

OPEX- netvolumen- mål for drikke- vandsselska- berne

Teknisk beregning af omkostningsækvivalenter til brug for OPEX-del af benchmarkingmodellen for 2019 og frem

September 2018

Version 2



**OPEX-netvolumenmål for drikkevandsselskaberne
Version 2**

Konkurrence- og Forbrugerstyrelsen

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

OPEX-netvolumenmål for drikkevandsselskaberne er udarbejdet af
Konkurrence- og Forbrugerstyrelsen.

September2018

Indhold

Kapitel 1	
Indledning.....	4
1.1 Arbejdet op til nu.....	4
1.2 Ændringer i denne version	5
Kapitel 2	
Metodebeskrivelse	6
2.1 Indledning.....	6
2.2 Mindste kvadraters metode	6
Kapitel 3	
Bestemmelse af omkostningsækvivalenter	13
3.1 Indledning.....	13
3.2 Boringer.....	13
3.3 Vandværker	16
3.4 Trykforøger- og trykreduktionsstationer	20
3.5 Rentvandsledninger og stik.....	24
3.6 Målere og kunder.....	28
3.7 Generel administration.....	31
Kapitel 4	
Samlet netvolumenmål.....	36
4.1 Indledning.....	36
4.2 Kvalitetskontrol	36

Kapitel 1

Indledning

Dette papir er en teknisk gennemgang af arbejdet med at revidere omkostningsækvivalenterne til brug for benchmarkingen af alle drikkevandsselskaber med en debiteret vandmængde over 800.000 m³. Omkostningsækvivalenterne skal bruges i forbindelse med benchmarkingen til de økonomiske rammer for 2019 og frem.

Omkostningsækvivalenterne skal udtrykke sammenhængen mellem selskabernes driftsomkostninger og costdrivere. Vi vil beskrive de statistiske metoder, der ligger til grund for udregningen af omkostningsækvivalenterne samt præsentere resultatet for hver enkelt costdriver.

1.1 Arbejdet op til nu

Til brug for udregningen af de nye omkostningsækvivalenter har vi haft brug for en stor mængde af data fra selskaberne.

I 2014 udarbejdede vi en konteringsvejledning og en costdrivervejledning. I løbet af 2015 har selskaberne indsamlet data om både driftsomkostninger og costdrivere på baggrund af de to vejledninger. Hen over sommeren 2016 indberettede 60 drikkevandsselskaber¹ deres driftsomkostninger for 2015 fordelt på costdriverne samt deres costdrivere i 2015.

Vi har efterfølgende arbejdet med at beregne de nye omkostningsækvivalenter. Til at starte med var det vigtigt for os, at datagrundlaget var af så god kvalitet som muligt. Vi satte derfor en omfattende kvalitetssikring i gang, hvor vi var i kontakt med selskaber, hvor der var tvivl om deres indberetning. Derefter kunne vi gå i gang med de endelige beregninger af omkostningsækvivalenterne, der har resulteret i dette papir.

Høring

I høringsperioden fra den 23. november til den 7. december 2017 har vi modtaget jeres høringssvar. Høringssvarene har givet stof til eftertanke, men har ikke medført ændringer. Høringsnotatet, som kort opsummerer de indkomne høringssvar samt vores kommentarer hertil ligger på vores hjemmeside sammen med dette papir².

¹ Svarende til 80 pct. af drikkevandsselskaberne med en debiteret vandmængde over 800.000 m³.

² <https://www.kfst.dk/vandtilsyn/benchmarking/revidering-af-opex-delen-af-bm-modellen/>

1.2 Ændringer i denne version

Vi er blevet opmærksomme på, at der har været uklarhed omkring definitionen af debiteret vandmængde. Derfor er omkostningsækvivalenten for generel administration blevet genberegnet i denne version. Data for debiteret vandmængde i 2015 er genindhentet, for at sikre at det er den samlede debiterede vandmængde, som ligger til grund for beregningen af omkostningsækvivalenten for generel administration. Ændringerne er beskrevet i afsnit 3.7.

Kapitel 2

Metodebeskrivelse

2.1 Indledning

I dette kapitel vil vi gennemgå de generelle statistiske overvejelser, vi har gjort os ved beregningerne af omkostningsækvivalenterne. Vi vil overordnet beskrive den statistiske metode ("mindste kvadraters metode"), der benyttes til at bestemme parametrene i en statistisk model.

Derudover vil vi beskrive, hvilke overvejelser der gøres i forbindelse med udregningen af en omkostningsækvivalent i form af outlieranalyse og transformation af data, og hvordan vi efterfølgende vurderer, om vi er kommet frem til den bedst mulige model.

