

Bilag 1: Beregning af omkostningsækvivalenter

Bilaget indeholder den tekniske beregning af omkostningsækvivalenterne til brug for benchmarkingen 2013.

FORSYNINGSSSEKRETARIATET FEBRUAR 2013

INDLEDNING	3
MINDSTE KVADRATERS METODE.....	3
Sammenhæng mellem omkostninger og costdriver	5
Normalfordeling af fejllid.....	6
Multikollinearitet.....	7
Ekstreme observationer og observationer med stor indflydelse	8
Transformationer	9
Valg af den bedste model	10
Datakvalitet	11
BESTEMMELSE AF OMKOSTNINGSÆKVIVALENTER FOR VANDFORSYNINGER ...	13
Boringer	13
Vandværker	16
Trykforøgerstationer	19
Rentvandsledninger	23
Stik	26
Kunder	29
BESTEMMELSE AF OMKOSTNINGSÆKVIVALENTER FOR SPILDEVAND	32
Ledninger	32
Pumpestationer	35
Regnvandsbassiner	39
Spildevandsbassiner	41
Minirenselanlæg.....	43
Renselanlæg.....	46
Slambehandling	53
Kunder	56

Indledning

Dette bilag har en relativ teknisk karakter og er primært en gennemgang af den statistiske metode, der ligger til grund for Forsyningssekretariatets beregning af omkostningsækvivalenterne til brug for benchmarkingen af vand- og spildevandsforsyningerne.

Omkostningsækvivalenterne udtrykker sammenhængen mellem forsyningernes driftsomkostninger, og de underliggende forhold der bestemmer, hvor store driftsomkostningerne er.

Det er vigtigt, at resultaterne af benchmarkingen er så præcise som mulig. Ulempen ved at sigte mod størst mulig præcision er dog, at de metoder der anvendes, kan være ret komplicerede. Forsyningssekretariatet har vurderet, at præcision i denne sammenhæng vægter højest.

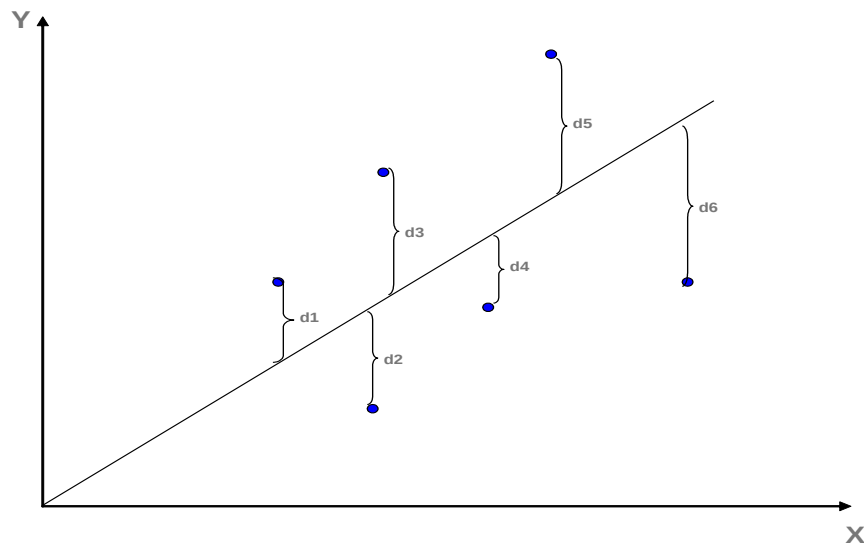
I det følgende beskrives først den metode, som Forsyningssekretariatet har brugt til at beregne omkostningsækvivalenterne. Dernæst er der en detaljeret gennemgang af beregningen af de seks omkostningsækvivalenter for vandforsyningerne og de otte omkostningsækvivalenter for spildevandsforsyningerne.

Forsyningssekretariatet har bedt samtlige selskaber indberette en række oplysninger om deres costdrivere. Selskaberne er således blevet bedt om at fordele deres driftsomkostninger ud på hver enkel costdriver samt angive de underliggende forhold, der er identificeret for hver costdriver, fx antal stik i byzone. Hvert selskabs oplysning indenfor de enkelte costdrivere betegnes som en observation.

Mindste kvadraters metode

For at finde en generel sammenhæng imellem omkostningerne og de underliggende forhold - kaldet omkostningsækvivalent - har Forsyningssekretariatet benyttet regressionsanalyse. I regressionsanalysen bliver der benyttet mindste kvadraters metode. Metoden består i at fastlægge en linje (sammenhæng), der minimerer summen af kvadratet på den lodrette afstand til linjen, jf. figur 1.

Figur 1: Mindste kvadraters metode



Figur 1 illustrerer seks forsyningers indberetning af driftsomkostninger (afsat på Y-aksen) og en tilhørende underliggende faktor, for eksempel udpumpet vandmængde (afsat på X-aksen). De er illustreret med de runde punkter. Mindste kvadraters metode placerer linjen i koordinatsystemet i figur 1, hvor summen af de kvadrerede afstande angivet med d1-d6 er så lille som mulig. Dette bevirker, at jo mere en observation adskiller sig fra resten af observationerne, jo mindre vægtning får denne observation. Mindste kvadraters metode giver således et mere sandsynligt bud på den korrekte sammenhæng i data, end et simpelt gennemsnit ville gøre.

En forudsætning for at benytte mindste kvadraters metode er, at sammenhængen imellem responsvariablen (Y-variablen) og den forklarende variabel (X-variablen) er lineær. Hvis den faktiske sammenhæng ikke er lineær, er det nødvendigt at foretage forskellige transformationer af data for at vurdere den faktiske sammenhæng imellem omkostningerne og de underliggende forhold, jf. afsnittet nedenfor om transformationer.

Yderligere er det en forudsætning for anvendelse af mindste kvadraters metode, at de forskelle (fejllid), der er mellem modellens forudsagte værdier af omkostningerne og den enkelte observations faktiske værdi, er normalfordelte. De forudsagte værdier ligger på linjen i koordinatsystemet i figur 1. En forudsætning, for at fejleddene er normalfordelte, er, at de har ens spredning omkring en middelværdi på 0, jf. afsnittet nedenfor om normalfordeling af fejllid.

Spredningen af en gruppe observationer udtrykker, hvor stor forskel der er på selskabernes indberetninger af omkostninger og costdrivere, når der tages hensyn til forskellen i størrelsen af de enkelte selskaber

Den generelle model, som Forsyningssekretariatet vil opstille for hver enkelt omkostningsækvivalent, er af formen:

$$(1) \quad Y = B_0 + B_1X_1 + \dots + B_NX_N$$

Ved hjælp af mindste kvadraters metode estimeres således den værdi af B'erne, der angiver den mest sandsynlige sammenhæng mellem Y og samtlige X'er.

De enkelte B-værdier angiver hældningen af den tendenslinje, der fastsættes i regressionsanalysen. Det vil sige, at en given B-værdi angiver ændringen i driftsomkostningerne (Y) ved en ændring på 1 i den forklarende variabel (X).

Sammenhæng mellem omkostninger og costdriver

For at sikre at modellen kan bruges, kontrolleres der for, om der er en god sammenhæng mellem omkostningerne og de underliggende forhold for hver costdriver. Der kontrolleres således for, om der er signifikans i modellen. Signifikans i en regressionsanalyse betyder, at det ikke kan afvises, at der er en sammenhæng mellem variabelen (Y) og de pågældende forklarende variable (X).

Forsyningssekretariatet har valgt at tage udgangspunkt i et standard 5 pct. signifikansniveau. Det vil sige, at hvis sandsynligheden er større end 5 pct. for at et givet B er lig med 0, vil det blive afvist, at der er en sammenhæng mellem den forklarende variabel (X) og responsvariabelen (Y). Det betyder, at det i denne situation afvises, at den forklarende variabel er med til at beskrive de samlede omkostninger forbundet med en costdriver.

Signifikansen beregnes på baggrund af antallet af observationer, den beregnede B-værdi og spredningen af observationerne omkring det pågældende B. Hvis spredningen er høj i forhold til størrelsen af B, eller der er få observationer, svækker det signifikansen.

Signifikansen af en variabel angives med både en t-værdi og en p-værdi. Både t-værdien og p-værdien beskriver, hvad sandsynligheden er for, at B er lig med 0. Selve t-værdien skal ligge udenfor intervallet $\pm 1,96$ for, at det kan afvises, at B er lig med 0. t-værdien kan desuden omregnes til en p-værdi. p-værdien angiver sandsynligheden for, at B er lig med 0 og skal således være mindre end 0,05 (5 pct.) for, at det kan accepteres, at den pågældende variabel indgår i modellen.

t- og p-værdierne fremgår af resultattabellerne i gennemgangen af hver enkel omkostningsækvivalent.

Normalfordeling af fejllid

Mindste kvadraters metode forudsætter, at fejllidene er normalfordelte med en middelværdi på nul.

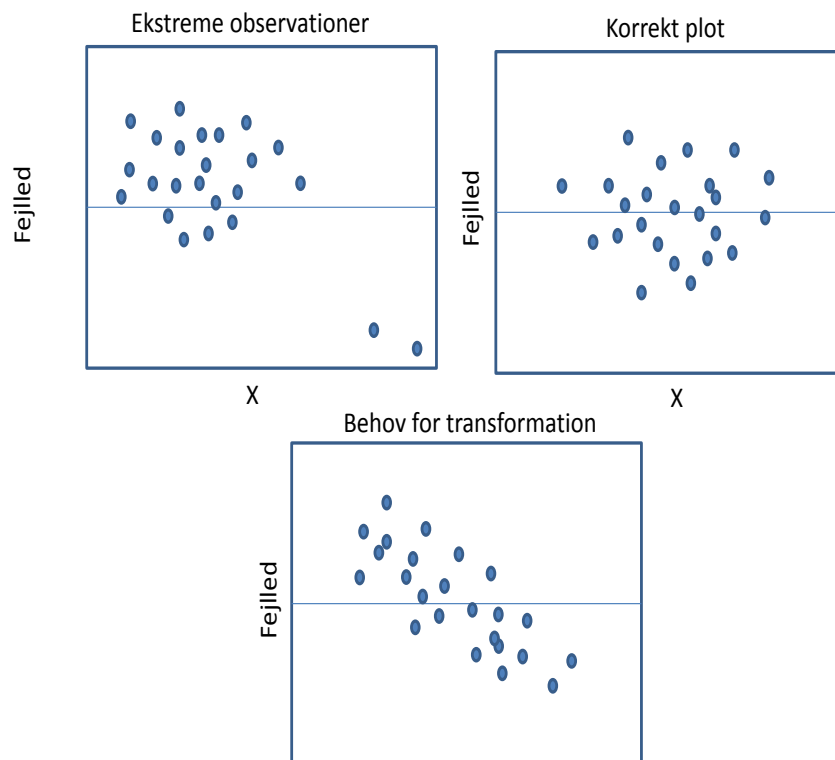
Det kan kontrolleres, om fejllidene er normalfordelte ved at plote fejllidene mod de observerede værdier af de enkelte X 'er. I praksis benyttes et mål, der kaldes de standardiserede fejllid, som i sidste ende udtrykker det samme som de faktiske fejllid, men derudover har en praktisk fordel med hensyn til kontrol af spredningen. Det skyldes, at fejllidene standardiseres i forhold til en normalfordeling. Standardiseringen indebærer derfor at 95 pct. af observationernes standardiserede fejllid skal ligge inden for spændet $[-1,96; 1,96]$.

For at betingelsen om normalfordelte fejllid er opfyldt, skal fejllidsploetterne vise en tilfældig fordeling omkring 0, jf. figur 2. Hvis antagelsen om normalfordeling af fejllidene ikke er opfyldt, kan det være tegn på, at modellen ikke er specificeret korrekt eller, at der er enkelte ekstreme observationer, der skævvrider resultaterne. Dette er illustreret i figur 2. Specificering af de enkelte modeller og håndteringen af ekstreme observationer er forklaret i de senere afsnit.

Yderligere kan der være generel usikkerhed omkring størrelsen af signifikansen for de enkelte B -værdier, hvis fejllidene danner en trompetform eller andre sammenhænge, der ikke umiddelbart ser tilfældige ud. Den beregnede spredning er i denne situation ikke udtryk for den faktiske spredning i hele datasættet, og der kan derfor opstå usikkerhed om hvorvidt sammenhængen mellem Y og X er signifikant. Forsyningssekretariatet har derfor beregnet robuste t værdier for en række af costdriverne for at sikre, at signifikansen holder. De robuste t -værdier tager hensyn til, at spredningen muligvis ser anderledes ud og kan både betyde, at signifikansen øges eller sænkes.

Det undersøges, om betingelsen omkring normalfordelte fejllid er opfyldt for hver af de beregnede omkostningsækvivalenter.

Figur 2: Fordeling af fejll



Note: Den horisontale linje i ovenstående figurer er lig med 0.

Multikollinearitet

En model, der skal beskrive en omkostningsfunktion for en given costdriver, kan i mange tilfælde bestå af flere variable (X_1 , X_2 , X_3 osv.) Eksempelvis vil omkostningsfunktionen for produktionen af en bil bestå af to variable - prisen på halvfabrikata og timeantallet til samling af bilen.

I modellem med flere variable kan der opstå problemer, hvis de forklarende variable (X 'erne), der indgår i modellen, indbyrdes er korrelerede. For eksempel er X_1 og X_2 korrelerede, hvis X_2 stiger, når X_1 stiger. Det betyder, at X_1 og X_2 beskriver den samme variation for Y . Dette problem betegnes multikollinearitet og kan give forkerte beregninger af B-værdierne og have betydning for beregningen af signifikans og B-værdiernes fortegn.

Det kan være svært at vurdere, hvornår multikollinearitet kan være et problem. Normalt vurderes en korrelation på over 0,7 mellem to parametre at være høj. Forsyningssekretariatet vil dog primært lægge vægt på, at modellerne giver intuitive resultater. Derfor vil det især blive undersøgt nærmere, hvis der forekommer sammenhænge, der ikke giver mening, eller hvis der forekommer parametre, som burde være signifikante, men som er insignifikante.

Ekstreme observationer og observationer med stor indflydelse

Enkelte observationer i et datasæt kan have en u hensigtsmæssig stor påvirkning af bestemmelsen af B-værdierne. Det er derfor nødvendigt at overveje, hvordan sådanne observationer skal behandles.

Forsyningssekretariatet har vurderet, at Cook's afstand er et passende mål til at identificere observationer med stor indflydelse på analysen. Metoden med Cooks d er i overensstemmelse med statistisk praksis og anvendes også som standard i det statistikprogram, der benyttes til regressionskørslerne¹.

Cook's afstand siger noget om, hvor meget B-værdierne ændres, hvis en given observation fjernes fra analysen. Hver eneste observation får således beregnet en Cook's afstand.

Der er ikke nogen faste statistiske regler for, hvornår en observations Cook's afstand er så stor, at observationen bør fjernes fra datasættet. Forsyningssekretariatet anvender en metode, der forslår, at observationer, som falder udenfor 50 pct.-fraktilen i en F-fordeling med $(p, n-p)^2$ frihedsgrader, bør fjernes fra datasættet. I en model med en forklarende variabel svarer dette til en Cook's afstand på ca. 0,5. I takt med at antallet af forklarende variable øges, bliver kravet til Cooks afstand sænket. Forsyningssekretariatet vil af forsigtighedshensyn holde fast i en Cooks afstand på 0,5.

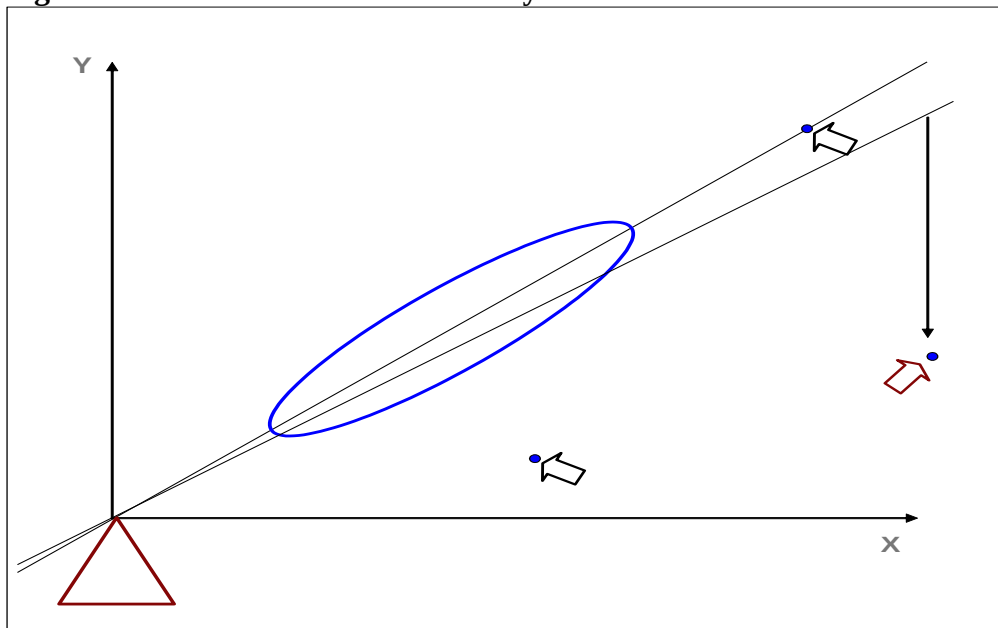
Figur 4 illustrerer, hvordan ekstreme observationer kan have indflydelse på modellens resultater. De ekstreme observationer er de observationer, der ikke ligger inden i den blå oval i figur 4 nedenfor. Figuren illustrerer, hvordan punktet ved den højrevendte pil kan påvirke linjen u hensigtsmæssigt meget, da den vil trække tendenslinien ned mod sig. Det skyldes, at en observation kan påvirke regressionsanalysen på samme måde som en vægtstang og dermed få stor betydning for resultaterne af regressionsanalysen.

