter. I mellemtiden er det nødvendigt at overveje, hvordan usikkerheden på omkostningsækvivalenterne kan håndteres. FVD og DANVA indgår gerne i en nærmere dialog med Forsyningssekretariatet om mulige løsninger.

Miljø- og servicemål

I den nuværende model tages der ikke hensyn til miljø- og servicemål i selskaberne. Dette er meget problematisk, da Forsyningssekretariatet har lagt en praksis om, at miljø- og servicemål, der var inkluderet i drift og vedligeholdelsesomkostningerne i 2003-05 ikke tages ud af prisloftet på drift og vedligehold. Det vil sige, at der i de omkostninger, der indgår i benchmarkingen, er betydelige omkostninger til miljø- og servicemål. I det nuværende udkast indeholder modellerne ikke nogen direkte variabler, der fanger forskelle i disse omkostninger. Da der er store forskelle på, hvilke målsætninger selskaberne har på dette område, er der behov for, at selskaberne kompenseres for at levere en højere service eller for at leve op til ekstra strenge miljøkrav. Det ligger også i intentionerne i Vandsektorloven, at selskaberne skal benchmarkes på miljø og kvalitet.

Det skaber desuden problemer, at reglerne omkring miljø- og servicemål ikke var særligt veldefinerede, da Forsyningssekretariatet indsamlede data til omkostningsækvivalenterne. Mange selskaber har fået afslag på ansøgning om en lang række miljø- og servicemål, og der pågår en klagesag på området. Desuden synes det at være et erkendt problem, at dokumentationskravene ikke var særligt klare, og at det har været vanskeligt for selskaberne at fremskaffe den relevante dokumentation til prisloftet for 2011. Det betyder, at man fremadrettet kan have en forventning om, at godkendte omkostninger til miljø- og servicemål vil stige de kommende år. Dette vil sætte omkostning-

sækvivalenternes validitet under yderligere pres fremover, da grundlaget for beregningerne i givet fald ikke vil afspejle de omkostninger, der aktuelt indgår i benchmarkingen.

Udeladte variable

Da der ikke har været mulighed for at indrette kontoplaner med videre på indsamling af data til den resultatorienterede benchmarking, har det ikke været muligt at fremskaffe en række relevante data. Det gælder derfor for mange af modellerne, at væsentlige variable ikke indgår i modellerne, hvilket rejser tvivl om hvorvidt de estimerede omkostningsækvivalenter er retvisende. I mange tilfælde vil man forvente korrelation mellem de udeladte variable og variable indeholdt i analyserne. Et grelt eksempel, hvor udeladte variable rimeligvis spiller en rolle er ledninger. Her har man ikke data på materiale, alder, dybde og dimension – alle variable vi ved i et vist omfang er korreleret med eksempelvis zonefaktoren. De udeladte variable skaber desuden problemer i form af, at en væsentlig del af variationen i data ikke bliver beskrevet.

Alder

Alder indgår ikke i nogen af modellerne. Mange selskaber rapporterer, at der er en klar sammenhæng mellem alderen på et aktiv og de tilhørende drift- og vedligeholdelsesomkostninger. Formentlig vil Forsyningssekretariatet foretage en generel korrektion for alder baseret på indberetningerne til POLKA, men det havde formentlig været mere hensigtsmæssigt at inkludere alder i beregningen af omkostningsækvivalenterne og så derefter have taget stilling til om alder skulle inkluderes i beregningen af omkostningsækvivalenterne eller skulle indsættes direkte i DEA-modellen. FVD og DANVA har i tidligere høringssvar gjort opmærksom på, at alder er en potentielt betydende faktor for mange af modellerne.

Valg af omkostningsfunktion

I flere af modellerne anvendes en log-lineær specifikation for at håndtere heteroskedasticitet eller stordriftsfordele. Det er i flere tilfælde uklart, hvorfor man har valgt de angivne omkostningsfunktioner frem for andre funktioner. Valg af omkostningsfunktion er meget afgørende for analyserne, men da datakvaliteten er lav, kan det være svært at foretage de rigtige valg.

Heteroskedasticitet

Flere af modellerne synes at have tendens til heteroskedasticitet. FVD og DANVA vil derfor opfordre Forsyningssekretariatet til at undersøge nærmere om det er relevant at bruge korrigerede koefficienter og robust regression.