Samtidig kan det også være tilfældet, at andre punkter der måske skiller sig en del ud fra mængden, ikke har særlig stor indflydelse på resultaterne, jf. de to punkter i figur 4 med venstrevendte pile. I forhold til vægtstangen i nulpunktet (ved den røde trekant) vil de ikke påvirke tendenslinien i nævneværdig grad, og punkterne har derfor ikke nogen væsentlig betydning for analysens resultater.

¹ Jf. standardbetingelser for Cook's D i statistikprogrammet "R", samt Gross, Jürgen, Linear Regression, Springer 2003.

² Her er n antal observationer og p er antal parametre.

Figur 3: Observationer med stor indflydelse



Ved beregningen af hver enkel omkostningsækvivalent kontrolleres ved hjælp af Cook's afstand for, om der er ekstreme observationer, der har en u hensigtsmæssig indflydelse på resultaterne og derfor bør fjernes fra datasættet.

Transformationer

Hvis det viser sig, at data ikke er lineært, kan problemet løses ved hjælp af forskellige transformationer af data. De mest udbredte transformationer er at tage logaritmen, kvadratroden eller andre potenser til variablene i ligningen.

Problemet med andre sammenhænge i data end lineære er, at mindste kvadraters metode kun kan bruges på et lineært datasæt. Derfor er det nødvendigt at vælge en transformation, der skaber en lineær sammenhæng i forhold til den forventede sammenhæng i det oprindelige datasæt.

Et eksempel kan være, at det forventes, at en produktion kan beskrives ved en klassisk Cobb-Douglas funktion af formen:

$$(2) \quad Y = B_0 X_1^{B_1} X_2^{B_2}$$

I dette tilfælde er det nødvendigt at lave en såkaldt "log-log"-transformation. Det vil sige, at logaritmen tages på begge sider af lighedstegnet. Det giver følgende lineære form:

$$(3) \quad \log Y = \log B_0 + B_1 \log X_1 + B_2 \log X_2$$

I denne form er det muligt at estimere B-værdierne ved hjælp af mindste kvadraters metode.

Årsagen til at den logaritmiske transformation er populær skyldes, at valget af omkostningsfunktion bliver bestemt af data. I den logaritmiske model vil størrelsen af B-værdien afgøre, hvorvidt der er tale om stordriftsfordele eller stordriftsulempen.

For at sikre, at ”log-log”-transformationen er den rigtige transformation, kan fejlledsplottene kontrolleres. Det vil sige, at modelkontrollen afslører, hvorvidt der er tale om log-normalfordelt data. Dette kan ses ved, at fejlleddene i den transformerede model er normalfordelte.

Da hele formålet med beregningen af omkostningsækvivalenterne drejer sig om at bestemme forholdet mellem Y og X’erne, er det nødvendigt at tilbagetransformere ligning (3). Tilbagetransformeringen er ikke så enkel, som det umiddelbart kunne se ud til. Det skyldes, at hvis ”log-log”-transformationen giver en normalfordeling af fejlleddene, så kan det ikke være tilfældet i parametrene oprindelige form. Det giver en skævhed, når den estimerede model tilbagetransformeres, og det er derfor nødvendigt at benytte en korrektionsfaktor, der justerer for dette. Korrektionsfaktoren kan bestemmes til³:

$$(4) \quad KF = \exp(\sigma^2/2)$$

Her betegner ”exp” den naturlige eksponentialfunktion og σ er spredningen af fejlleddet. Korrektionsfaktoren ganges med hele højre siden af den aktuelle model.

Hvis man i stedet foretager en transformation af data, således at variablene Y og X f.eks. er opløftet i en potens kan korrektionsfaktoren bestemmes til:

$$(5) \quad KF = \sigma^2$$

Her betegner σ spredningen af fejlleddet. Korrektionsfaktoren lægges til på højre side af den aktuelle model.

Valg af den bedste model

For flere af costdriverne er der indsamlet information, som giver mulighed for at konstruere forskellige former af de enkelte omkostningsækvivalenter. Dette betyder, at der ofte er flere modeller, som potentielt kan blive lagt til grund for beregning af netvolumenmålet. Udvalgelsen af den bedste model kræver en konkret vurdering af de enkelte modeller.

Samlet set vurderer Forsyningssekretariatet nedenstående tre kriterier:

³ Kilde: Don M. Miller, *Reducing Transformation Bias in Curve Fitting*.

- Intuitiv og teoretisk baggrund.
- Statistiske kriterier.
- Antallet af forsyninger.

1) Der er **intuitiv og teoretisk baggrund** for at antage den sammenhæng, som undersøges: Dette har i flere tilfælde ført til transformationer af data. Valget af den bedste model vil derfor i første omgang afhænge af, om der er en intuitiv mening med modellen. Størrelsen af de estimerede B-værdier tages også i betragtning. Det giver for eksempel ikke mening at have en model for trykforøgere, hvor de store pumper er billigere end de små pumper.

2) Der er en bedre opfyldelse af de **statistiske kriterier** vurderet på baggrund af:

Højere forklaringsgrad i modellen målt ved den statistiske værdi, " R^2 ". R^2 kan bruges til at vurdere, hvor godt modellen samlet set beskriver de driftsomkostninger, som forsyningerne har fordelt på de enkelte costdrivere. Herudover anvendes resultatet af F-testet til at angive sandsynligheden for, at der er sammenhæng i hele modellen.

Yderligere lægges der vægt på signifikansen af de enkelte koefficienter og de tilhørende fejlsplots, jf. ovenfor.

I nogle tilfælde har Forsyningssekretariatet sammenlignet resultaterne af modellernes forudsagte værdier overfor de faktiske omkostninger. Dette bruges som et supplement til vurderingen af de statistiske kriterier

3) Endelig har Forsyningssekretariatet også lagt vægt på, **Antallet af Forsyninger der** indgår i beregningen så der ikke fjernes for mange observationer i beregningen af de enkelte omkostningsækvivalenter på baggrund af størrelsen af Cook's afstand.

Datakvalitet

For at sikre de mest retvisende resultater af regressionsanalyserne har det været nødvendigt at foretage en kvalitetssikring af det datasæt, som Forsyningssekretariatet har modtaget.

Forsyningssekretariatet har valgt at fjerne forsyninger helt fra datasættet, hvis der har været store mangler i den enkelte indberetning.

Forsyningssekretariatet har yderligere foretaget en grundig kvalitetssikring af data. Forsyningssekretariatet har blandt andet kontaktet mange forsyninger for at få en præcis redegørelse for deres indberetninger, når der har været uoverensstemmelser i indberetningerne eller andre forhold, som krævede en gennemgang.

Herudover har der for enkelte af de underliggende forhold været så stor tvivl om datakvaliteten, at det har været nødvendigt at udelukke disse i beregningerne.

Et af de primære formål med genberegningen af omkostningsækvivalenterne er at forøge datakvaliteten. I den forbindelse skal det bemærkes, at ca. 80 % af de totale omkostninger fordeler sig på enten direkte eller funktionshenførbare omkostninger, hvilket indikerer en rimelig sikker omkostningsfordeling jf. tabel 1 og 2.

Tabel 1: Omkostningsfordeling for drikkevandsforsyninger

Direkte henførbare driftsomkostninger	Funktionshenførbare driftsomkostninger	Øvrige driftsomkostninger
0,44	0,34	0,22

Tabel 2: Omkostningsfordeling for spildevandsforsyninger

Direkte henførbare driftsomkostninger	Funktionshenførbare driftsomkostninger	Øvrige driftsomkostninger
0,60	0,23	0,17

Herudover er det Forsyningssekretariatets vurdering, at det er forsyningerne selv, der bedst ved, hvordan den skønsmæssige fordeling af de øvrige omkostninger i den enkelte forsyning bør foretages. Derfor er det ikke umiddelbart til skade for datakvaliteten, at skønnene er fordelt på baggrund af forskellige principper. Dog kan der naturligvis være fejl i et datasæt som dette, som eksempelvis tastefejl og fejlskøn.

Den regressionsanalyse, som Forsyningssekretariatet benytter, medfører at sådanne fejl kun får begrænset betydning. Analysen tager således hensyn til den slags datafejl under en forholdsvis ustringent antagelse om, at disse fejl er tilnærmelsesvis normalfordelte omkring 0. Derudover indebærer regressionsanalysen, at observationer der adskiller sig meget fra resten af observationerne, får mindre vægt i beregningen af B-værdierne.

Endelig har Forsyningssekretariatet foretaget en række følsomhedsanalyser af omkostningsækvivalenterne for at teste deres robusthed og herunder også datakvaliteten, samt gennemført en række ekstra analyser af skævvridninger i netvolumenmålet jf. bilag 3.

Bestemmelse af omkostningsækvivalenter for vandforsyninger

Forsyningssekretariatet har beregnet omkostningsækvivalenter for de seks costdrivere:

- Boringer
- Vandværker
- Trykforøgerstationer
- Rentvandsledninger
- Stik
- Kunder

Beregningen af de enkelte omkostningsækvivalenter bliver beskrevet nedenfor. Først beskrives den model, som Forsyningssekretariatet efter en række analyser, har vurderet, er den mest velegnede model til at beskrive driftsomkostningerne for den enkelte costdriver. Herefter beskrives de forudgående analyser, der har ført til valget af modellen.

Boringer

Forsyningssekretariatet har vurderet, at den mest velegnede model, til at beskrive de samlede driftsomkostninger ved at drive boringer, er en eksponentiel sammenhæng af formen $Y = X_1^{B_1}$.

For at kunne behandle eksponentielle sammenhænge med *mindste kvadraters metode* er det nødvendigt, at transformere omkostningsfunktionen som herefter bliver en log-log-sammenhæng. Den mest velegnede model fremgår af ligning (6), hvor Y angiver de samlede driftsomkostninger ved at drive boringer og X_1 er den totale oppumpede vandmængde:

$$(6) \log Y = B_1 \log X_1$$

I funktionen ovenfor indgår parameteren B_1 . B_1 siger noget om skalatendensen ved at drive boringer. Der kan således være stordriftsfordele (B -værdi < 1) eller stordriftsulemper (B -værdi > 1) for boringer.

I tabel 3 nedenfor præsenteres resultaterne af modellen i ligning (6).

Tabel 3: Regressionsanalysens resultater, boringer

Variabel	B-værdi	Spredning	t-værdi	Pr(> t)
Log (oppumpet vandmængde)	0.932065	0.005635	165.4	<2e-16
Antal observationer: 145 Justeret $R^2 = 0,99$				

Omkostningsækvivalenten for boringer

Idet ligning (6) er en såkaldt log-log-funktion, er det nødvendigt at foretage en afsluttende korrektion, når modellen tilbagetransformeres. Tilbagetransformationen er nødvendig for, at omkostningsækvivalenten kan anvendes til at beregne netvolumenbidraget fra boringer, jf. side 10.

Den endelige omkostningsækvivalent for boringer består af den oppumpede vandmængde (X_1) og en ekstra korrektionsfaktor, jf. afsnittet vedr. transformationer ovenfor. Korrektionsfaktoren beregnes som $KF = \exp(\sigma^2/2) = 1,518$. Omkostningsækvivalenten ved at drive boringer har dermed følgende udseende, jf. tabel 3:

$$(7) \quad Y = 1,518X_1^{0,9321}$$

Modellen angiver dermed, at de årlige driftsomkostninger ved at drive boringer kan beskrives ved, at det koster 1,518 kr. per m^3 vand opløftet i en potens på 0,9321 om året at pumpe en m^3 vand op af en boring.

Kvalitetskontrol af omkostningsækvivalenten

Det vurderes at resultaterne i tabel 3 er tilfredsstillende, idet der er overensstemmelse med hvad der blev forventet. Den positive værdi af B_1 betyder således, at driftsomkostningerne stiger med den oppumpede vandmængde. Det er yderligere rimeligt at forvente en lille skalafordel i en meget produktionsorienteret del af forsyningsnettet. Desuden ses det af tabel 3, at koefficienten i modellen er signifikant.

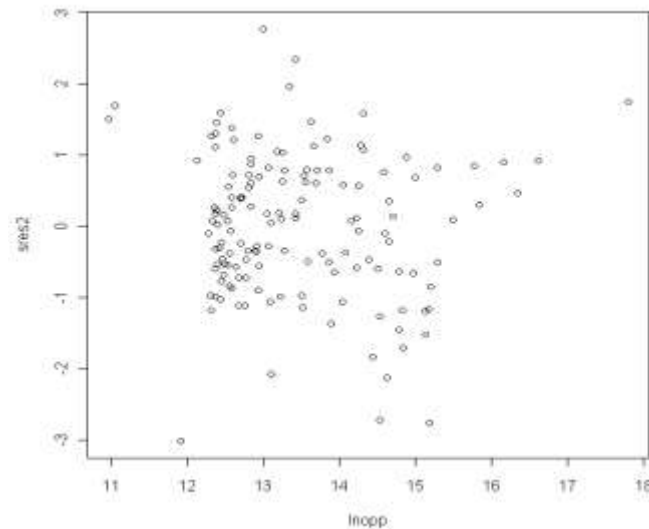
Der er endvidere kontrolleret for, hvorvidt betingelserne for at bruge regressionsanalysens resultater er opfyldt for modellen. Kontrollen foretages ved at plote de standardiserede fejllid mod de observerede værdier af den oppumpede vandmængde.

Figur 4 nedenfor viser de standardiserede fejllid plottet mod de observerede værdier af den oppumpede vandmængde. Plottet viser en umiddelbar tilfældig fordeling af fejllidene omkring 0, hvilket indikerer, at normalfordelingsbetingelsen er opfyldt.

Yderligere lader der ikke til at være tegn på heteroskedasticitet. Der er nogle få observationer, der falder uden for intervallet $[-1,96; 1,96]$ men det er

umiddelbart under 5 pct. af observationerne og dermed hvad der kan forventes, jf. figur 4.

Figur 4: Fejlplots for boringer



Der er ikke nogen enkelte observationer, der har stor indflydelse på B-værdierne. Dette måles med Cook's afstand som beskrevet i afsnittet *"Ekstreme observationer og observationer med stor indflydelse"* ovenfor. Der er dog fjernet en enkelt observation, som adskilte sig i så væsentlig grad, at det forstyrrede sammenhængen i residualplottet, hvorfor denne observation også er fjernet fra analysen.

Modellens baggrund

Forsyningssekretariatet har indsamlet data for fire variable for costdriveren boringer. Disse variable er:

- A. Antal boringer,
- B. Total oppumpet vandmængde,
- C. Samlet løftehøjde og
- D. Total kapacitet.

Disse data er indsamlet ud fra en forventning om, at de fire variable har indflydelse på størrelsen af de samlede omkostninger tilknyttet en vandforsynings boringer.

En gennemgang af indberetningerne viser, at flere forsyninger ikke har tilfredsstillende data på kapaciteten af deres boringer. Kapaciteten er derfor ikke medtaget i modellen.

Forsyningssekretariatet er endvidere blevet opmærksom på, at flere forsyninger indberetter boringernes dybdemeter svarende til værdien angivet i pris- og levetidskataloget. Dette er ikke i overensstemmelse med definitionen i *Vejledning til indberetning af oplysninger til genberegning af omkostningsækvivalenter*. Det betyder, at nogle forsyninger har angivet den

korrekte løftehøjde og andre forsyninger har angivet løftehøjden baseret på boringernes dybdemeter. Forskellen betyder, at analyser baseret på disse data vil resultere i forkerte resultater, hvorfor Forsyningssekretariat har udelukket løftehøjde fra regressionen.

En indledende kontrol viser, at det ikke er muligt at inddrage både antallet af boringer og den oppumpede vandmængde. Det skyldes, at der er en meget stor sammenhæng mellem de to variable. Korrelationen udgør således ca. 0,97. Forsyningssekretariatet kan derfor ikke anvende begge variable i regressionen til bestemmelse af omkostningsækvivalenten. Resultaterne af en analyse med begge variable resultere i, at antallet af boringer ikke er signifikant med en p-værdi på over 0,9. Det svarer til, at sandsynligheden for at der ikke er en sammenhæng mellem antallet af boringer og den oppumpede vandmængde er over 90 pct.

Forsyningssekretariatet vurderer desuden, at den oppumpede vandmængde er det mest intuitive og præcise mål til at beskrive driftsomkostningerne forbundet med at drive boringer. Forsyningssekretariatet har også forsøgt at lave en model med antallet af boringer med den samme logaritmiske transformation. Denne model rammer dog meget skævt for de forsyninger, som kun har meget få boringer i forhold til oppumpet vandmængde. Yderligere giver denne model en lidt lavere værdi af R^2 , og det er nødvendigt at fjerne en enkelt outlier. Samlet vurderes det derfor bedst at lave en model, der bruger den totale oppumpede vandmængde.

Den funktionelle form af omkostningsækvivalenten for boring er en eksponentiel funktion. Muligheden for at lave en lineær sammenhæng mellem omkostningerne og den oppumpede vandmængde er også blevet undersøgt. Regressionskørsler af faktiske værdier mod modellens forudsagte værdier har dog vist, at den eksponentielle form giver den bedste beskrivelse af data. Endelig skal det bemærkes, at Forsyningssekretariatet har testet for om der faste omkostninger ved at drive boringer, hvilket ikke er tilfældet.

Vandværker

Driftsomkostningerne ved at drive vandværker forventes at afhænge positivt af mængden af udpumpet vand fra vandværkerne. Det forventes yderligere, at jo mere behandling vandet undergår, jo dyrere vil det være at behandle det. Det er også en umiddelbar forventning, at den største forskel bør være på, om der foretages behandling i forhold til ingen behandling.

Forsyningssekretariatet har vurderet at, den mest velegnede model, til at beskrive de samlede driftsomkostninger ved at drive vandværker, er den lineære ligning (8), hvor Y angiver de årlige driftsomkostninger ved at drive vandværker, X_1 er mængden af udpumpet vand fra vandværker uden behandling, X_2 er mængden af udpumpet vand fra vandværker med almindelig behandling og X_3 er mængden af udpumpet vand fra vandværker med avanceret behandling:

$$(8) \quad Y = B_1X_1 + B_2X_2 + B_3X_3$$

B-værdierne udtrykker enhedsomkostningerne pr. m³ for hver vandbehandlingstype.

I tabel 4 nedenfor præsenteres resultaterne af modellen fra ligning 8.

Tabel 4: Regressionsanalysens resultater, vandværker

Variabel	B-værdi	Spredning	t-værdi	Pr(> t)
Type 1	0,748	0,16706	4,475	1.55e-05
Type 2	1,613	0,05932	27,194	< 2e-16
Type 3	1,774	0,12367	14,341	< 2e-16
Antal observationer: 145 Justeret R ² = 0,89				

Omkostningsækvivalenten for vandværker

Omkostningsækvivalenten ved at drive vandværker har følgende udseende, jf. tabel 4:

$$(9) \quad Y = 0,748X_1 + 1,613X_2 + 1,774X_3$$

Dermed angiver modellen, at det koster 0,748 kr. om året for en forsyning at udpumpe 1 m³ vand uden nogen form for behandling. Herudover viser modellen, at det koster 1,613 kr. om året at udpumpe 1 m³ vand med almindelig behandling og det koster 1,774 kr. om året at udpumpe 1 m³ vand med avanceret behandling.

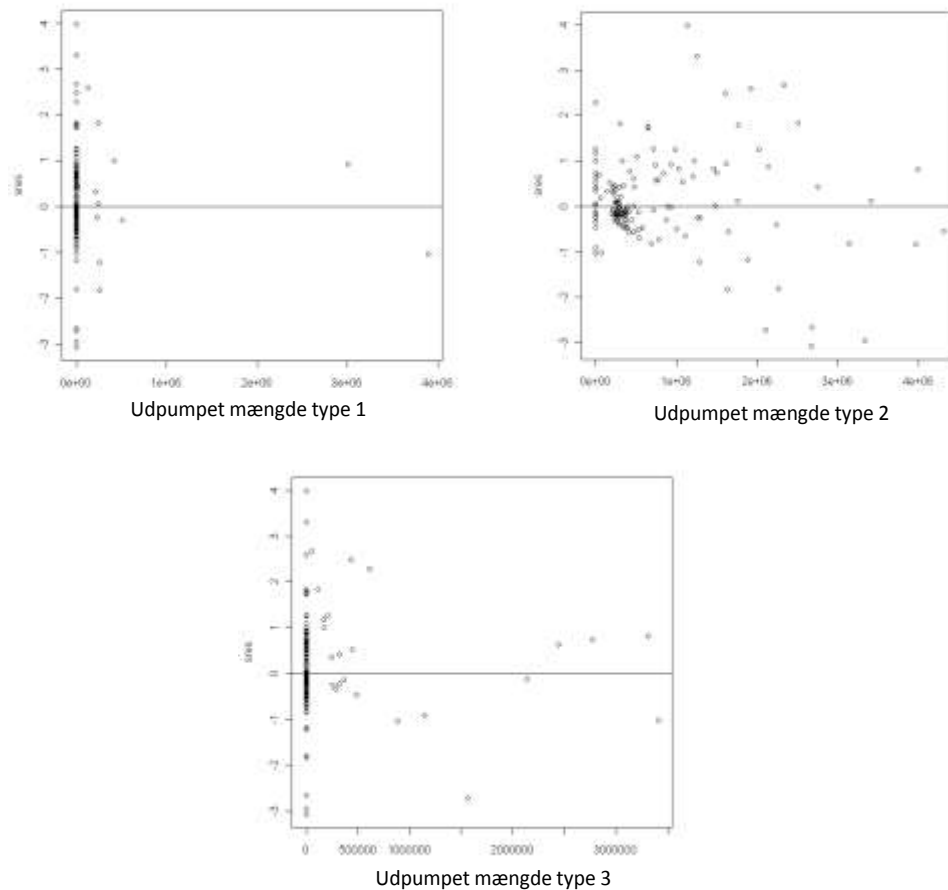
Kvalitetskontrol af omkostningsækvivalenten

Resultaterne i tabel 4 er tilfredsstillende, idet B-værdierne som forventet er positive, hvilket betyder at driftsomkostningerne stiger med en større udpumpet vandmængde. Yderligere bekræftes det, at jo mere behandling der foretages af vandet, jo dyrere er det at drive vandværk. Det fremgår af modellen, at type 2 og type 3 omkostningsmæssigt er tæt på hinanden. Dette indikerer, at forskellen ved vandbehandling primært vedrører forskellen på, hvorvidt der laves en egentlig behandling eller der ikke foretages nogen behandling. Det ses også af tabel 4, at alle koefficienterne i modellen er signifikante.

Herudover er der kontrolleret for, hvorvidt betingelserne for at bruge regressionsanalysens resultater er opfyldt for modellen. Kontrollen foretages

ved at plotte de standardiserede fejlede mod de observerede værdier af de tre typer af vandbehandling. Disse plots fremgår i figur 5 nedenfor:

Figur 5: Fejlledsplot for vandværker



Plottene viser umiddelbart ingen tegn på problemer, idet fejlleddene fordeler sig rimeligt pænt omkring 0. Resultaternes robusthed ville naturligvis være højere, hvis der var et stort antal observationer for alle typer af vandbehandling. Type 1 og type 3 har relativt få observationer, hvilket skyldes, at få forsyninger foretager en anden behandling end almindelig vandbehandling.

Det har været nødvendigt at fjerne syv observationer med stor indflydelse på B-værdierne. Dette er målt med Cook's afstand som beskrevet i afsnittet ”*Ekstreme observationer og observationer med stor indflydelse*” ovenfor.

Modellens baggrund

Forsyningssekretariatet har indsamlet data for fire variable for hvert vandværk:

- A. Udpumpet vandmængde
- B. Typen af behandling og
- C. Kapacitet

Indsamlingen af data er foretaget på baggrund af en forventning om, at de 4 variable har indflydelse på størrelsen af de samlede driftsomkostninger for en vandforsynings vandværker.

Forsyningssekretariatet har vurderet, at det ikke vil være retvisende at fastsætte faste omkostninger ved at drive et vandværk. Dette skyldes, at der ikke er intuitiv mening i at fastsætte identiske faste omkostninger for forsyningernes vandværker, da der er markante størrelsesforskelle.

I kvalitetssikringen af indberetningerne har det vist sig, at flere forsyninger ikke har leveret tilfredsstillende data om kapaciteten af deres vandværker. Derfor er kapacitet ikke medtaget.

Yderligere er Forsyningssekretariatet blevet opmærksom på, at flere forsyninger har fordelt driftsomkostningerne ud på de enkelte vandværker efter, hvor stor udpumpet mængde de har og ikke med hensyn til typen af vandbehandling. Dette er ikke i overensstemmelse med definitionen i *Vejledning til indberetning af oplysninger til genberegning af omkostningsækvivalenter*. Det betyder, at nogle forsyninger har angivet driftsomkostningerne efter de enkelte vandværker hvor der er taget hensyn til forskellen i typen af behandling, mens andre har angivet driftsomkostningerne fordelt efter udpumpet vandmængde.

Hvis disse data blev anvendt i analysen, ville det skævvride resultaterne af en regressionsanalyse, der er baseret på selskabernes omkostningsfordeling på de enkelte vandværker. Derfor har Forsyningssekretariatet samlet hvert selskabs omkostninger og mængder med forskellig behandling til et samlet udtryk for selskabets samlede vandværker, jf. ligning 9.

Trykforøgerstationer

Driftsomkostningerne ved at drive trykforøgerstationer i forskellige kapacitetsintervaller forventes at afhænge af antallet af trykforøgerstationer. Herudover forventes det, at trykforøgerstationer med en stor kapacitet driver større driftsomkostninger end mindre trykforøgerstationer.

Forsyningssekretariatet har vurderet, at den mest velegnede model, til at beskrive de samlede omkostninger ved at drive trykforøgerstationer, er den lineære ligning (10), hvor

- Y angiver de årlige driftsomkostninger ved at drive trykforøgerstationer,
- X_1 angiver antallet af trykforøgerstationer i intervallet 0-50 m³/t,
- X_2 er antallet af trykforøgerstationer i intervallet 51-100 m³/t,
- X_3 er antallet af trykforøgerstationer i intervallet 101-200 m³/t,
- X_4 er antallet af trykforøgerstationer i intervallet 201-600 m³/t og
- X_5 er kapaciteten af trykforøgerstationer i intervallet 601-max m³/t:

$$(10) \quad Y = B_1X_1+B_2X_2+B_3X_3+B_4X_4+B_5X_5$$

B-værdierne udtrykker enhedsomkostningerne for hver kategori af trykforøgerstationer.

Resultaterne af modellen i ligning (10) fremgår af tabel 5 nedenfor.

Tabel 5: Regressionsanalysens resultater, Trykforøgerstationer

Variabel	B-værdi	Spredning	t-værdi	Pr(> t)
0-50 m ³ /t	35.453	6.931	5,115	1,45e-06
51-100 m ³ /t	56.469	22.456	2,515	0,0135
101-200 m ³ /t	144.773	30.084	4,812	5,13e-06
201-600 m ³ /t	178.851	32.993	5,421	3,93e-07
601-max m ³ /t	440	63	7,02	2,43e-10
Antal observationer: 108 Justeret R ² = 0,8238				

Omkostningsækvivalenten for trykforøgerstationer

Omkostningsækvivalenten ved at drive trykforøgerstationer har følgende udseende, jf. tabel 5:

$$(11) \quad Y = 35.453X_1+56.469X_2 + 144.773X_3+178.851X_4 + 440X_5$$

Modellen angiver dermed, at en trykforøgerstation i intervallet 0-50 m³/t koster 35.453 kr. at drive om året, en trykforøgerstation i intervallet 50-100 m³/t koster 56.469 kr. at drive om året, en trykforøgerstation i intervallet 100-200 m³/t koster 144.773 kr. at drive om året, en trykforøgerstation i intervallet 200-600 m³/t koster 178.851 kr. at drive om året og en trykforøgerstation i intervallet 600-max m³/t koster 404 kr. pr. m³/t at drive pr. m³/t.

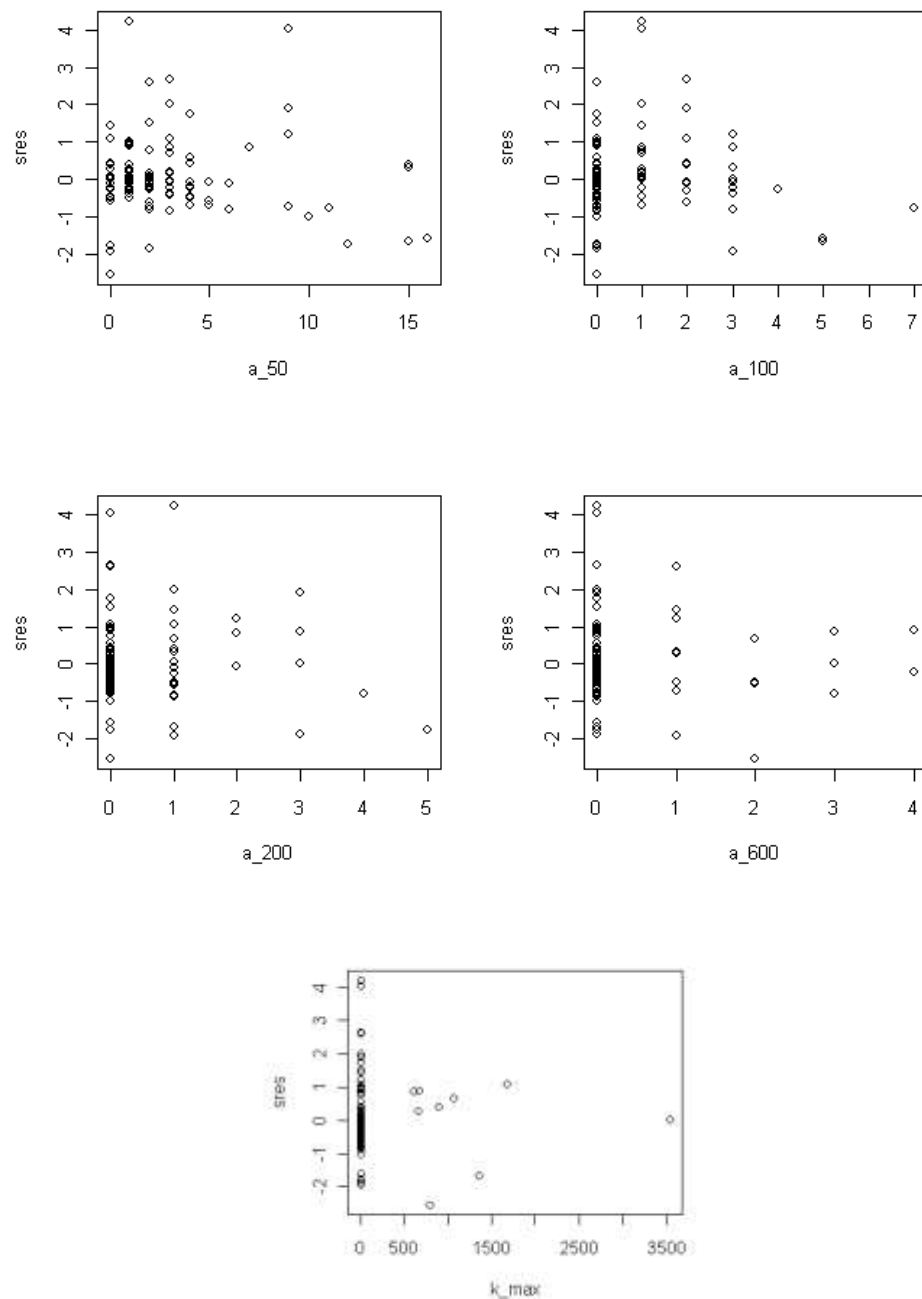
Kvalitetskontrol af omkostningsækvivalenten

Resultaterne i tabel 5 er tilfredsstillende, idet B-værdierne som forventet er positive, hvilket betyder at driftsomkostningerne stiger med et højere antal trykforøgerstationer, og at driftsomkostningerne ved at drive en stor trykforøgerstation er højere end ved at drive en mindre trykforøgerstation. Det ses desuden af tabel 5, at alle koefficienterne i modellen er signifikante.

Herudover er der kontrolleret for, hvorvidt betingelserne for at bruge regressionsanalysens resultater er opfyldt for modellen. Kontrollen foretages ved at plotte de standardiserede fejllid mod de observerede værdier af de

fem intervaller af trykforøgerstationer. Disse plots fremgår i figur 6 nedenfor:

Figur 6: Fejlledsplot for trykforøgerstationer



Plottene viser umiddelbart ingen tegn på problemer, idet fejlleddene fordeler sig pænt omkring 0. Resultaternes robusthed ville naturligvis være højere, hvis der var et stort antal observationer i hvert interval for trykforøgerstationerne. De to største intervaller har relativt få observationer, hvilket skyldes, at kun få forsyninger har store trykforøgerstationer.

Det har været nødvendigt at fjerne enkelte observationer med stor indflydelse på B-værdierne. Dette er målt med Cook's afstand som beskrevet i afsnittet "*Ekstreme observationer og observationer med stor indflydelse*" ovenfor.

Modellens baggrund

Forsyningssekretariatet har indsamlet data for 10 variable for forsyningernes trykforøgere:

- Antal pumper i kategori 0-50 m³/t
- Kapacitet af pumper i kategorien 0-50 m³/t
- Antal pumper i kategori 51-100 m³/t
- Kapacitet af pumper i kategorien 51-100 m³/t
- Antal pumper i kategori 101-200 m³/t
- Kapacitet af pumper i kategorien 101-200 m³/t
- Antal pumper i kategori 201-600 m³/t
- Kapacitet af pumper i kategorien 201-600 m³/t
- Antal pumper i kategori 601-max m³/t
- Kapacitet af pumper i kategorien 601-max m³/t

Der er blevet testet en række alternative regressioner ved at sammensætte antal trykforøgerstationer og kapaciteter på forskellige måder. Blandt andet er der udført regressioner med flere og færre intervalinddelinger for antallet af trykforøgerstationer. Herudover er der udført regressioner på baggrund af trykforøgerstationernes samlede kapaciteter med flere forskellige intervalinddelinger. Derudover er der udført en regression, som tester for, hvorvidt der er faste omkostninger ved at drive trykforøgerstationer.