Håndtering af outliers

I udkastet er det angivet, at alle observationer med en Cook's afstand på over 1 fjernes fra datasættet, mens der laves en individuel vurdering af den enkelte observation ved en Cook's afstand mellem 0,5 og 1. Forsyningssekretariatet har dermed sat en meget høj grænse for, hvornår en observation betragtes som outlier. Dette harmonerer meget dårligt med, at der er stor usikkerhed omkring kvaliteten af data. Hvis der er usikkerhed om en observation er korrekt, er det ikke rimeligt, at man lader denne observation påvirke resultatet i større omfang. Desuden bør man følge op på alle observationer, der har stor indflydelse på regressionerne for at sikre sig, at de indberettede data er pålidelige. Det fremgår ikke af udkastet, om man har fulgt en sådan procedure.

FVD og DANVA vil foreslå, at man vælger en noget mere forsigtig tilgang, hvor alle observationer med en Cook's afstand større end $4/n$ ekskluderes, hvor n er lig antal observationer i regressionen. Finder man denne tilgang for forsigtig, kan man overveje et kriterium, der hedder $4/(n-p-1)$, hvor n er lig antal observationer og p er lig antal parametre i modellen. Metoden er nærmere beskrevet i Belsley et al. (1980) og blev brugt i forbindelse med regulering af el og gas i Tyskland, se evt. Agrell and Bogetoft (2008). De ovennævnte referencer er uddybet i bilag 3.

Man bør desuden inddrage andre kriterier end Cook's afstand i vurderingen af, om nogle af selskaberne under ovennævnte grænse er outliers. Man kan med fordel benytte DFBETAs, DFFITS og covariance ratio. En kort beskrivelse af metoderne ses i bilag 3.

Herudover anbefaler FVD og DANVA, at man undersøger kvaliteten af de data, der ligger bag observationer, der beholdes i modellerne, men som påvirker resultaterne væsentligt. Det er i den forbindelse ikke alene vigtigt at undersøge om de indberettede tal er korrekte, men også om det pågældende selskab eventuelt har nogle forhold, der gør at det er uegnet til at indgå i analyserne.

Det er i udkastet generelt ikke beskrevet, hvordan man efterfølgende vil håndtere outliers. En undtagelse er på stik, hvor man har anvist en model for håndtering af stik i indre city, der virker fornuftig. Forsyningssekretariatet bør specificere, hvordan man i de enkelte modeller vil håndtere outliers efterfølgende, eventuelt med udgangspunkt i den model, der er valgt på stik. Såfremt man vælger blot at sætte outliers ind i den estimerede model, giver man som udgangspunkt disse outliers en ufordelagtig behandling. I den endelige udgave bør man derfor angive hvilke selskaber, der er outliers i modellen, og hvordan disse selskabers omkostninger estimeres efterfølgende. Det er væsentligt, at selskaber, der kommer ud som outliers er informerede om dette, så de kan forholde

sig til det i de individuelle afgørelser. Nogle selskaber er outliers i mange modeller, og det er derfor vigtigt, at Forsyningssekretariatet finder en fornuftig måde at håndtere outliers på, da det ellers kan få store konsekvenser for "outlier selskaberne".

Anvendte kriterier og rækkefølgen af disse

Generelt savnes der en redegørelse for valg af kriterier og for den relative vigtighed af disse i forbindelse med udvælgelsen af variabler i modellerne. Det er for eksempel uklart, hvilket kriterium, der er brugt til at afgøre om modellerne skal indeholde et B_0 – og der synes ikke altid at være brugt det samme kriterium gennem udkastet. Endvidere er det uklart hvilken rækkefølge R^2 , rimelige fortegn og størrelsesrelationer og signifikante koefficienter, har i forhold til hinanden. Det er også uklart om Forsyningssekretariatet har benyttet sekventielt valg af cost drivers eller hvordan man har analyseret forskellige kombinationer af parametre i modellerne. Disse forhold kan have stor betydning for de endelige modeller og de overordnede kriterier bør derfor fremgå. Alternativt kan man få indtryk af, at man lidt usystematisk analyserer data indtil man har en nogenlunde plausibel model. En sådan metode kan indeholde visse risici, da man risikerer at modellerne passer godt på de aktuelle data, men ikke særligt godt når man sætter nye data, eksempelvis 2010 data, ind i modellerne.

Specifikke kommentarer til omkostningsækvivalenterne på spildevand:

Ledning

Tabel 12 i udkastet viser, at der er markant forskel på omkostningerne ved at drive ledning i land, by og city zone. Forsyningssekretariatet vælger at forkaste modellen da modellen har en p-værdi på 0,0777 for land. Dette synes at være baseret på en noget stram fortolkning af signifikans-kriteriet på bekostning af hvor stor en del af variationen modellen forklarer. DANVA opfordrer derfor Forsyningssekretariatet til at overveje at acceptere en høj p-værdi på land. Mange selskaber har ud fra en ingeniørmæssig betragtning rapporteret, at det er forventet, at det er billigere at drive ledninger i land-zone end ledninger i by-zone. Ud fra denne betragtning er modellen i tabel 12 alt andet lige at foretrække fremfor modellen i tabel 13. Man kan også diskutere om ikke det er bedre at håndtere Indre city separat på samme måde som man gør på stik på vand (dog med et vægtet gennemsnit). I modellen i tabel 13 ekskluderes 3 ud af 6 selskaber med indre city som outliers fra modellen, heraf ekskluderes de to selskaber, der har flest ledninger i indre city. Dermed bliver det meget diskutabelt hvor godt modellen i tabel 13 repræsenterer indre city.

Et særligt forhold under ledninger er, at bassinledninger medregnes som almindelige ledninger, selvom der er betydelige omkostninger forbundet med drift af disse. Bassinledninger har en meget stor dimension, og når dimension ikke indgår i modellen, stiller det selskaber med bassinledninger ekstra dårligt.

Det er ikke overraskende, at det er vanskeligt at finde en passende model for ledninger, da modellen formentlig lider af at flere betydende parametre ikke indgår i modellen. Der er ikke taget hensyn til materiale, dybde, antal tilslutninger, dimension og alder. Alle faktorer, der påvirker drift- og vedligeholdelsesomkostningerne i betydelig grad. Der er således grund til at tro, at modellen på ledninger lider under både ringe datakvalitet og manglende variable, hvorfor resultaterne bør fortolkes meget varsomt. Der kan stilles spørgsmål ved om det overhovedet er zonefaktoren, der driver de fleste omkostninger eller denne faktor blot er korreleret med andre og eventuelt mere betydende faktorer som dimension, materiale og alder.

Pumper

I modellen for pumper har man igen en meget konsekvent fortolkning af p-værdien. I tabel 14 har kun kategori e en høj p-værdi, mens de øvrige er signifikante eller tæt på at være signifikante. DANVA har lavet en regression med kategorierne a, b, c, d+e, f, hvor kategori a og b er klart signifikante, mens de øvrige har en p-værdi på 0,04-0,06. I bilag 1 ses en udskrift af regressionen. Bemærk der ikke er foretaget kvalitetssikring af data eller omfattende modelkontrol, hvorfor udskriften er til inspiration. Beregningen viser, at der er grund til at tro, at pumper fremadrettet bør indeles i minimum 4 kategorier. Om man skal vælge en lettere overfittet model med mange variable eller en stærkt forenklet model er et vanskeligt spørgsmål, da ingen af løsningerne er optimale. DANVA hælder til, at den lettere overfittede model i dette tilfælde ser mest troværdig ud. Igen kræver de usikre resultater dog, at man tager hensyn i fortolkningen af data.

DANVA vil gerne gøre Forsyningssekretariatet opmærksom på, at kategori f indeholder et stort spænd i pumpestørrelse, hvilket er vanskeligt for modellen at beskrive. Når man slår kategori d, e, f sammen bliver spændet meget stort. Det forhold bør som minimum medtages i eventuelle subjektive vurderinger. Især da den model DANVA har kørt viser, at kategori f er omtrent signifikant.

Åbne bassiner

Der er stor forskel i størrelsen på bassiner mellem selskaberne, hvilket ikke er indeholdt i modellen. DANVA har foretaget en regression, der viser, at justeret R^2 kan hæves til

cirka 0,42, hvis der inkluderes en parameter, der hedder: areal/antal bassiner. Modellen hedder

$$Y = B1 * \text{antal bassiner} + B2 * (\text{areal/antal bassiner}).$$

I bilag 2 er der en udskrift fra det benyttede statistikprogram. DANVA har ikke foretaget en grundig modelkontrol eller kvalitetssikring af data, hvorfor udskriften er tænkt som inspiration til videre arbejde.

Der er stor forskel på bassiner og deres vedligehold, hvorfor det ikke er overraskende, at den ret simple model ikke forklarer en ret stor del af variationen i data. Blandt andet har det betydning, hvor meget man benytter bassiner over året. Et andet forhold er, at bassinerne skal renses efter en årrække. Hvis ikke dette regnes som en reinvestering, kan det påvirke resultaterne.