Regressionerne udført på baggrund af trykforøgerstationernes kapaciteter giver ikke intuitive og fortolkningsmæssige gode resultater. Når kapaciteterne inddeles efter kapacitetsintervaller er det ikke muligt at verificere resultaterne intuitivt. Det betyder, at resultaterne fortolkningsmæssigt ikke understøttes af en klar og fornuftig sammenhæng. Derfor blev der udført en regression udelukkende på trykforøgerstationernes samlede kapacitet. Denne model er meget simpel, hvorfor den eksempelvis ikke kan tage højde for, at nogle selskaber driver mange små trykforøgerstationer eller nogle selskaber driver færre store trykforøgerstationer.

Resultaterne af de gennemførte regressioner peger på, at det er regressionen i ligning (11), der forklarer selskabernes driftsomkostninger ved at drive trykforøgerstationer bedst. Denne model er også den mest robuste. I denne regression har det været nødvendigt at slå antallet af trykforøgerstationer i intervallet fra 200-400 m³/t og antallet af trykforøgerstationer i intervallet 400-600 m³/t sammen til en samlet kategori på 200-600 m³/t. Herudover bemærkes det, at resultaterne viser, at der ikke er en signifikant fast omkostning ved at drive trykforøgerstationer.

Regressionen i ligning (11) er blandt andet valgt, fordi det er denne regression, som bedst tager hensyn til selskabernes individuelle forhold i trykforøgerstationerne. Opgørelsen på antal indenfor et givent interval vil eksempelvis fange, hvis et selskab har to små trykforøgerstationer i stedet for en stor trykforøgerstation. Dog opgøres trykforøgerstationer over 600 m³/t med kapaciteten. På den måde tager modellen højde for at trykforøgerstationer i denne kategori kan variere meget.

Denne model følger overordnet set intuitionen, at det er dyre at drive et større antal små trykforøgerstationer end at drive et færre antal store trykforøgerstationer. Dette kan eksempelvis illustreres ved, at det ifølge denne model vil koste 70.906 kr. (2*35.453) om året at drive to trykforøgerstationer med en kapacitet på 50 m³/t. Til sammenligning vil det koste 56.469 kr. om året at drive én trykforøgerstation med en kapacitet på 100 m³/t.

Modellen understøtter således, at nogle forsyninger skal pumpe meget vand over store afstande, eller at det er nødvendigt med flere små trykforøgerstationer for at tage hensyn til topologien i området.

Rentvandsledninger

Driftsomkostningerne ved at drive rentvandsledninger forventes at afhænge af ledningernes længde (meter) samt volumen (m³) fordelt på de fire zoner land, by, city og indre city. Det forventes, at store ledninger er dyrere at drive end små ledninger, og at omkostningerne stiger i et tættere befolket område.

Forsyningssekretariatet har vurderet, at den mest velegnede model til at beskrive de samlede driftsomkostninger ved at drive rentvandsledninger er beskrevet i ligning (12), hvor:

- Y angiver de årlige driftsomkostninger ved at drive rentvandsledninger,
- X₁ angiver meter rentvandsledning i land,
- X₂ angiver meter rentvandsledning i by,
- X₃ angiver meter rentvandsledning i city og
- X₄ angiver meter rentvandsledning i indre city:

$$(12) \quad Y = B_1(X_1+X_2)+B_2(X_3+X_4)$$

B-værdierne udtrykker enhedsomkostningerne for hver zonekategori af rentvandsledninger.

Resultaterne af modellen fra ligning (12), fremgår af tabel 6 nedenfor:

Tabel 6: Regressionsanalysens resultater, rentvandsledning

Variabel	B-værdi	Spredning	t-værdi	Pr(> t)
Land + By	4,0474	0.4049	9.996	<2e-16
City + Indre city	85,3695	5.7370	14.880	<2e-16
Antal observationer: 152 Justeret R ² = 0,8173				

Omkostningsækvivalenten for rentvandsledninger

Omkostningsækvivalenten ved at drive rentvandsledninger har følgende udseende, jf. tabel 6:

$$(13) \quad Y = 4,047(X_1+X_2) + 85,370(X_3+X_4)$$

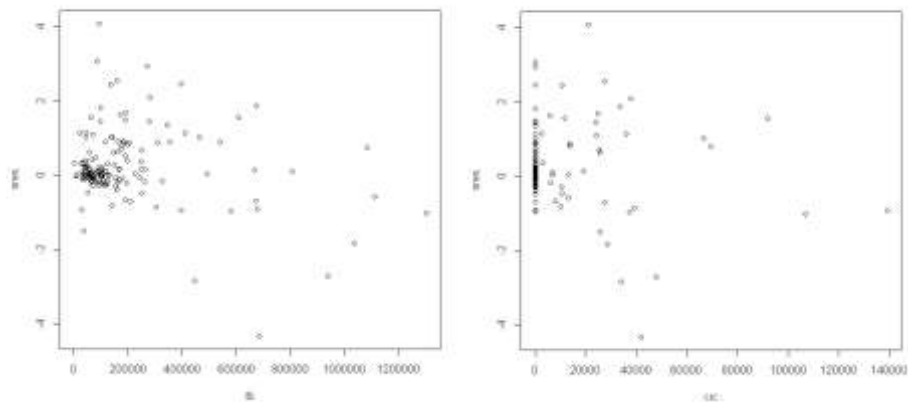
Dermed angiver modellen, at det koster 4,047 kr. om året at drive en meter rentvandsledning i land eller by og, at det koster 85,370 kr. om året at drive en meter rentvandsledning i city eller indre city.

Kvalitetskontrol af omkostningsækvivalenten

Resultaterne i tabel 6 er tilfredsstillende, idet B-værdierne følger den økonomiske intuition – det vil sige, at driftsomkostningerne stiger, jo længere rentvandsledningerne er i de fire zoner. Dertil kommer, at alle koefficienterne er signifikante.

Der er herudover kontrolleret for, hvorvidt betingelserne for at bruge regressionsanalysens resultater er opfyldt for modellen. Kontrollen foretages ved at plotte de standardiserede fejlede mod de observerede værdier af de to kategorier af rentvandsledninger. Disse plots fremgår i figur 7 nedenfor:

Figur 7: Fejlledsplot for rentvandsledninger



Plottene viser umiddelbart ingen tegn på problemer, idet fejlleddene fordeler sig pænt omkring 0. Antallet af observationer er større for kategorien land og by, idet næsten alle selskaber har rentvandsledninger i disse kategorier, men det er ikke alle selskaberne, som har rentvandsledninger i kategorierne city og indre city.

Det har været nødvendigt at fjerne enkelte observationer med stor indflydelse på B-værdierne. Dette er målt med Cook's afstand som beskrevet i afsnittet *"Ekstreme observationer og observationer med stor indflydelse"* ovenfor.

Modellens baggrund

Forsyningssekretariatet har indsamlet data for 8 variable for costdriveren rentvandsledning. Disse variable er:

- Meter rentvandsledning < 110mm i land,
- Meter rentvandsledning > 110mm i land,
- Meter rentvandsledning < 110mm i by
- Meter rentvandsledning > 110mm i by
- Meter rentvandsledning < 110mm i city og
- Meter rentvandsledning > 110mm i city og
- Meter rentvandsledning < 110mm i indre city:
- Meter rentvandsledning > 110mm i indre city:

Forsyningssekretariatet har udført en række regressioner for at teste alternative sammenhænge mellem selskabernes driftsomkostninger ved at drive rentvandsledninger og længden af rentvandsledninger i forskellige zonekategorier.

Derudover er der blandt andet gennemført regressioner, som inddrager rentvandsledningernes volumen. Endelig er der blevet testet for stordriftsfordele i længden af rentvandsledninger og for stordriftsfordele i typen af rentvandsledninger.

Regressionerne, der inddrager ledningernes dimensioner, gav ikke fornuftige og robuste resultater. Ledningernes dimensioner er derfor udeladt af modellen.

Forsyningssekretariatet har vurderet på baggrund af en række regressionsanalyser, at der ikke er grundlag for at inddrage stordriftsfordele i omkostningsækvivalenten for ledninger. Det skyldes, at der ikke er en overbevisende tendens i data, som bekræfter en sammenhæng mellem driftsomkostningerne og stordriftsfordele.

Resultaterne af de gennemførte regressioner peger på, at det er regressionen i ligning (13), der bedst og mest robust forklarer selskabernes driftsomkostninger ved at drive rentvandsledninger. I denne regression har det været nødvendigt at slå længden af rentvandsledningerne i land og by sammen og slå længden af rentvandsledningerne i city og indre city sammen. Denne model er fornuftig, idet modellen kan håndtere, at selskaberne har rentvandsledninger i forskellige zonekategorier. Herudover kan modellen håndtere, at selskaberne har forskellige typer af forsyningsområder, hvor indbyggertætheden kan variere. Dette skyldes, at regressionen bygger på længden af ledninger i forskellige zonekategorier.

Stik

Driftsomkostningerne ved at drive stik forventes i udgangspunktet at afhænge af antallet af stik og zoneplaceringen. Beregningen af omkostningsækvivalenten for stik kræver en mere kompliceret beregning end for de andre costdrivere for vandforsyninger. Det er nødvendigt at tage forsyninger med stik i indre cityzonen ud af datasættet og i første omgang kun beregne driftsomkostninger forbundet med stik i zonerne land, by og city på baggrund af regressionsanalysen.

Dette har været nødvendigt, da variationen i driftsomkostningerne for de forsyninger, som har stik i indre cityzone, har adskilt sig meget fra de øvrige forsyninger. Det betyder, at fem ud af de seks selskaber, som har stik i indre city har en høj Cook's D og derfor er blevet identificeret som outliers i modellen. Idet der kun er et enkelt selskab med stik i indre city tilbage i observationssættet, vil en regression indeholdende en variabel for stik i indre cityzone ikke være repræsentativ.

Forsyningssekretariatet har derfor udført en regression med variable for stik i land-, by- og cityzonerne. Herefter har Forsyningssekretariatet estimeret driftsomkostningerne ved at drive stik i indre city-zone på baggrund af de selskaber, som har stik i indre city-zone.

Den mest velegnede model til at beskrive de samlede omkostninger ved at drive stik er ligning (14), hvor Y angiver de samlede driftsomkostninger forbundet med stik, X_1 angiver antal stik i landzone, X_2 angiver antal stik i byzone, X_3 angiver antal stik i cityzone og X_4 angiver antal i indre cityzone:

$$(14) \quad Y = B_1(X_1+X_2) + B_2X_3 + B_3X_4$$

Ovenstående betyder, at modellen beskriver de samlede omkostninger ved at drive stik. Dette sker ved at beregne en stykpris (B-værdien) for hvert stik i de forskellige zoner. Dernæst beregnes summen af samtlige stykpriser for stikkene fordelt på de forskellige zoner.

Forsyningssekretariatet er nået frem til modellen i ligning (14) i to skridt:

Første skridt er at estimere driftsomkostningerne ved at drive stik i land-, by- og city-zone, hvilket gøres i følgende regression:

$$(15) \quad Y = B_1(X_1+X_2) + B_2X_3$$

Y angiver de samlede driftsomkostninger forbundet med stik, X_1 angiver antal stik i landzone, X_2 angiver antal stik i byzone og X_3 angiver antal stik i cityzone. Til denne regression er der kun anvendt data for selskaber uden stik i indre city-zone.

I tabel 7 nedenfor præsenteres resultaterne af modellen fra ligning (15):

Tabel 7: Regressionsanalysens resultater, stik

Variabel	B-værdi	Spredning	t-værdi	Pr(> t)
Land + By	135,76	12,29 (17,29)	11,044 (7,86)	<2e-16 (1,04e-12)
City	430,18	114,22 (175,59)	3,766 (2,45)	0,000246 (0,015)
Antal observationer: 138 Justeret $R^2 = 0,6933$				

Tal i parenteser angiver robuste værdier

Andet skridt er at estimere driftsomkostningerne ved at drive stik i indre cityzone, hvilket gøres i det følgende:

For at beregne omkostningsækvivalenten for stik i indre cityzonen beregnes først værdien af de stik, som forsyningerne med indre cityzone har i de øvrige to kategorier (land+by og city). Denne værdi beregnes på baggrund af resultaterne af modellen i ligning (15), som fremgår i tabel 7.

Når denne værdi er fastlagt beregnes en værdi for de driftsomkostninger, som forsyningerne har ved at drive deres stik i indre city. Dette gøres ved at opgøre forskellen på de forudsagte værdier fra ligning (15) og forsyningens samlede indberettede omkostninger til at drive deres stik.

På baggrund af hver forsynings beregnede omkostninger til stik i indre city zone beregnes en gennemsnitlig udgift til stik i indre cityzonen, jf. tabel 8 nedenfor.

Tabel 8: Stik i indre city

Omkostninger i alt	Værdi af øvrige zoner	Forskel	Antal stik	Gennemsnit
65.278	490.367	-425.089	32	-13.284
13.356.326	6.049.046	7.307.280	513	14.244
8.666.168	1.043.512	7.622.656	2.398	3.179
6.593.348	2.515.064	4.078.284	2.975	1.371
5.800.167	6.792.917	-992.750	6.058	-164
21.943.972	6.969.796	14.974.176	7.870	1.903
		Gns. omkostning pr. stik		1.208,10

Omkostningsækvivalenten for stik

Omkostningsækvivalenten ved at drive stik har herefter følgende udseende, jf. tabel 7 og 8:

$$(16) \quad Y = 135,8(X_1 + X_2) + 430,2X_3 + 1208,1X_4$$

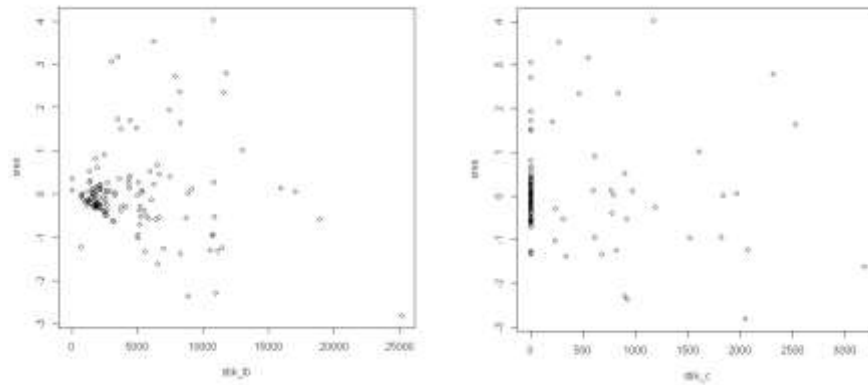
Dermed angiver modellen, at et stik i land- eller by-zone koster 135,76 kr. at drive om året, et stik i cityzone koster 430,18 kr. at drive om året og et stik i indre cityzone koster 1208,10 kr. at drive om året.

Kvalitetskontrol af omkostningsækvivalenten

Resultaterne i tabel 7 er rimelige, idet B-værdierne som forventet er positive, hvilket betyder at driftsomkostningerne stiger ved et større antal stik i de fire zoner. Herudover er det ifølge modellen også som forventet dyrere at drive stik i zoner med høj befolkningstæthed og en mere kompliceret infrastruktur. Dertil kommer, at alle koefficienterne er signifikante.

Herudover er der kontrolleret for, hvorvidt betingelserne for at bruge regressionsanalysens resultater er opfyldt for modellen. Kontrollen foretages ved at plote de standardiserede fejled mod de observerede værdier af de to kategorier af ledninger. Disse plots fremgår i figur 8 nedenfor:

Figur 8: Fejlledsplot for stik



Plottene udviser heteroskedasticitet for stik i (land+by) kategorien. Der er derfor beregnet robuste t-værdier, som tager hensyn til, at variansen af fejlleddene er uens. Beregningen viser, at der stadig er en signifikant sammenhæng imellem målere og omkostninger jf. tabel 7.

Det har været nødvendigt at fjerne enkelte observationer med stor indflydelse på B-værdierne. Dette er målt med Cook's afstand som beskrevet i afsnittet "*Ekstreme observationer og observationer med stor indflydelse*" ovenfor.

Modellens baggrund

Forsyningssekretariatet har udført en række regressioner for at teste alternative sammenhænge mellem selskabernes driftsomkostninger til at drive stik og antallet af stik i de forskellige zoner. Der er blandt andet gennemført regressioner, hvor stikkene er fordelt på alle fire zoner. Derudover er der gennemført en række regressioner, som slår de forskellige zoner sammen i forskellige kombinationer. Disse regressioner har dog ikke givet gode resultater.

Resultaterne af de gennemførte analyser og regressioner peger på, at det er modellen i ligning (19), der bedst forklarer selskabernes driftsomkostninger ved at drive stik i de forskellige zoner. I regressionen for selskabernes stik i land, by og city har det været nødvendigt at slå antallet af stik i land- og byzone sammen til én variabel. Herudover har det været nødvendigt at estimere driftsomkostningerne ved at drive et stik i indre city-zone ved et gennemsnit pr. stik for de selskaber, som har stik i indre city-zone. Denne model er fornuftig, idet modellen kan håndtere selskaber med et stort og et lille antal stik. Herudover giver modellen fornuftige resultater, idet stikkernes fordeling på forskellige zoner vil fange eventuelle fordele eller ulemper ved at have stik i tyndt eller tæt befolkende forsyningsområder.