Renseanlæg

Det er meget overraskende for selskaberne, at renseanlæg kan beskrives ud fra kun en variabel, faktisk belastning i PE. DANVA har tidligere anbefalet væsentlig flere faktorer og det er overraskende, at disse ikke bliver signifikante i modellerne. Det tyder kraftigt på, at datakvaliteten ikke er i orden. Modellen er simpelthen for simpel.

Mange selskaber har haft problemer med at fordele omkostningerne ud på de enkelte renseanlæg, da der ikke har foreligget konkrete registreringer. Det er tydeligt ud fra data, at en del selskaber har fordelt omkostningerne ud på de enkelte renseanlæg ud fra belastning i PE. I udkastet mangler en beskrivelse af hvor udbredt denne praksis er. Det er klart problematisk, at omkostningerne er fordelt ud fra PE, når den eneste forklarende variabel er PE! Det giver måske et godt fit, men man kan ikke sige, det er en retvisende praksis. Det er desuden særligt problematisk for renseanlæg, da der er stordriftsfordele. Når man fordeler omkostninger ud fra PE, ved vi altså, at det ikke er retvisende.

Foruden fordelingen af omkostninger er der også problemer med datakvaliteten. Når man måler PE sker det med en betydelig stikprøveusikkerhed afhængig af antallet af stikprøver taget på det enkelte renseanlæg. Ved vurdering af et selskabs faktiske driftsomkostninger til renseanlæg i forhold til de forventede gennemsnitlige driftsomkostninger beregnet ved hjælp af omkostningsækvivalenten, skal der tages højde for, at angivelse af den faktiske organiske belastningsgrad sker med en usikkerhed på op til 20 % af den angivne værdi. Den usikkerhed, der er på måling af PE, kan derfor medføre, at et selskab det ene år er effektivt og det næste er ineffektivt. Medmindre man tager forbe-

hold i den senere fortolkning af DEA-modellens resultater, skaber man altså et falskt effektiviseringspotentiale.

Den usikkerhed hvormed PE er bestemt, sænker altså datakvaliteten, men der er yderligere et væsentligt forhold i denne forbindelse, nemlig at PE kan svinge ganske betydeligt fra år til år pga. forskellig udledning fra eksempelvis industrier. Da belastningen ikke er kendt på forhånd er selskaberne nødt til at have et vist beredskab klart, hvilket naturligvis giver ekstra udgifter. Udsving i PE på op til 10-15 % vil forekomme for mange selskaber. Kombinationen af stikprøveudsving og udsving i belastningen taler for, at PE og de tilhørende omkostninger bør bestemmes over en årrække og ikke fra år til år.

Forsyningssekretariatet finder en betydelig størrelsesfordel ved drift af renseanlæg. DANVA er enig i vurderingen af, at der er betydelige størrelsesfordele ved at drive store renseanlæg, men umiddelbart virker den beregnede størrelsesfordel som værende noget markant. I en model, hvor DANVA vurderer, at der er udeladt mange betydende variable, er det svært at stole på beregningen.

Det er overraskende, at slamdisponering ikke bliver signifikant i modellen. Et eksempel på, at slamdisponering burde være en signifikant variabel i modellen, fremgår af følgende beregning fra VandCenter Syd:

"Vi er bekendt med at priserne på slutdisponering af spildevandsslam varierer afhængig af løsninger: a) direkte i landbrug ca. 250,- kr/ton b) kompostering ca. 550 kr/ton c) forbrænding ca. 800 kr/ton. I VandCenter Syd har vi i 2010 slutdisponeret 27.000 ton slam. Omkostningsmæssigt kan slutdisponeringen dermed variere fra billigste løsning = 6,75 mio. kr. til dyreste løsning = 21,6 mio. kr. Altså en difference på 14,9 mio. kr. afhængig af løsningsvalg på slutdisponering af spildevandsslammet. Når vi beregner netvolumen for vores 3 primære renseanlæg får vi et netvolumenbeløb på 57,1 mio. kr. Differensbeløbet på slutdisponering udgør altså 26 % af det samlede netvolumenbeløb på vores renseanlæg".

En mulig forklaring på, at så betydningsfuld en post ikke bliver signifikant kan være, at nogle selskaber køber forbrændingsydelsen ude i byen og bogfører den under drift, mens andre selskaber har egne forbrændingsanlæg og bogfører en del af udgiften under investeringer. En anden forklaring kan være, hvis Forsyningssekretariatet har forsøgt at estimere slamdisponering som en dummy variabel, hvorved slamdisponering bliver et fast beløb pr. renseanlæg. Denne fremgangsmåde giver ikke mening når man sammenligner minirenselanlæg med Lynettefællesskabet, hvorfor slamdisponering ikke

kan indgå som en dummy variabel. En lignende problemstilling gør sig gældende for de øvrige variabler under renseanlæg.