Kunder

Det forventes at, driftsomkostningerne ved at håndtere selskabernes kunder afhænger af antallet af målere. Dette skyldes, at antallet af målere

giver en indikation af, hvor mange kunder selskaberne har i deres forsyningsområder.

Forsyningssekretariatet har vurderet at den mest velegnede model til at beskrive de samlede omkostninger ved at håndtere kunder er en lineær sammenhæng, jf. ligning (17). Y angiver de årlige driftsomkostninger forbundet ved at håndtere kunder og X_1 angiver antallet af målere:

$$(17) \quad Y = B_1 X_1$$

B-værdien udtrykker enhedsomkostningerne for målerne.

I tabel 9 nedenfor præsenteres resultaterne af modellen i ligning (17):

Tabel 9: Regressionsanalysens resultater, Kunder

Variabel	B-værdi	Spredning	t-værdi	Pr(> t)
Målere	149,84	6,587()	22,75 ()	<2e-16()
Antal observationer:151 Justeret $R^2=0,69$				

Tal i parenteser er de robuste værdier

Omkostningsækvivalenten for kunder

Omkostningsækvivalenten ved at håndtere kunder har følgende udseende jf. tabel 9:

$$(18) \quad Y = 149,8 X_1$$

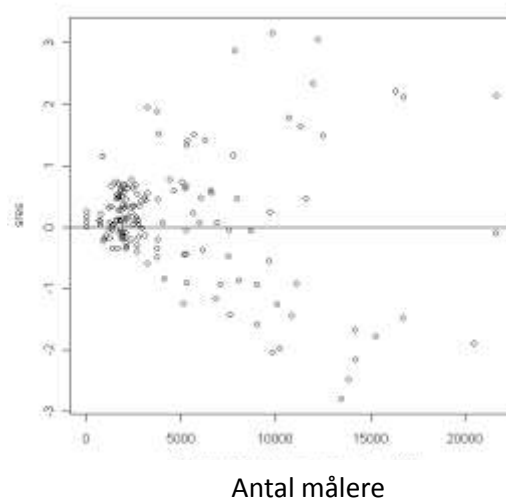
Dermed angiver modellen, at en måler koster 149,8 kr. at drive om året.

Kvalitetskontrol af omkostningsækvivalent

Resultatet i tabel 9 er tilfredsstillende, idet B-værdien som forventet er positiv, hvilket betyder, at driftsomkostningerne stiger ved et større antal målere. Desuden ses det af tabel 9, at koefficienten i modellen er signifikant.

Herudover er der kontrolleret for, hvorvidt betingelserne for at bruge regressionsanalysens resultater er opfyldt for modellen. Kontrollen foretages ved at plotte de standardiserede fejllid mod de observerede værdier for målerne. Dette plot fremgår i figur 9 nedenfor:

Figur 9: Fejlledsplot for målere



Plottet udviser en vis grad af heteroskedasticitet. Der er derfor beregnet robuste t-værdier, som tager hensyn til, at variansen af fejleddene er uens. Beregningen viser, at der stadig er en signifikant sammenhæng imellem målere og omkostninger jf. tabel 9.

Det har været nødvendigt at fjerne 6 observationer med stor indflydelse på B-værdierne. Dette måles med Cook's afstand, som beskrevet i afsnittet *"Ekstreme observationer og observationer med stor indflydelse"* ovenfor. Yderligere er der fjernet en enkelt observation, der forstyrrede billedet i fejlledsplottet.

Modellens baggrund

Resultaterne af den gennemførte regression giver et fornuftigt resultat. Resultatet er signifikant, og det følger den økonomiske intuition. Forsyningssekretariatet bemærker, at sekretariatet ikke har testet andre sammenhænge end den ovenstående, idet selskaberne i 2012 kun er blevet bedt om at indberette måleroplysninger. Til sidste års beregning af omkostningsækvivalenter indberettede selskaberne ligeledes antallet af husstande, hvilket ikke resulterede i et lige så godt resultat som antallet af målere.

Bestemmelse af omkostningsækvivalenter for Spildevand

Forsyningssekretariatet har beregnet en omkostningsækvivalent for hver af de otte costdrivere:

- Ledninger
- Pumpestationer
- Regnvandsbassiner
- Spildevandsbassiner
- Minirenselanlæg
- Renselanlæg
- Slambehandling
- Kunder

Beregningen af de enkelte omkostningsækvivalenter bliver gennemgået nedenfor. Først beskrives den model, som Forsyningssekretariatet efter en række analyser, har vurderet, er den mest velegnede model til at beskrive driftsomkostningerne for den enkelte costdriver. Herefter beskrives de forudgående analyser, der har ført til valget af modellen.

Ledninger

Driftsomkostningerne ved at drive ledninger forventes i udgangspunktet at afhænge af ledningernes længde samt ledningernes volumen fordelt på de fire zoner land, by, city og indre city.

Den mest velegnede model til at beskrive de samlede omkostninger ved at drive ledninger er den lineære sammenhæng i ligning (19). Y angiver de årlige driftsomkostninger ved at drive ledninger, X₁ angiver meter ledning i land, X₂ angiver meter ledning i by, X₃ angiver meter ledning i city og X₄ angiver meter ledning i indre city:

$$(19) \quad Y = B_1(X_1+X_2)+B_2(X_3+X_4)$$

B-værdierne udtrykker enhedsomkostningen for hver zonekategori af ledningerne.

I tabel 10 nedenfor præsenteres resultaterne af modellen fra ligning (19):

Tabel 10: Regressionsanalysens resultater, ledninger

Variabel	B-værdi	Spredning	t-værdi	Pr(> t)
Land + By	4,2790	0,4323	9,898	1,34e-15
City + Indre city	87,0884	8,1367	10,703	<2e-16
Antal observationer: 83 Justeret R ² = 0,8558				

Omkostningsækvivalenten for ledninger

Omkostningsækvivalenten ved at drive ledninger har følgende udseende, jf. tabel 10:

$$(20) \quad Y = 4,279(X_1 + X_2) + 87,088(X_3 + X_4)$$

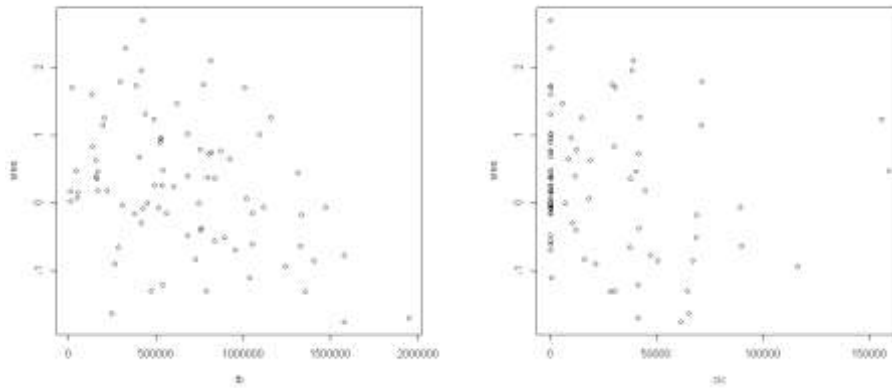
Dermed angiver modellen, at en meter ledning i land eller by koster 4,279 kr. at drive om året og en meter ledning i city eller indre city koster 87,088 kr. at drive om året.

Kvalitetskontrol af omkostningsækvivalenten

Resultaterne i tabel 10 er rimelige, idet B-værdierne som forventet er positive, hvilket betyder at driftsomkostningerne stiger jo længere ledningerne er i de fire zoner. Herudover er det dyrere at drive ledninger i tættere befolkede forsyningsområder. Dertil kommer, at alle koefficienterne er signifikante.

Der er herudover kontrolleret for, hvorvidt betingelserne for at bruge regressionsanalysens resultater er opfyldt for modellen. Kontrollen foretages ved at plote de standardiserede fejllid mod de observerede værdier af de to kategorier af ledninger. Disse plots fremgår i figur 10 nedenfor:

Figur 10: Fejlledsplot for ledninger



Plottene viser umiddelbart ingen tegn på problemer, idet fejlleddene fordeler sig pænt omkring 0. Antallet af observationer er størst for kategorien land og by, idet næsten alle selskaber har ledninger i disse kategorier, men det er ikke alle selskaberne, som har ledninger i kategorierne city og indre city.

Det har været nødvendigt at fjerne enkelte observationer med stor indflydelse på B-værdierne. Dette er målt med Cook's afstand som beskrevet i afsnittet "Ekstreme observationer og observationer med stor indflydelse" ovenfor.

Modellens baggrund

Forsyningssekretariatet har indsamlet data for 8 variable for costdriveren ledning. Disse variable er:

- Meter ledning < 200mm. i land,
- Meter ledning > 200mm. i land,
- Meter ledning < 200mm. i by
- Meter ledning > 200mm. i by
- Meter ledning < 200mm. i city
- Meter ledning > 200mm. i city
- Meter ledning < 200mm. i indre city:
- Meter ledning > 200mm. i indre city:

Forsyningssekretariatet har udført en række regressioner for at teste alternative sammenhænge mellem forsyningernes driftsomkostninger ved at drive ledninger og selskabernes længder af ledninger i forskellige zonekategorier. Derudover er der blandt andet gennemført regressioner, som inddrager ledningernes volumen. Endelig er der blevet testet for stordriftsfordele i længden af ledninger og for stordriftsfordele i typen af ledninger.

Regressionerne, der inddrager ledningernes dimensioner, gav ikke fornuftige og robuste resultater. Ledningernes dimensioner er derfor udeladt af modellen.

Forsyningssekretariatet har, på baggrund af en række regressionsanalyser, vurderet, at der ikke er grundlag for at inddrage stordriftsfordele i omkostningsækvivalenten for ledninger. Det skyldes, at der ikke er en overbevisende tendens i data, som bekræfter en sammenhæng mellem driftsomkostningerne og stordriftsfordele.

Resultaterne af de gennemførte regressioner peger på, at det er regressionen i ligning (20), der bedst forklarer selskabernes driftsomkostninger ved at drive ledninger. I denne regression har det været nødvendigt at slå længden af ledningerne i land og by sammen og slå længden af ledningerne i city og indre city sammen. Herudover kan modellen håndtere, at forsyningerne har forskellige typer af forsyningsområder, hvor indbyggertætheden kan variere. Dette skyldes, at regressionen bygger på længden af ledninger i forskellige zonekategorier.

Pumpestationer

Driftsomkostningerne ved at drive pumpestationer i forskellige kapacitetsintervaller forventes at afhænge lineært af antallet af pumpestationer. Det forventes herudover, at pumpestationer med en stor kapacitet driver større driftsomkostninger end mindre pumpestationer.

Forsyningssekretariatet har vurderet, at den mest velegnede model til at beskrive de samlede driftsomkostninger ved at drive pumpestationer er ligning (22), hvor:

- Y angiver de årlige driftsomkostninger ved at drive pumpestationer,
- X_1 angiver antallet af husstandspumper,
- X_2 er antallet af pumpestationer i intervallet 0-10 m³/t,
- X_3 er antallet af pumpestationer i intervallet 10-100 m³/t,
- X_4 er antallet af pumpestationer i intervallet 100-600 m³/t og
- X_5 er kapaciteten af pumpestationer i intervallet 600-max m³/t.

$$(21) \quad Y = B_1X_1 + B_2X_2 + B_3X_3 + B_4X_4 + B_5X_5$$

B-værdierne udtrykker enhedsomkostningerne ved at drive hver kategori af pumpestationerne.

I tabel 11 nedenfor præsenteres resultaterne af modellen fra ligning (21).

Tabel 11: Regressionsanalysens resultater, pumpestationer

Variabel	B-værdi	Spredning	t-værdi	Pr(> t)
Husstandspumper	6.628	1.892,7	3,502	0,000763
0-10 m ³ /t	13.891	3.706,9	3,747	0,000339
11-100 m ³ /t	24.337	4.330,5	5,620	2,77e-07
101-600 m ³ /t	102.864	37.365	2,753	0,007325
601-max m ³ /t	597	151,3	3,945	0,000172
Antal observationer: 84 Justeret R ² = 0,8135				

Omkostningsækvivalenten for pumpestationer

Omkostningsækvivalenten ved at drive trykforøgerstationer har følgende udseende, jf. tabel 11:

$$(22) \quad Y = 6.628X_1 + 13.891X_2 + 24.337X_3 + 102.864X_4 + 597X_5$$

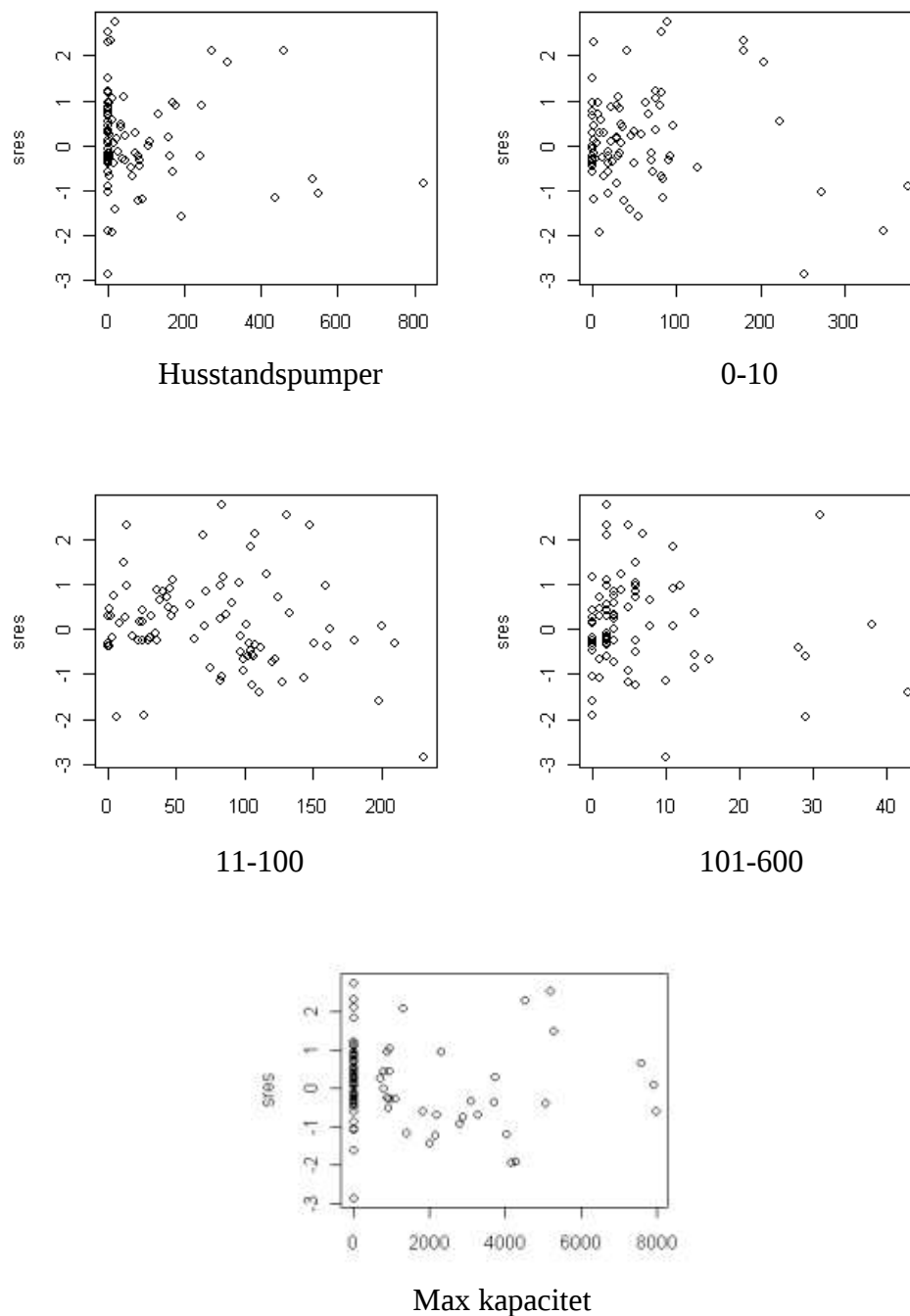
Modellen angiver dermed, at en husstandspumper koster 6.628 kr. at drive om året, en pumpestation i intervallet 0-10 m³/t koster 13.891 kr. at drive om året, en pumpestation i intervallet 11-100 m³/t koster 24.337 kr. at drive om året, en pumpestation i intervallet 101-600 m³/t koster 102.864 kr. at drive om året og en pumpestation i intervallet 601-max m³/t koster 597 kr. pr. m³/t at drive pr. m³/t.

Kvalitetskontrol af omkostningsækvivalenten

Resultaterne i tabel 11 er rimelige, idet B-værdierne indebærer, at driftsomkostningerne stiger ved et højere antal pumpestationer og driftsomkostningerne ved at drive en stor pumpestation er højere end driftsomkostningerne ved at drive en mindre pumpestation. Dertil kommer, at alle koefficienterne er signifikante.