VandCenter Syd har også regnet på betydningen af fosforkrav, der også synes at burde være i modellen:

"Variationerne i fosforkrav tilgodeses ikke i ækvivalentberegningen. Det generelle krav til fosforrensning er ned til maksimalt 1,5 mg/l, men på vores 3 primære renseanlæg har vi krav om fosforrensning ned til 0,5 mg/l i udløbet. Det kan vi kun overholde ved ekstraordinært at rense spildevandet i et kontaktfiler, og det betyder naturligvis ekstraordinære omkostninger i forhold til et "normalt" renseanlæg. Vi har opgjort de ekstraordinære omkostninger til fosforrensning til ca. 3,2 mio. kr. årligt i 2010-priser. Sat i forhold til vores netvolumenbeløb på 57,1 mio. kr. på de primære renseanlæg udgør den ekstra fosforrensning altså 6 %".

Beregningen understøtter, at fosfor i udløb er et betydende forhold, der bør tages hensyn til. En mulig årsag til, at fosfor ikke bliver signifikant kan være datakvaliteten. I rådatasættet, er der mange observationer, der er urealistisk høje, hvilket tyder på mange deciderede fejlindberetninger. Endvidere bemærkes det, at i følge opgørelsen lever kun 60 % af renseanlæggene op til det nationale minimumskrav om en maksimal udledning på 1,5 mg/l, hvilket ikke kan passe. Der er ikke enheder i rådatasættet, så ovenstående kan bero på en fejlfortolkning af data fra DANVAs side, men DANVA opfordrer Forsyningssekretariatet til at kigge nærmere på dette forhold.

Et andet forhold, der ikke indgår i modellen er kapacitetsudnyttelsen. De marginale omkostninger ved at rense en PE stiger når belastningsgraden på anlægget falder. Det koster altså at have ledig kapacitet på anlæggene. Ledig kapacitet kan opstå, hvis belastningen ikke stiger så hurtigt i området som man forudsatte, da man byggede anlægget. Store industrier, der lukker, kan også give ledig kapacitet, ligesom nogle industrier har en stærkt varierende udledning over året, eksempelvis fiskeindustrien, og derfor kræver en stor ledig kapacitet. Et andet eksempel er sommerhusområder, der kræver en stor kapacitet et par måneder om året. De fleste renseanlæg er desuden dimensioneret med en vis overkapacitet, så man ikke skal bygge et nyt anlæg, hvis der kommer et par nye industrier i oplandet. I den nuværende model, indgår kapacitet ikke pga. korrelation, hvilket er meget uheldigt. Et renseanlæg har mange faste udgifter, og hvis selskabet af den ene eller anden grund er nødt til at stille med et stort beredskab, eksempelvis pga. sommerhusområder, så bliver selskabet ikke kompenseret. Selskaber, der af den ene

eller anden grund, har et generelt overdimensioneret renseanlæg bliver ligeledes ikke kompenseret.

De ovenfor nævnte forhold: slamdisponering, kapacitet og fosforkrav synes alle at være tæt på signifikans i forskellige modeller, men hvor det er svært at se en samlet model, der på tilfredsstillende vis kan forklare forholdene. Dette kan dog ses som en klar indikation på, at de nævnte faktorer alle er forhold man bør tage højde for i fremtidige modeller.

Modellen for renseanlæg er meget forsimplet. Der er en hel række forhold, som selskaberne har konkrete erfaringer med påvirker omkostningerne, men som ikke fanges af modellen. Muligvis er datakvaliteten for dårlig, men DANVA opfordrer Forsyningssekretariatet til at undersøge, om der kan være andre supplerende forklaringer, der muliggør en bedre model allerede med de nuværende data. Alternativt må modellen fortolkes meget varsomt.

Kunder

Problemstillingen beskrevet under vand nedenfor, gælder også for spildevand.

Specifikke kommentarer til omkostningsækvivalenterne på vand:

Råvandsledninger

Drift af råvandsledninger minder om drift af rentvandsledninger. Det virker derfor mere naturligt at slå råvandsledning sammen med rentvandsledninger fremfor at lægge råvandsledninger ind under vandværker. Rent konteringsmæssigt er det nemmere at styre på den måde, da det er de samme folk, der vedligeholder råvandsledninger og rentvandsledninger. FVD og DANVA har ikke regnet på konsekvenserne af dette, men det virker mere logisk.

I det hele taget kan man overveje, hvor godt selskaberne rent regnskabsmæssigt kan adskille råvandsledninger, rentvandsledninger og stik. FVD og DANVA formoder, at der ligger et betydeligt antal skøn bag denne fordeling og at der derfor kan være en del variation mellem råvandsledninger, rentvandsledninger og stik, der ikke fanges i modellen på nuværende tidspunkt. Man kunne overveje at slå alle tre kategorier sammen og acceptere, at modellen derved bliver meget simpel. Dette ville dog være meget uheldigt for de selskaber, der har mange stik pr. km rentvandsledning. Kort sagt er FVD og DANVA meget betænkelige ved datakvaliteten på dette område, men omvendt er stik en væ-

sentlig faktor, hvilket taler for man beholder den nuværende model og er varsom i fortolkningen af den. Og at der hurtigt findes et bedre grundlag for benchmarking på dette område.

Vandværk

Modellen er meget simpel og tager ikke hensyn til bl.a. overkapacitet. Mange vandværker opretholder en overkapacitet, således der er forsyningssikkerhed, hvis et værk skal repareres. Mange selskaber har også ledig kapacitet andre selskaber skal kunne trække på i en nødsituation (dette er i mange tilfælde fastlagt i den kommunale vandforsyningsplan). Det er dyrt at have ledig kapacitet, hvorfor det burde være med i modellen. Forsyningssekretariatet finder, at der er stordriftsulemper ved store vandværker, hvilket intuitivt ikke er åbenlyst. Det kunne indikere, at modellen lider af udeladte variable, der er korreleret med størrelse. Bud på udeladte variable kunne være længde af råvandsledninger, overkapacitet og servicemål på drikkevands- og forsyningssikkerhed.

Rentvandsledning

Der kan anføres mange af de samme forhold for rentvandsledning som gælder for ledning på spildevand. Materiale, alder og dimension er alle vigtige variable på rentvandsledning, og ligesom på spildevand giver det store problemer, at de ikke er inkluderet.

Igen ser man, at estimatet for land er mindre end estimatet for by, og at land næsten er signifikant. Og igen forkastes modellen i tabel 6. Desuden er der problemer med størrelsen på estimatet for indre city, som man må forvente er større end estimatet for city. Da estimatet på indre city er signifikant, og land i øvrigt er næsten signifikant, understreger det, at der formentlig er problemer med dels den underliggende datakvalitet og dels at der er problemer med udeladte variable. Det kunne overvejes at ekskludere indre city fra modellen og beregne den separat som man gør på stik, og se om det hjælper på modellen. Beholder man modellen i tabel 7 ender man op med en stærkt forsimplet model, hvilket man bør være opmærksom på i anvendelsen af resultaterne.

Stik

Forsyningssekretariatet forkaster modellen i tabel 8, da man ikke vil acceptere en u-formet kurve, hvor der er større omkostninger ved at drive et stik i landzone. Det kan dog ikke afvises, at det er dyrere at drive et stik i landzone, da transportomkostningerne er højere, stikkene potentielt er længere og de går evt. gennem gårdspladser og lignende. Det er dog vanskeligt for FVD og DANVA at vurdere om kurven for stik rent faktisk er u-formet. Det kan ikke afvises, men omvendt heller ikke bekræftes. Og de underlig-

gende data på stik er ret upræcise, som det er beskrevet under "råvandsledning" ovenfor. Derfor er det en vanskelig at vurdere om tabel 8 eller 9 er den korrekte model.

Et andet forhold omkring stik er, at nogle selskaber ikke kun vedligeholder selve stikket, men også vedligeholder jordledningen koblet til stikket. Selskaber der har denne praksis får ikke beskrevet deres omkostninger fuldt ud i modellen.

Bemærk i øvrigt, at der er en regnefejl i tabel 10. Her bør man helt klart tage et vægtet gennemsnit i stedet for et simpelt gennemsnit. Den gennemsnitlige omkostning pr. stik bliver dermed 2044 kr. i stedet for 1398 kr.

Kunder

Forsyningssekretariatet skal være opmærksom på, at der er stor forskel på hvor mange husstande, der er tilknyttet en måler, og at omkostningerne til kunder ikke kun drives af antal målere, men også af antal husstande. Den foreslåede model giver især problemer for selskaber, der forsyner områder med mange lejlighedskomplekser, da disse ofte kun har en måler til mange husstande. De selskaber, der har mange husstande pr. måler bliver således underkompenseret i modellen. Endvidere har selskaber med mange erhvervs-kunder, herunder især ved salg af vand til andre forsyninger, en væsentlig højere udgift på disse mere specielle kunder. Dette forhold kompenseres heller ikke i modellen.