Der er herudover kontrolleret for, om betingelserne for at bruge regressionsanalysens resultater er opfyldt for modellen. Kontrollen foretages ved at plotte de standardiserede fejled mod de observerede værdier af de fem intervaller af pumpestationer. Disse plots fremgår i figur 11 nedenfor:

Figur 11: Fejlplot for pumpestationer



Plottene viser umiddelbart ingen tegn på problemer, idet fejleddene fordeler sig pænt omkring 0. Resultaternes robusthed ville naturligvis være højere, hvis der var et stort antal observationer i hvert interval for pumpestationerne. De to største intervaller har relativt få observationer, hvilket skyldes, at kun få forsyninger har store pumpestationer.

Det har været nødvendigt at fjerne enkelte observationer med stor indflydelse på B-værdierne. Dette er målt med Cook's afstand som beskrevet i afsnittet "*Ekstreme observationer og observationer med stor indflydelse*" ovenfor.

Modellens baggrund

Forsyningssekretariatet har indsamlet data for 13 variable for costdriveren pumpestationer. Disse variable er:

- Antal husstandspumper
- Antal pumpestationer med kapacitet 0-10 m³/t
- Kapacitet af pumpestationer i kategorien 0-10 m³/t
- Antal pumper i med kapacitet 11-100 m³/t
- Kapacitet af pumpestationer i kategorien 11-100 m³/t
- Antal pumpestationer i kategorien 101-300 m³/t
- Kapacitet af pumper i kategorien 101-300 m³/t
- Antal pumper med kapacitet 301-600 m³/t
- Kapacitet af pumper i kategorien 301-600 m³/t
- Antal pumper med kapacitet 601-1000 m³/t
- Kapacitet af pumper i kategorien 601-1000 m³/t
- Antal pumper med kapacitet 1000-max m³/t
- Kapacitet af pumper i kategorien 1000-max m³/t

Der er blevet testet en række alternative regressioner ved at sammensætte antallet af pumpestationer og pumpestationernes kapaciteter på forskellige måder. Blandt andet er der udført regressioner med flere og færre intervalinddelinger for antallet af pumpestationer. Herudover er udført regressioner på baggrund af pumpestationernes samlede kapaciteter med flere forskellige intervalinddelinger. Derudover er der udført en regression, som tester for, hvorvidt der er faste omkostninger ved at drive pumpestationer.

Regressionerne udført på baggrund af pumpestationernes kapaciteter giver ikke intuitive og fortolkningsmæssige gode resultater. Når kapaciteterne inddeles efter kapacitetsintervaller er det ikke muligt at verificere resultaterne intuitivt. Det betyder, at resultaterne fortolkningsmæssigt ikke understøttes af en klar og fornuftig sammenhæng. Derfor blev der udført en regression udelukkende på antallet af husstandspumper og de øvrige pumpestationers samlede kapacitet. Denne model er meget simpel, hvorfor den ikke kan tage højde for, om forsyningerne driver mange små pumpestationer, eller om forsyningerne driver færre store pumpestationer.

Resultaterne af de gennemførte regressioner peger på, at det er regressionen i ligning (22), der forklarer selskabernes driftsomkostninger ved at drive pumpestationer bedst. Denne model er også den mest robuste. I denne regression har det været nødvendigt at slå antallet af pumpestationer i intervallet fra 100-300 m³/t og antallet af pumpestationer i intervallet 300-

600 m³/t sammen til en samlet kategori på 100-600 m³/t. Derudover har det været nødvendigt at slå antallet af pumpestationer i intervallet fra 600-1000 m³/t og antallet af pumpestationer i intervallet max m³/t sammen til en samlet kategori på max m³/t (over 600 m³/t). Pumpestationer over 600 m³/t opgøres med kapaciteten. På den måde tager modellen højde for at pumpestationer i denne kategori kan variere meget.

Regressionen i ligning (22) er blandt andet valgt, fordi det er denne regression, som bedst tager hensyn til selskabernes individuelle forhold vedrørende pumpestationer. Opgørelsen på antal pumper indenfor et givent interval vil eksempelvis fange, hvis et selskab har to små pumpestationer i stedet for en stor pumpestation. Overordnet set følger denne model forventningen om at det er dyrere at drive et større antal små pumpestationer end at drive et færre antal store pumpestationer. Dette kan eksempelvis illustreres ved, at det ifølge denne model vil koste 146.022 kr. (6*24.337) om året at drive seks pumpestationer med en kapacitet på 100 m³/t. Til sammenligning vil det koste 102.864 kr. om året at drive én pumpestation med en kapacitet på 600 m³/t. Modellen kan derfor håndtere selskabernes forskellige befolkningstætheder i forsyningsområderne.

Regnvandsbassiner

Driftsomkostningerne ved at drive regnvandsbassiner forventes i udgangspunktet at afhænge lineært af antallet af regnvandsbassiner og regnvandsbassinernes størrelse.

Forsyningssekretariatet vurderer, at den mest velegnede model til at beskrive de samlede driftsomkostninger ved at drive regnvandsbassiner er ligning (24), hvor Y er omkostningerne forbundet med regnvandsbassiner, og X₁ er antal bassiner:

$$(23) \quad Y = B_1 X_1$$

B-værdien udtrykker enhedsomkostningen for et regnvandsbassin.

Resultaterne af modellen fra ligning (23) fremgår af tabel 12 nedenfor:

Tabel 12: Regressionsanalysens resultater, regnvandsbassiner

Variabel	B-værdi	Spredning	t-værdi	Pr(> t)
Antal Bassiner	13.523	2.101 (1616,8)	6,438 (8,38)	1,52e-08 (5,31e-12)
Antal observationer: 68 Justeret R ² = 0,373				

Tal i parenteser angiver robuste værdier

Omkostningsækvivalenten for regnvandsbassiner

Den endelige model for omkostningsækvivalenten er således beskrevet ved modellen i ligning (23). Den endelige omkostningsækvivalent for

regnvandsbassiner består dermed af antallet af regnvandsbassiner, som angivet i ligning (24).

$$(24) \quad Y = 13.523X_1$$

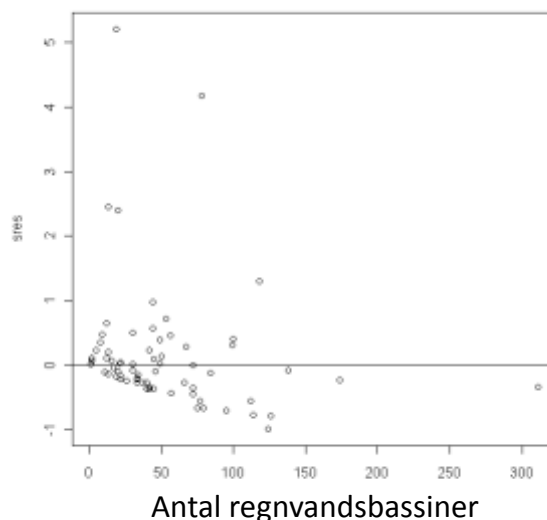
Forsyningerne får således forklaret driftsomkostningerne i forbindelse med at drive regnvandsbassiner ud fra, hvor mange regnvandsbassiner den enkelte forsyning har. Ovenstående resultat illustrerer, at når en forsyning driver et ekstra regnvandsbassin, så stiger de årlige driftsomkostninger med 13.523 kr.

Kvalitetskontrol af omkostningsækvivalenten

Resultaterne i tabel 12 er rimelige, idet B-værdien som forventet er positiv, hvilket betyder, at driftsomkostningerne stiger med antallet af regnvandsbassiner. Desuden er koefficienten signifikant.

Der er herudover kontrolleret for, hvorvidt betingelserne for at bruge regressionsanalysens resultater er opfyldt for modellen. Kontrollen foretages ved at plote de standardiserede fejlede mod de observerede værdier af antal regnvandsbassiner. Dette plot fremgår af figur 12 nedenfor:

Figur 12: Fejledeplot for regnvandsbassiner



Plottet udviser en hvis grad af heteroskedasticitet. Derfor er der beregnet robuste t-værdier, som tager hensyn til, at variansen af fejleddene er uens. Beregningen viser, at der stadig er en signifikant sammenhæng imellem målere og omkostninger jf. tabel 12.

Der er ikke nogen enkelte observationer, der har stor indflydelse på B-værdien, hvorfor der ikke er fjernet observationer som outliers. Dette måles med Cook's afstand, som beskrevet i afsnittet "Ekstreme observationer og observationer med stor indflydelse" ovenfor.

Modellens baggrund

Forsyningssekretariatet har indsamlet data for to variable for costdriveren regnvandsbassiner - antal regnvandsbassiner og det samlede areal af regnvandsbassinerne. Indsamlingen af data er foretaget på baggrund af en forventning om, at begge variable har indflydelse på størrelsen af de samlede omkostninger for en spildevandsforsynings regnvandsbassiner.

En indledende kontrol viser dog, at der er en høj korrelation mellem antal regnvandsbassiner og det samlede areal af regnvandsbassiner. Korrelationen er 0,87, hvilket er meget højt. Den høje korrelation betyder, at der vil være en høj sandsynlighed for multikollinearitet i en regression, hvor begge disse variable indgår. Det vil sige, at antal bassiner og arealet af bassiner forklarer den samme andel af variationen i driftsomkostningerne forbundet med at drive regnvandsbassiner. Begge variable kan derfor ikke umiddelbart medtages i omkostningsækvivalenten for regnvandsbassiner.

Forsyningssekretariatet har dog undersøgt, hvorvidt forholdet mellem arealet af bassinerne og antal bassiner kan anvendes. Det vil sige, det er undersøgt, hvorvidt de gennemsnitlige omkostninger per bassin kan forklares af den gennemsnitlige størrelse af et bassin for hver forsyning. Denne model havde dog en meget ringe forklaringsgrad samtidig med, at intuitionen bag er usikker. Forsyningssekretariatet vurderer derfor, at dette ikke er en anvendelig model.

Forsyningssekretariatet vurderer desuden, at antal bassiner er det mest præcise mål til at beskrive driftsomkostningerne forbundet med regnvandsbassiner. Endelig har Forsyningssekretariatet analyseret, hvorvidt der er væsentlige faste omkostninger ved at drive regnvandsbassiner. Regressionen gav ikke gode resultater. Modellen i ligning (24) indeholder derfor ikke en fast omkostning.

Spildevandsbassiner

Driftsomkostningerne ved at drive spildevandsbassiner forventes i udgangspunktet at afhænge lineært af antallet af regnvandsbassiner og regnvandsbassinernes størrelse.

Den mest velegnede model til at beskrive de samlede driftsomkostninger ved at drive spildevandsbassiner er ligning (25), hvor Y angiver de årlige driftsomkostningerne ved at drive spildevandsbassiner og X_1 er den samlede volumen af spildevandsbassinerne:

$$(25) \quad Y = B_1 X_1$$

B-værdien udtrykker enhedsomkostningerne for hvert spildevandsbassin.

Resultaterne af modellen fra ligning (25) præsenteres i tabel 13 nedenfor:

Tabel 13: Regressionsanalysens resultater, spildevandsbassiner

Variabel	B-værdi	Spredning	t-værdi	Pr(> t)
Samlet volumen	19,737	3,821 (3,61)	5,165 (5,46)	2,55e-06 (8,24e-7)
Antal observationer: 65 Justeret $R^2 = 0,2832$				

Tal i parenteser angiver robuste værdier

Omkostningsækvivalenten for spildevandsbassiner

Omkostningsækvivalent ved at drive spildevandsbassiner har følgende udseende, jf. tabel 13:

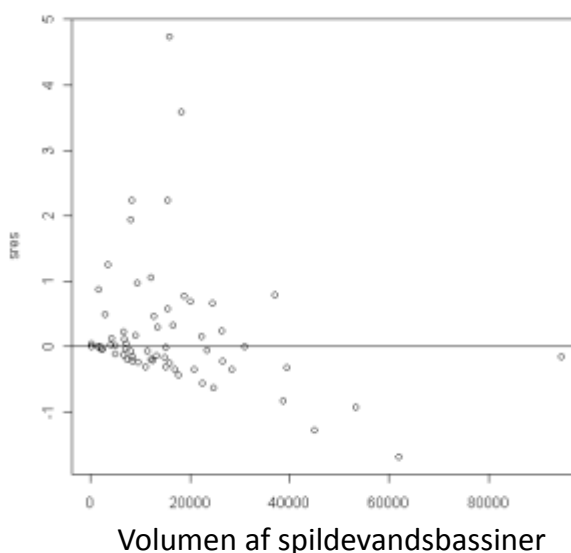
$$(26) \quad Y = 19,74X_1$$

Kvalitetskontrol af omkostningsækvivalenten

Resultaterne i tabel 13 er rimelige, idet B-værdien følger den økonomiske intuition, at driftsomkostningerne stiger med den samlede volumen af spildevandsbassinerne. Herudover er koefficienten signifikant.

Herudover er der kontrolleret for, hvorvidt betingelserne for at bruge regressionsanalysens resultater er opfyldt for modellen. Kontrollen foretages ved at plote de standardiserede fejlede mod de observerede værdier af antal regnvandsbassiner. Dette plot fremgår af figur 13 nedenfor:

Figur 13: Fejledeplot for spildevandsbassiner



Plottet udviser en hvis grad af heteroskedasticitet. Der er derfor beregnet robuste t-værdier, som tager hensyn til, at variansen af fejleddene er uens. Beregningen viser, at der stadig er en signifikant sammenhæng imellem målere og omkostninger jf. tabel 12.

Der har været nødvendigt at fjerne fire enkelte observationer, der har stor indflydelse på B-værdien. Dette måles med Cook's afstand som beskrevet i afsnittet *"Ekstreme observationer og observationer med stor indflydelse"* ovenfor.

Modellens baggrund

Forsyningssekretariatet har indsamlet data for to variable for costdriveren spildevandsbassiner. Indsamlingen af data er foretaget på baggrund af en forventning om, at begge variable har indflydelse på størrelsen af de samlede omkostninger tilknyttet en forsynings spildevandsbassiner både hver for sig og samlet.

En indledende kontrol viser, at korrelationen mellem antal spildevandsbassiner og det samlede volumen af spildevandsbassiner er 0,44. Det er derfor ikke umiddelbart noget til hinder for at begge variable kan medtages i omkostningsækvivalenten for spildevandsbassiner. Regressionsanalyserne giver dog ikke signifikante resultater, når begge variable medtages. Ligesom det var tilfældet med regnvandsbassiner, er det undersøgt, hvorvidt de gennemsnitlige omkostninger per bassin kan forklares af den gennemsnitlige volumen af et bassin for hver forsyning. Denne model har dog en ringe forklaringsgrad og intuitionen bag var usikker. Forsyningssekretariatet vurderer derfor, at dette ikke ville være en anvendelig model.

Forsyningssekretariatet har derfor undersøgt, hvorvidt antal bassiner eller volumen af spildevandsbassinerne beskriver driftsomkostningerne ved spildevandsbassinerne bedst. Analyserne viser, at det kun volumen af spildevandsbassiner der er signifikant, når begge variable inddrages. Det er derfor Forsyningssekretariatets vurdering, at det samlede volumen af spildevandsbassiner er det mest præcise mål til at beskrive driftsomkostningerne forbundet med spildevandsbassiner.

Endelig har Forsyningssekretariatet analyseret, hvorvidt der er væsentlige faste omkostninger ved at drive spildevandsbassiner. Intuitivt burde driftsomkostningerne ved at drive spildevandsbassiner afhænge af hvor store (eller hvor mange) spildevandsbassiner en forsyning driver. Forklaringsgraden (R^2) i denne model er meget lav og F-testet angiver at både B_0 og B_1 ligeså godt kunne være nul, hvilket bekræfter intuitionen om, at modellen ikke skal indeholde faste omkostninger.

Minirensesanlæg

Driftsomkostningerne ved at drive minirensesanlæg forventes i udgangspunktet at afhænge lineært af antallet af minirensesanlæg og minirensesanlæggenes kapacitet.

Forsyningssekretariatet har vurderet, at den mest velegnede model til at beskrive de samlede driftsomkostninger ved at drive minirensesanlæg er

ligning (27), hvor Y angiver de årlige driftsomkostningerne ved at drive spildevandsbassiner, og X_1 er antallet af minirenseanlæg:

$$(27) \quad Y = B_1 X_1$$

B-værdien udtrykker enhedsomkostningerne for et minirenseanlæg.

Resultaterne af modellen fra ligning (27) fremgår af tabel 14 nedenfor:

Tabel 14: Regressionsanalysens resultater, minirenseanlæg

Variabel	B-værdi	Spredning	t-værdi	Pr(> t)
Antal	2.540	812,4	3,127	0,00474
Antal observationer: 24 Justeret $R^2 = 0,2677$				

Omkostningsækvivalenten for minirenseanlæg

Omkostningsækvivalenten ved at drive minirenseanlæg har følgende udseende, jf. tabel 14:

$$(28) \quad Y = 2.540 X_1$$

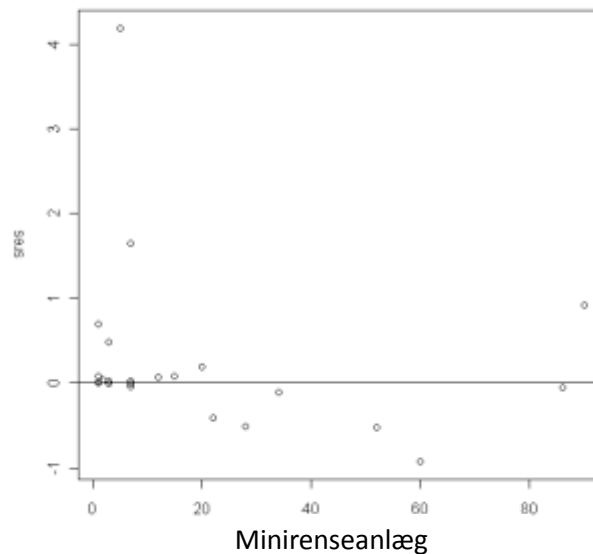
Ovenstående resultat betyder, at når antallet af minirenseanlæg stiger med en, så stiger de årlige driftsomkostninger med 2.540 kr. om året.