Angående valg af omkostningsfunktion, er det uklart om en lineær model beskriver data bedst eller om man eventuelt skulle overveje en logaritmisk model.

Venlig hilsen



Direktør Bent Soelberg
FVD



f/ Direktør Carl-Emil Larsen
DANVA

Bilag 1
Pumper

Der udføres en regressionanalyse på
 $Y=b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5$
hvor

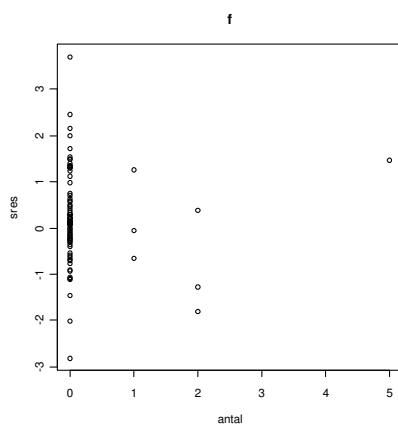
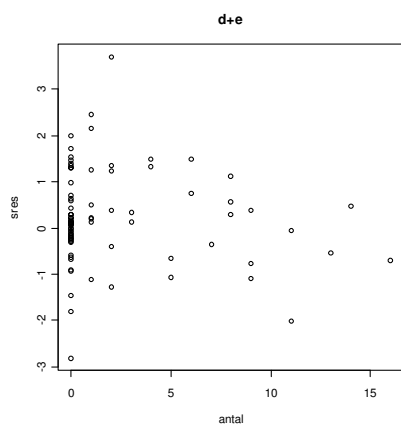
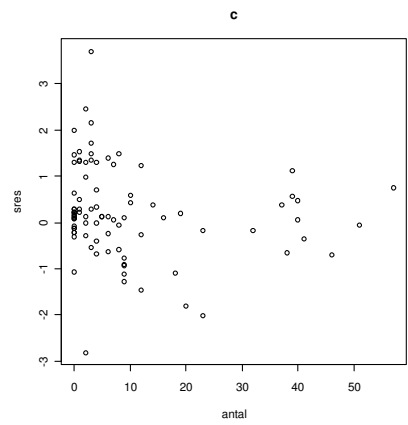
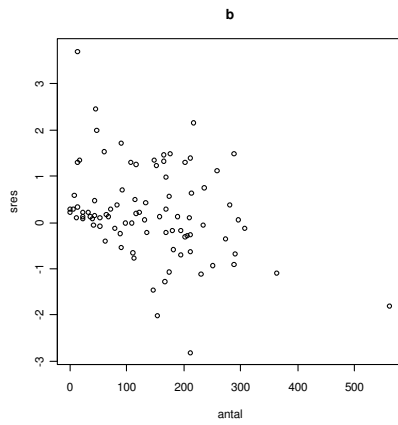
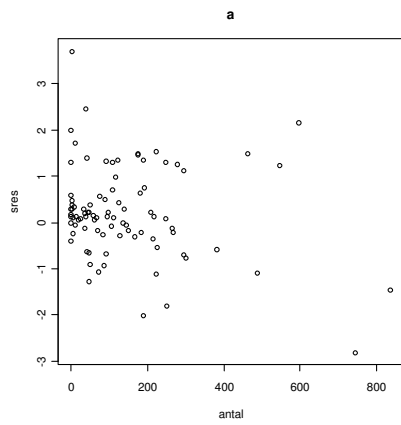
x_1 = kategori a, x_2 = kategori b, x_3 = kategori c, x_4 = kategori d+e, x_5 = kategori f
Kategori a - f henviser til kategorierne i høringsudkastet.