Kvalitetskontrol af omkostningsækvivalenten

Resultaterne i tabel 14 er rimelige, idet B-værdien som forventet er positiv, hvilket betyder, at driftsomkostningerne stiger med det samlede antal af minirenseanlæggene. Desuden er koefficienten signifikant.

Der er herudover kontrolleret for, hvorvidt betingelserne for at bruge regressionsanalysens resultater er opfyldt for modellen. Kontrollen foretages ved at plotte de standardiserede fejllid mod de observerede værdier af antal regnvandsbassiner. Dette plot fremgår af figur 14 nedenfor:

Figur 14: Fejlplot for minirenselanlæg



Plottet i figur 14 nedenfor viser en umiddelbar tilfældig fordeling omkring nul, hvilket betyder, at normalfordelingsbetingelsen er opfyldt. Yderligere lader der ikke til at være tegn på heteroskedasticitet. Der er en enkelt observation, der falder uden for intervallet $[-1,96; 1,96]$, men det er under 5 % af observationerne, og dermed hvad der kan forventes. Det skal dog bemærkes, at der kun er 24 observationer i datasættet for minirenselanlæg.

Der har været nødvendigt at fjerne to enkelte observationer, der har stor indflydelse på B-værdien. Dette måles med Cook's afstand som beskrevet i afsnittet *"Ekstreme observationer og observationer med stor indflydelse"* ovenfor.

Modellens baggrund

Forsyningssekretariatet har indsamlet data for to variable for costdriveren minirenselanlæg - antal minirenselanlæg og den samlede kapacitet på minirenselanlæggene. Indsamlingen af data er foretaget på baggrund af en forventning om, at begge variable har indflydelse på størrelsen af de samlede omkostninger tilknyttet en spildevandsforsynings minirenselanlæg både hver for sig og samlet.

En indledende kontrol viser dog, at der er en høj korrelation mellem antal minirenselanlæg og den samlede kapacitet på minirenselanlæggene. Korrelationen er 0,98, hvilket er meget højt. Den høje korrelation betyder, at en regression der indeholder begge variable med høj sandsynlighed vil opleve multikollinearitet. Det vil sige, at antal minirenselanlæg og den samlede kapacitet af minirenselanlæggene forklarer den samme andel af variationen i driftsomkostningerne forbundet med minirenselanlæg. Begge variable kan derfor ikke medtages i omkostningsækvivalenten for minirenselanlæg.

Forsyningssekretariatet har derfor undersøgt, hvorvidt det er antal minirensanlæg eller kapaciteten af minirensanlæggene der beskriver driftsomkostningerne forbundet med minirensanlæggene bedst.

Forholdet mellem kapaciteten af minirensanlæggene og antal minirensanlæg varierer meget fra forsyning til forsyning. Det vil sige, at en samlet kapacitet på 30 PE kan dække over et minirensanlæg på 30 PE eller 10 minirensanlæg på 3 PE. På baggrund af det indberettede data er det Forsyningssekretariatets vurdering, at forskellen på driftsomkostningerne på minirensanlæg afhænger mest af antallet af minirensanlæg og mindre af den samlede kapacitet af rensanlæggene. Denne model er i stand til at håndtere, at nogle selskaber har mange små minirensanlæg, hvilket i nogen grad er sammenfaldende med hvor tyndt befolket forsyningsområdet er.

Forsyningssekretariatet bemærker dog også, at driftsomkostningerne henført til minirensanlæg udgør en meget lille andel af de samlede fordelte driftsomkostninger. Forsyningssekretariatet bemærker desuden, at der er meget få observationer om minirensanlæg. Det er kun 26 forsyninger, der har indberettet anvendelige oplysninger om minirensanlæg. Forsyningssekretariatet har derfor undersøgt muligheden for at anvende et simpelt gennemsnit til at beskrive driftsomkostningerne. Resultatet af dette er dog, at det simple gennemsnit relativt tildeler den samme andel af omkostningerne til hver forsyning som den foreslåede model.

Endelig har Forsyningssekretariatet testet for, om der er væsentlige faste omkostninger ved at drive minirensanlæg. Denne analyse viser, at B_0 (de faste omkostninger) ikke er signifikant. Det vil sige, at B_0 ikke er med til at beskrive de samlede driftsomkostninger forbundet med at drive minirensanlæg i væsentlig grad, hvorfor B_0 er udeladt af modellen.

Rensanlæg

Driftsomkostningerne ved at drive rensanlæg forventes i udgangspunktet at afhænge lineært af den behandlede spildevandsmængde, behandlingsmetoden og zoneplaceringen.

Forsyningssekretariatet har vurderet, at den mest velegnede model til at beskrive de samlede driftsomkostninger ved at drive rensanlæg er ligning (29), hvor Y angiver driftsomkostningerne ved at rensningen spildevand. X 'erne er gennemsnittet af rensanlæggets belastning og kapacitet målt i PE

- X_1 er rensanlæg med M-rensning,
- X_2 er rensanlæg med MK-rensning,
- X_3 er rensanlæg med MBN-rensning,
- X_4 er rensanlæg med MBNK-rensning,
- X_5 er rensanlæg med MBNKD-rensning beliggende i landzone og
- X_6 er rensanlæg med MBNKD-rensning beliggende i byzone:

$$(29) \quad Y^{0,336} = B_1 X_1^{0,2284} + B_2 X_2^{0,2284} + B_3 X_3^{0,2284} + B_4 X_4^{0,2284} + B_5 X_5^{0,2284} + B_6 X_6^{0,2284}$$

B-værdierne afspejler enhedsomkostningerne pr. PE for hver rensemetode.

Der er i modellen fastsat en stordriftsfordel på $0.6798 = 0,2284/0,336$ ved at drive renseanlæg. Denne er baseret på Box-Cox metoden. Denne metode er udført iterativt, fordi det kun er muligt at beregne, hvordan Y variabelen skal transformeres. Det betyder, at det er nødvendigt at stille ligningen omvendt op hver anden gang beregningen foretages. Metoden foreslår således en ny transformation af Y parameteren, hver gang der er lavet en transformation af Y. Beregningen gentages frem til der forekommer en konvergens.

Hensynet til stordriftsfordelen betyder, at det er nødvendigt at transformere data for at sikre at sammenhængen er lineær. Dette gøres ved at opløfte både Y og X'erne i forskellige potenser, jf. afsnittet vedr. transformationer ovenfor.

I tabel 15 nedenfor præsenteres resultaterne af modellen i ligning (29).

Tabel 15: Regressionsanalysens resultater, renseanlæg

Variabel	B-værdi	Spredning	t-værdi	Pr(> t)
M	11,8830	1,8807	6,318	5,3e-10
MK	14,7191	0,5829	25,250	<2e-16
MBN	14,9187	0,6394	23,334	<2e-16
MBNK	15,1809	0,2884	52,631	<2e-16
MBNKD i landzone	16,0802	0,1953	82,319	<2e-16
MBNKD i byzone	16,3323	0,1786	91,452	<2e-16
Antal observationer: 497 Justeret $R^2 = 0,9747$				

Omkostningsækvivalenten for renseanlæg

Det er en særlig situation for costdriveren renseanlæg, at der for hvert renseanlæg skal beregnes et netvolumenbidrag. Derfor opstilles der seks enkelte udtryk for omkostningsækvivalenten renseanlæg. Bemærk, at der herudover ligges et enkelt beløb på 373,65 til hvert selskabs netvolumen, der har renseanlæg, uanset om de har et eller flere renseanlæg.

$$(30) \ Y = 1.581,92X_1^{0,6798} + 2.991,14X_2^{0,6798} + 3.113,49X_3^{0,6798} + 3.279,19X_4^{0,6798} + 3.891,82X_5^{0,6798} + 4.076,24X_6^{0,6798} + 373,65$$

Den endelige omkostningsækvivalent for renseanlæg består således af typen af renseanlæg, delvist af placeringen af renseanlægget i landzone og byzone og en korrektionsfaktor.

Modellen angiver dermed, at et selskab der har renseanlæg, har et fast årlig drift på 373,65 kr. samt at der er stordriftsfordele på renseanlæg. De angivne PE-værdier bliver derfor opløftet i 0,6798 inden der ganges en faktor på. For et renseanlæg med mekanisk rensning (M) er faktoren 1.581,92, for et renseanlæg med mekanisk og biologisk rensning (MB) er faktoren 2.991,14, for et renseanlæg med mekanisk og biologisk rensning samt enten kemisk rensning eller denitrifikation (MBN) er faktoren 3.113,49, for et renseanlæg med mekanisk rensning, biologisk rensning og nitrifikation samt enten kemisk rensning eller denitrifikation (MBNK) er faktoren 3.279,19, for et renseanlæg beliggende i landzone med mekanisk, biologisk og kemisk rensning samt nitrifikation og denitrifikation (MBNKD) er faktoren 3.891,82 og for et renseanlæg i byzone med mekanisk, biologisk og kemisk rensning samt nitrifikation og denitrifikation (MBNKD) er faktoren 4.079,24.

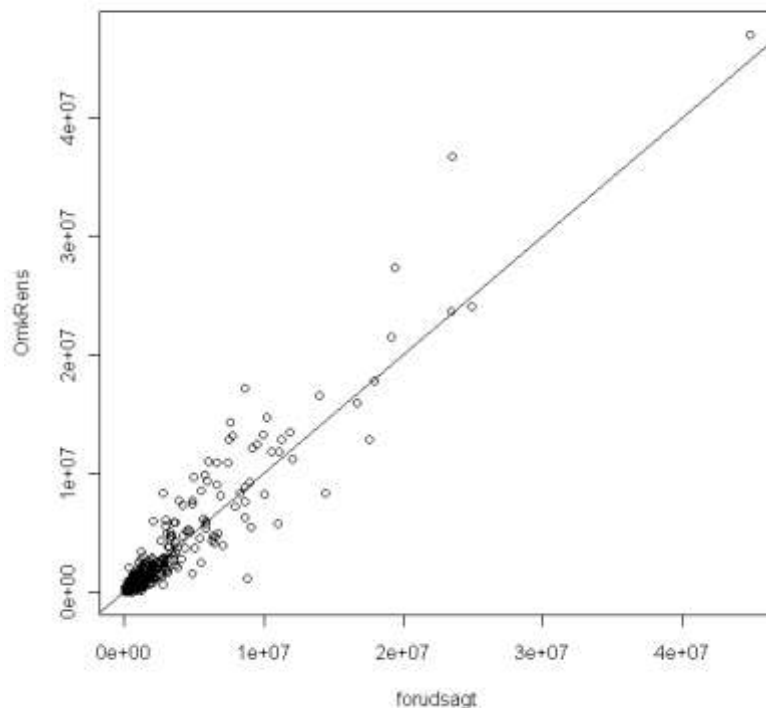
Korrektionsfaktoren beregnes som $KF = \sigma^2 = 373,6489$.

Kvalitetskontrol af omkostningsækvivalenten

Resultaterne i tabel 15 er rimelige, idet B-værdierne som forventet er positive, hvilket betyder, at driftsomkostningerne stiger, jo flere renseprocesser spildevandet skal igennem og når renseanlægget er placeret i landzone for visse renseprocesser. Herudover er koefficienterne signifikante.

Figur 15 nedenfor viser et plot af de forudsagte driftsomkostninger og de faktiske driftsomkostninger:

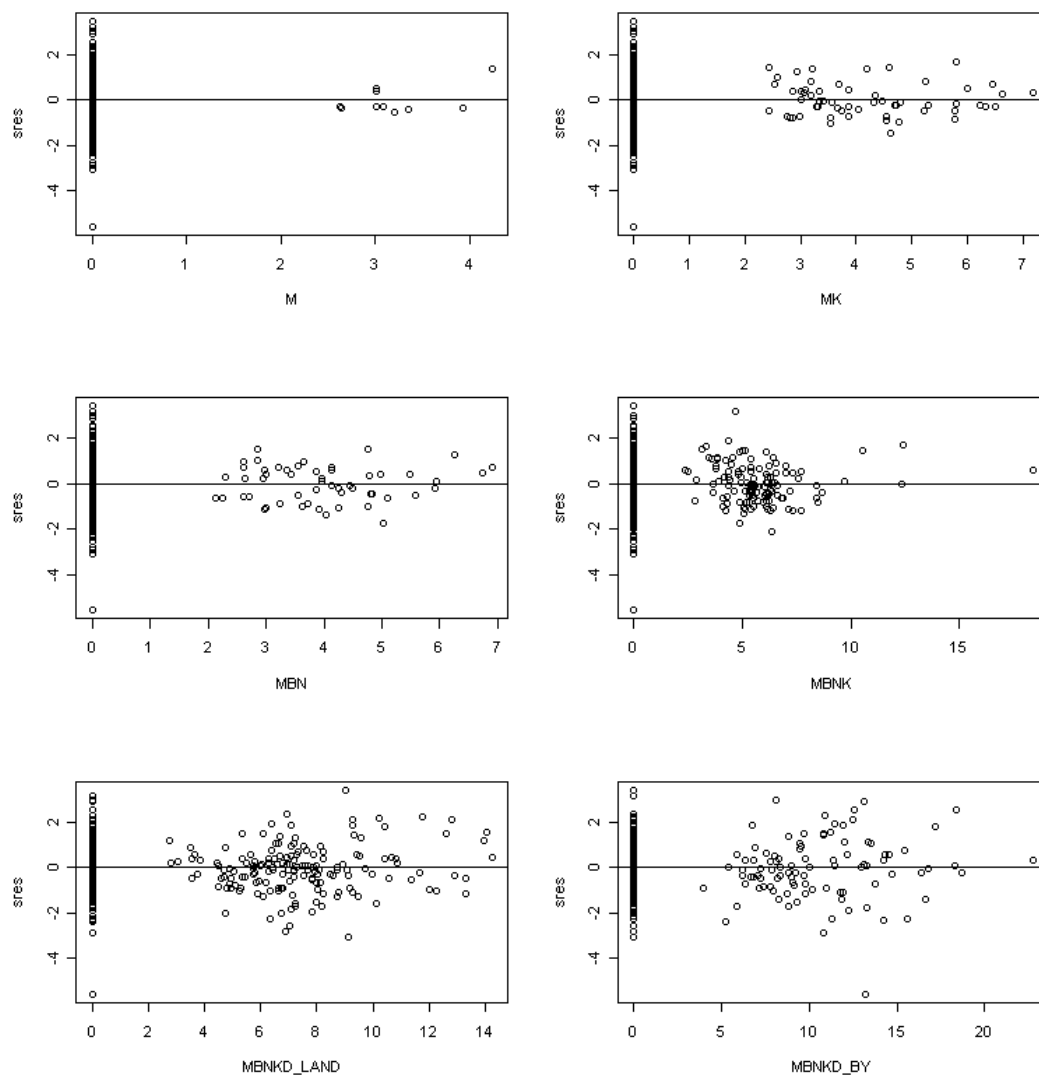
Figur 15: De faktiske omkostninger og de forudsagte omkostninger



Plottet i figur 15 ser meget pænt ud, idet der er et stort sammenfald mellem modellens forudsagte driftsomkostninger og forsyningernes faktiske driftsomkostninger.

Herudover er der kontrolleret for, hvorvidt betingelserne for at bruge regressionsanalysens resultater er opfyldt for modellen. Betingelserne kontrolleres ved at plote de standardiserede fejllid mod de observerede værdier af gennemsnittet af renseanlæggets kapacitet og belastning målt i PE for henholdsvis renseanlæg af type M, renseanlæg af type MK/MBK/MB, renseanlæg af type MBN, renseanlæg af type MBNK/MBND, renseanlæg af type MBNKD beliggende i landzone og renseanlæg af type MBNKD beliggende i byzone. Plottene fremgår af figur 16 nedenfor:

Figur 16: Fejllsplots for renseanlæg



Plottene i figur 16 viser de standardiserede fejlls. Plottene viser en umiddelbar tilfældig fordeling omkring nul, hvilket betyder, at normalfordelingsbetingelsen er opfyldt.

Yderligere lader der ikke til at være tegn på heteroskedasticitet. Der er nogle få observationer, der falder uden for intervallet $[-1,96; 1,96]$, men det er umiddelbart under 5 pct. af observationerne, og dermed hvad der kan forventes.

Der er ikke nogen enkelte observationer, der har stor indflydelse på B-værdien. Dette måles med Cook's afstand som beskrevet i afsnittet *"Ekstreme observationer og observationer med stor indflydelse"* ovenfor.