Ved et outlierkriterium på Cook's distance > 1 fjernes tre virksomheder (26, 64 og 91),
og der fås følgende regressionsanalyse:

```
Residuals:
      Min       1Q   Median       3Q      Max
-6181156  -610479   319562  1550027  8998823

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
x1          8274      1742    4.749 8.46e-06 ***
x2         16586      2654    6.250 1.69e-08 ***
x3         49282     26236    1.878  0.0638 .
x4        177366     87512    2.027  0.0459 *
x5        787760     401127    1.964  0.0529 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2435000 on 83 degrees of freedom
Multiple R-squared:  0.8384,    Adjusted R-squared:  0.8286
F-statistic: 86.11 on 5 and 83 DF,  p-value: < 2.2e-16
```

Hvilket giver modellen: $Y=8274x_1 + 16586x_2 + 49282x_3 + 177366(x_4 + x_5) + 787760x_6$

Bilag 2

Åbne bassiner

Derudføres en regressionanalyse på

$$Y = b_1x_1 + b_2x_2$$

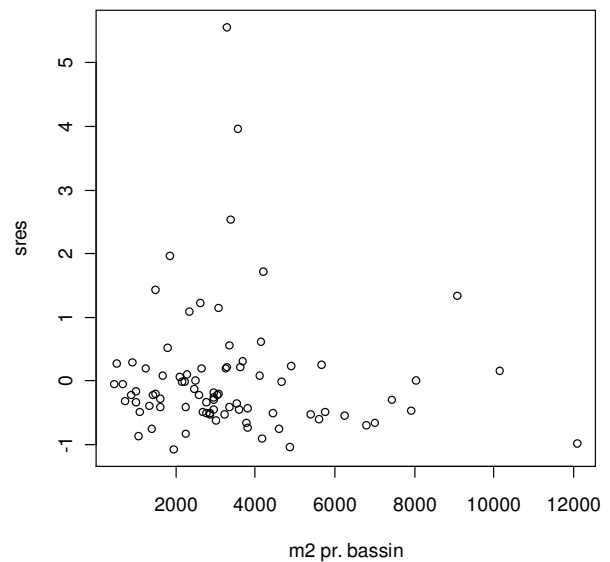
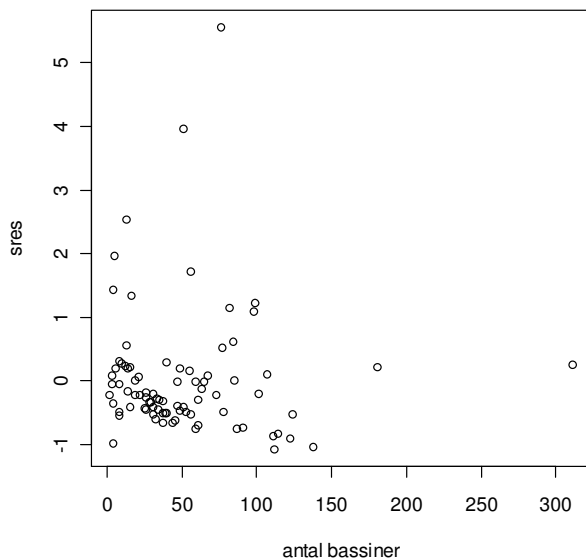
hvor x_1 = antal bassiner og x_2 = m^2 pr. bassin

Ved et outlierkriterium på Cook's distance > 1 fjernes en virksomhed (83), og der fås følgende regressionsanalyse:

```
Residuals:
    Min       1Q   Median       3Q      Max
-1173504 -552960 -246415  230403  6185155

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
x1     99.99      40.25   2.484  0.0150 *
x2  9717.53   2341.76   4.150  8.09e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1122000 on 82 degrees of freedom
Multiple R-squared:  0.4385,    Adjusted R-squared:  0.4248
F-statistic: 32.02 on 2 and 82 DF,  p-value: 5.284e-11
```



Hvilket giver modellen: $Y=99,99x_1 + 9717,53x_2$

Bilag 3

Metoder udover Cooks afstand brugt i reguleringen af gas og el i Tyskland (distributionselskaber):

DFBETAs: Forskellen mellem regressionskoefficienterne beregnet på baggrund af alle data og koefficienterne, når en enhed fjernes og skaleret med standardafvigelsen. Cut-off for DFBETA er normalt $2/\sqrt{n}$, hvor n som før er antal observationer.

DFFITS: Måler på de ændringer i den forudsagte værdi som følger af at eliminere en observation. Cut-off er normalt $2\sqrt{p/n}$

Covariance ratio: Måler på ændringen i determinanten af covariansmatricen ved eliminering af en enhed. Cut-off sker normalt når den afviger fra 1 med mere end $3n/p$

Læs evt. mere herom i nedenstående publikation af Agrell and Bogetoft (2008).

Referencer nævnt i teksten:

Agrell, P.J., and P.Bogetoft, PROJECT GERNER IV, Electricity and Gas DSO Benchmarking, WHITE PAPER, BNetzA, 2008.

Belsley, D. A., E. Kuh and R. E. Welsch (1980). *Regression Diagnostics*. New York: Wiley.