Modellens baggrund

Forsyningssekretariatet har indsamlet data for 13 variable for costdriveren pumpestationer. Disse variable er:

- Renseanlæggets dimensionerede organiske kapacitet i PE
- Renseanlæggets faktiske belastning i PE baseret på prøver af BOD5
- Indløbsmængden på renseanlægget målt i m³/år
- Renseanlæggets zoneplacering (Kun land og by ifølge kommuneplanen)
- Typen af rensning: M, B, K, N, D
- Mængde af fosfor i indløb i kg
- Mængde af fosfor i udløb i kg
- Mængde af kvælstof i indløb i kg
- Mængde af kvælstof i udløb i kg
- Angivelse af renseanlæggets udlederkrav mht. (BOD, COD, kvælstof og fosfor)

Indsamlingen af data er foretaget på baggrund af en forventning om, at disse variable har indflydelse på størrelsen af de samlede driftsomkostninger ved at drive renseanlæg.

En indledende kontrol viser, at syv af variablene er højt korrelerede. Det er renseanlæggets kapacitet, renseanlæggets belastning, renseanlæggets vandindløbsmængde, fosfor i indløb, fosfor i udløb, kvælstof i indløb og kvælstof i udløb. Den parvise korrelation mellem disse syv variable varierer fra 0,79 til 0,98. Denne høje korrelation betyder, at en regression med alle syv variable med høj sandsynlighed vil medføre multikollinearitet, idet disse syv variable forklarer den samme andel af variationen i driftsomkostningerne forbundet med renseanlæg. Der skal derfor udvises forsigtighed ved inddragelse af enhver kombination af disse syv variable i omkostningsækvivalenten for renseanlæg.

Forklaringen på korrelationen kan være, at måleenheden PE kan måles på flere måder.⁴ Renseanlæggets belastning kan for eksempel måles ud fra fosfor og kvælstof i indløb og udløb. Forsyningssekretariatet har på baggrund af dette kun valgt at inddrage renseanlæggets belastning og renseanlæggets kapacitet ud af de 7 korrelerede variable.

Af de øvrige seks variable beskriver fire af disse variable udlederkrav til henholdsvis COD, BOD, fosfor og kvælstof. Forsyningssekretariatet har analyseret de udlederkrav, der er skærpet i forhold til de nationale udlederkrav angivet i § 17 i bekendtgørelse om spildevandstilladelser m.v. efter miljøbeskyttelseslovens kapitel 3 og 4.⁵

Regressionsanalyserne viser, at udlederkravene ikke har signifikant betydning for størrelsen af driftsomkostningerne. Dette kan skyldes flere ting, f.eks. at data om udlederkrav ikke er godt nok eller at mange forsyninger renser spildevandet bedre end krævet, og derfor har omkostninger der svarer til, at de havde skærpet udlederkrav.

⁴ Ved 1 personækvivalent (PE) forstås 21,9 kg organisk stof/år målt som det biokemiske iltforbrug (BI5), 4,4 kg total kvælstof/år eller 1,0 kg total fosfor/år. (§ 4, stk. 5 i BEK nr. 1448 af 11/12/2007)

⁵ BEK nr. 1448 af 11/12/2007

Forsyningssekretariatet har derfor vurderet, at udlederkrav ikke skal medtages i omkostningsækvivalenten for renseanlæg.

De sidste to variable er zoneplaceringen af renseanlægget samt typen af rensning. Disse to variable er medtaget i den endelige omkostningsækvivalent for renseanlæg.

Samlet set indgår der fire variable i omkostningsækvivalenten for renseanlæg; belastning, kapacitet, typen af rensning og zoneplacering.

Forsyningssekretariatet har valgt, at gennemsnittet af et renseanlægs belastning og kapacitet i PE, skal indgå som parameter for renseanlæggets størrelse. Dette valg er foretaget på baggrund af omkostningsækvivalenten for renseanlæg i benchmarkingen for 2012. Forsyningssekretariatet har undersøgt hvorvidt en 40/60- eller en 30/70-fordeling forklarer flere af omkostningerne end en 50/50-fordeling. Analyserne viser dog, at forskellen er så lille, at der ikke er noget, der peger på, at den ene model er bedre end den anden. Forsyningssekretariatet har derfor valgt at bibeholde et almindeligt gennemsnit af renseanlæggets belastning og kapacitet.

Typen af renseanlæg indgår i omkostningsækvivalenten således, at jo flere renseprocesser et renseanlæg har, jo dyrere er det at drive. Zoneplaceringen indgår også i den endelige omkostningsækvivalent. Dog indgår denne kun delvist. Analyserne viser, at renseanlæg med en til fire renseprocesser er dyrere at drive i landzone end i byzone. Data viser dog, at disse renseanlæg oftest er beliggende i landzone og meget få er derfor placeret i byzone. Dette betyder, at der tillægges mere værdi til landzonen, da der er flere observationer her. Forsyningssekretariatet har derfor valgt, at zoneplaceringen ikke skal inddrages for renseanlæg med op til fire renseprocesser. For renseanlæg med fem renseprocesser (MBNKD) viser analyserne dog, at renseanlæg placeret i byzone er dyrere end renseanlæg placeret i landzone. Dette hænger fint sammen med intuitionen og resultatet afhænger sandsynligvis også af, at disse renseanlæg er næsten ligeligt fordelt på landzone og byzone. Der er derfor et solidt datagrundlag her.

Datagrundlaget viser tydeligt, at der er markante stordriftsfordele ved renseanlæg. For at anvende en lineær regression skal sammenhængen dog være lineær. Forsyningssekretariatet har derfor foretaget en transformation af data, således at sammenhængen er lineær. Denne transformation er vist i ligning (30).

Omkostningsækvivalenten for renseanlæg består af gennemsnittet af renseanlæggets kapacitet og belastning i PE, typen af rensning der foretages (M, MK, MBN, MBNK/MBND, MBNKD). Såfremt renseanlægget har MBNKD-rensning afhænger omkostningsækvivalenten også af renseanlæggets zoneplacering.

Endelig har Forsyningssekretariatet testet for, om der er væsentlige faste omkostninger ved at drive renseanlæg. Denne analyse viser, at B_0 (de faste omkostninger) ikke er signifikant. Det vil sige, at B_0 ikke er med til at beskrive de samlede driftsomkostninger forbundet med at drive renseanlæg i væsentlig grad, hvorfor B_0 er udeladt af modellen.

Slambehandling

Driftsomkostningerne ved at behandle slam forventes i udgangspunktet, at afhænge lineært af mængden af slam der skal behandles, og af den metode der bruges til at behandle slammet.

Den mest velegnede model til at beskrive de samlede driftsomkostninger ved at behandle slam er ligning (31), hvor Y angiver driftsomkostningerne forbundet med slambehandling, X_1 er tons tørstof A-slam og X_2 er tons tørstof B- og C-slam:

$$(31) \quad Y = B_1X_1 + B_2X_2$$

B -værdierne afspejler enhedsomkostningerne for hver behandlingstype.

resultaterne af modellen i ligning (31), der bedst beskriver forsyningernes driftsomkostninger ved at behandle slam præsenteres i tabel 16 nedenfor..

Tabel 16: Regressionsanalysens resultater, slambehandling

Variabel	B-værdi	Spredning	t-værdi	Pr(> t)
Tons A	3.965,40	402,8	9,846	1,45e-13
Tons B og Tons C	4.747,70	481,1	9,869	1,33e-13
Antal observationer: 55 Justeret $R^2 = 0,8047$				

Omkostningsækvivalenten for slambehandling

Omkostningsækvivalenten ved at behandle slam har følgende udseende, jf. tabel 16:

$$(32) \quad Y = 3.965,4X_1 + 4.747,7X_2$$

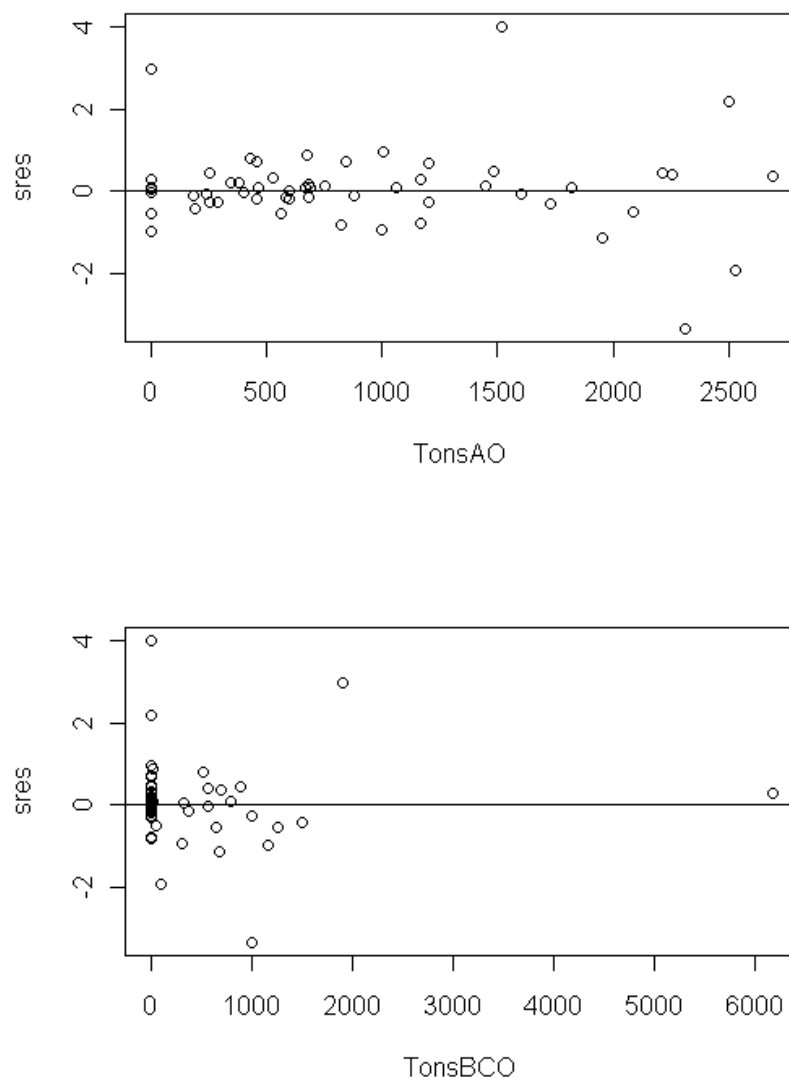
Dermed angiver modellen, at det koster 3.965,40 kr. om året at behandle et tons tørstof A-slam og at det koster 4.747,70 kr. om året at behandle et tons tørstof B- eller C-slam.

Kvalitetskontrol af omkostningsækvivalenten

Resultaterne i tabel 16 er tilfredsstillende, idet B-værdierne er positive, og dermed følger forventningen om, at driftsomkostningerne stiger med en større mængde slam, der skal behandles. Det er også som forventet dyrere at behandle B- eller C-slam i stedet for A-slam. Desuden er koefficienterne signifikante.

Der er herudover kontrolleret for, hvorvidt betingelserne for at bruge regressionsanalysens resultater er opfyldt for modellen. Kontrollen foretages ved at plote de standardiserede fejllid mod de observerede værdier af tons tørstof A-slam og tons tørstof B- og C-slam.

Figur 17: Fejllidplots for slambehandling



Plottene i figur 17 viser, at de standardiserede fejllid umiddelbart fordeler sig tilfældigt omkring nul, hvilket betyder, at normalfordelingsbetingelsen er opfyldt.

Yderligere lader det ikke til, at der er tegn på heteroskedasticitet. Der er nogle få observationer, der falder uden for intervallet $[-1,96; 1,96]$, men det er umiddelbart under 5 pct. af observationerne og dermed hvad der kan forventes.

Det har været nødvendigt at fjerne to enkelte observationer, der har stor indflydelse på B-værdien. Dette måles med Cook's afstand som beskrevet i afsnittet "*Ekstreme observationer og observationer med stor indflydelse*" ovenfor.

Der er dog yderligere 13 observationer, som adskilte sig i så væsentlig grad, at det forstyrrede sammenhængen i residualplottet, jf. nedenfor. Disse observationer er også fjernet fra analysen.

Modellens baggrund

Forsyningssekretariatet har indsamlet data for tre variable for costdriveren slambehandling; antal tons tørstof af A-, B- og C-slam. Indsamlingen af data er foretaget på baggrund af en forventning om, at disse variable har indflydelse på størrelsen af en spildevandsforsynings samlede omkostninger til slambehandling.

En indledende kontrol viser, at den indbyrdes korrelation mellem de tre variable er lav. Alle tre variable kan derfor som udgangspunkt indgå på lige fod i omkostningsækvivalenten for slambehandling.

Regressionsanalyserne har dog efterfølgende vist, at når alle tre variable indgår, bliver B-slam dyrere at behandle end C-slam. Forsyningssekretariatet har derfor valgt at sammenlægge B-slam og C-slam til én kategori. Dette er valgt, da regressionerne viser, at parametrene for B-slam og C-slam ligner hinanden mere end parameteren for A-slam. Der er derudover en forventning om, at der er stor forskel på omkostningerne til A-slam og B-slam og mindre forskel på omkostningerne til B-slam og C-slam.

Regressionsanalyserne viste efter sammenlægningen af de to kategorier en klar sammenhæng i data. Dog var der nogle uhensigtsmæssigheder i residualplottene, som der ikke umiddelbart kunne findes en forklaring på. Forsyningssekretariatet har valgt at fjerne de observationer fra datasættet som forårsager uhensigtsmæssighederne i plottene, da de kan betegnes som ekstreme observationer, da de havde en stor indflydelse på resultaterne.

Endelig har Forsyningssekretariatet testet for, om der er væsentlige faste omkostninger ved at drive renseanlæg. Denne analyse viser, at B_0 (de faste omkostninger) ikke er signifikant. Det vil sige, at B_0 ikke er med til at beskrive de samlede driftsomkostninger forbundet med slambehandling i væsentlig grad, hvorfor B_0 er udeladt af modellen.

Kunder

Driftsomkostningerne ved at håndtere forsyningernes kunder forventes at afhænge af antallet af målere. Dette skyldes, at flere kunder må forventes at give flere kundeforholdsmæssige omkostninger, og at antallet af målere afspejler, hvor mange kunder forsyningerne har.

Den mest velegnede model til at beskrive de samlede omkostninger ved at håndtere kunder er fundet at være den lineære ligning (33), hvor Y angiver de årlige driftsomkostninger forbundet ved at håndtere kunder og X_1 angiver antallet af målere:

$$(33) \quad Y = B_1 X_1$$

B-værdien afspejler enhedsomkostningen ved at håndtere en måler.

Resultaterne af modellen i ligning (33), der bedst beskriver forsyningernes driftsomkostninger ved at håndtere kunder præsenteres i tabel 17 nedenfor:

Tabel 17: Regressionsanalysens resultater, kunder

Variabel	B-værdi	Spredning	t-værdi	Pr(> t)
Målere	120,843	7.622 (9,22)	15.85(13,10)	<2e-16 (2,2e-16)
Antal observationer:80 Justeret $R^2=0,76$				

Omkostningsækvivalenten for kunder

Omkostningsækvivalenten ved at håndtere kunder har følgende udseende, jf. tabel 17:

$$(35) \quad Y = 120,8 X_1$$

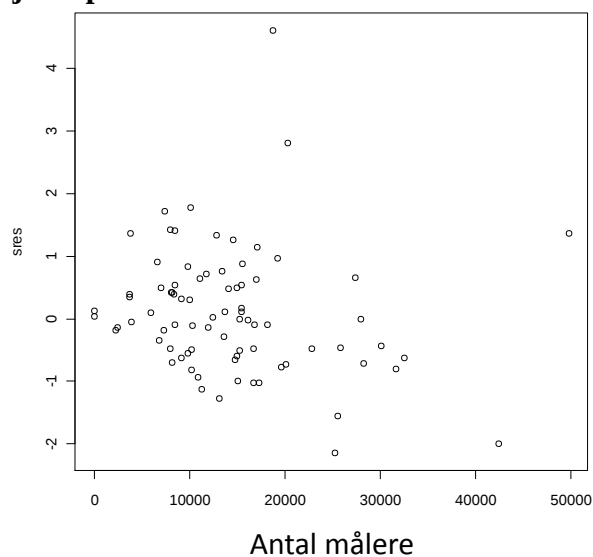
Dermed angiver modellen, at en måler koster 120,8 kr. at drive om året.

Kvalitetskontrol af omkostningsækvivalent

Resultatet i tabel 17 er tilfredsstillende, idet B-værdien som forventet er positiv, hvilket betyder at driftsomkostningerne stiger med et større antal målere. Desuden ses det af tabel 17, at parameteren i modellen er signifikant.

Herudover er der kontrolleret for, hvorvidt betingelserne for at bruge regressionsanalysens resultater er opfyldt for modellen. Kontrollen foretages ved at plotte de standardiserede fejle mod de observerede værdier målerne. Dette plot fremgår i figur 18 nedenfor:

Figur 18: Fejlledsplot for målere



Plottet udviser en vis grad af heteroskedasticitet. Der er derfor beregnet robuste t-værdier som tager hensyn til, at variansen af fejleddene er uens. Beregningen viser, at der stadig er en signifikant sammenhæng imellem målere og omkostninger jf. tabel 17.

Det har været nødvendigt at fjerne 4 observationer med stor indflydelse på B-værdierne. Dette måles med Cook's afstand som beskrevet i afsnittet *"Ekstreme observationer og observationer med stor indflydelse"* ovenfor.

Modellens baggrund

Forsyningssekretariatet bemærker, at sekretariatet ikke har testet andre sammenhænge end den ovenstående, idet selskaberne i 2012 kun er blevet bedt om at indberette måleroplysninger. Der er ingen tegn på stordrift.