2.2 Mindste kvadraters metode

Formålet med OPEX-modellen er at finde sammenhæng mellem omkostninger og tilhørende underliggende forhold. Sagt på en anden måde ønsker vi at forklare, hvordan y ændres når x ændres. Til dette anvender vi regressionsanalyser. Mere specifikt er der benyttet en metode, der hedder "mindste kvadraters metode". I denne metode antages det, at der er en lineær sammenhæng mellem omkostninger og de underliggende forhold. Dette kan præsenteres i modellen nedenfor:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n + u \quad (1)$$

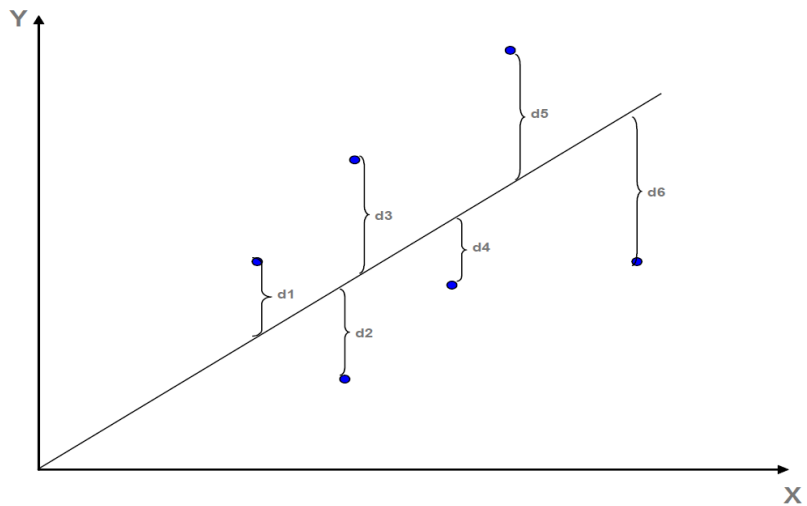
Her er X de underliggende forhold, Y er omkostningerne, n er antallet af underliggende forhold, som bruges til at forklare omkostningerne, β_0 er linjens skæring med y -aksen og $\beta_1 \dots \beta_n$ repræsenterer påvirkningen af et underliggende forhold på omkostningerne (Y), hvor alt andet er fastholdt.

Det sidste led u er et fejld. Fejleddet indeholder alt det, som ikke er forklaret ved de underliggende forhold. Antagelser omkring fejleddet er vigtige for validiteten af den model, som estimeres med mindste kvadraters metode. Dette uddybes i afsnittet om normalfordeling af fejld.

Med mindste kvadraters metode estimeres de værdier for β 'erne, som giver den statistisk bedste sammenhæng mellem omkostningerne og de underliggende forhold. Det er vigtigt at være opmærksom på, at regressionen (1) ikke er lig med alle de indberettede omkostninger, men at den derimod giver et gennemsnit for ændringerne i Y , når X ændres. Denne sammenhæng kan præsenteres som i Figur 2.1. Med mindste kvadraters metode findes den linje, som bedst forklarer sammenhængen mellem en række afhængige variable og en række uafhængige variable.

Den bedste linje betragtes i mindste kvadraters metode som den linje, der skaber de mindste afstande – ”mindste kvadrater” – mellem punkterne og linjen. Dette ses i Figur 2.1.

Figur 2.1 Mindste kvadraters metode



Ud fra dette princip kan de bedste værdier for β 'erne i (1) beregnes, således at vi får modellen:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_n X_n \quad (2)$$

Her repræsenterer ”hatten” over β 'erne, at det er de værdier, som er estimeret i modellen.

Signifikant model

For hvert underliggende forhold er det nødvendigt at undersøge, om der er en god sammenhæng mellem denne og omkostningerne. Det skal undersøges, om der er signifikans i modellen. I en regressionsanalyse betyder det, at det skal undersøges, om de underliggende forhold har en signifikant effekt på omkostningerne. I praksis undersøges det, om de estimerede $\hat{\beta}$ er signifikante.

Vi tager udgangspunkt i et signifikansniveau på 5 pct. Det vil sige, at hvis sandsynligheden for, at et $\hat{\beta}$ er lig med 0, er større end 5 pct., vil det som udgangspunkt blive afvist, at der er en sammenhæng mellem det givne underliggende forhold og omkostningerne. Dette vil som oftest medføre, at den valgte model tages op til genovervejelse og fx, at underliggende forhold lægges sammen, eller at data transformeres. Signifikansen beregnes på baggrund af antallet af observationer, størrelsen af $\hat{\beta}$ -værdien samt spredningen af observationerne. En stor spredning og få observationer vil svække signifikansen af det enkelte underliggende forhold.

I præsentationstabellerne, som fremgår i kapitel 3 med de estimerede modeller for hver cost-driver, fremgår både p- og t-værdierne. De præsenterer som udgangspunkt det samme, nemlig signifikansniveauet for den enkelte $\hat{\beta}$. De fundne t-værdier skal være så langt fra 0 som muligt. Som tommelfingerregel skal t-værdien være udenfor intervallet $[-1,96; 1,96]$ for at være signifikant. Intervallet for t-værdien ændres i forhold til antallet af frihedsgrader i modellen. Jo færre frihedsgrader, jo længere skal t-værdien være fra 0 for at være signifikant. Dette tager p-værdien højde for, hvorfor denne er nemmere at fortolke som et mål for estimatets signifikans. De fundne p-værdier skal være mindre end 0,05 (5 pct.) for, at sammenhængen accepteres som signifikant.

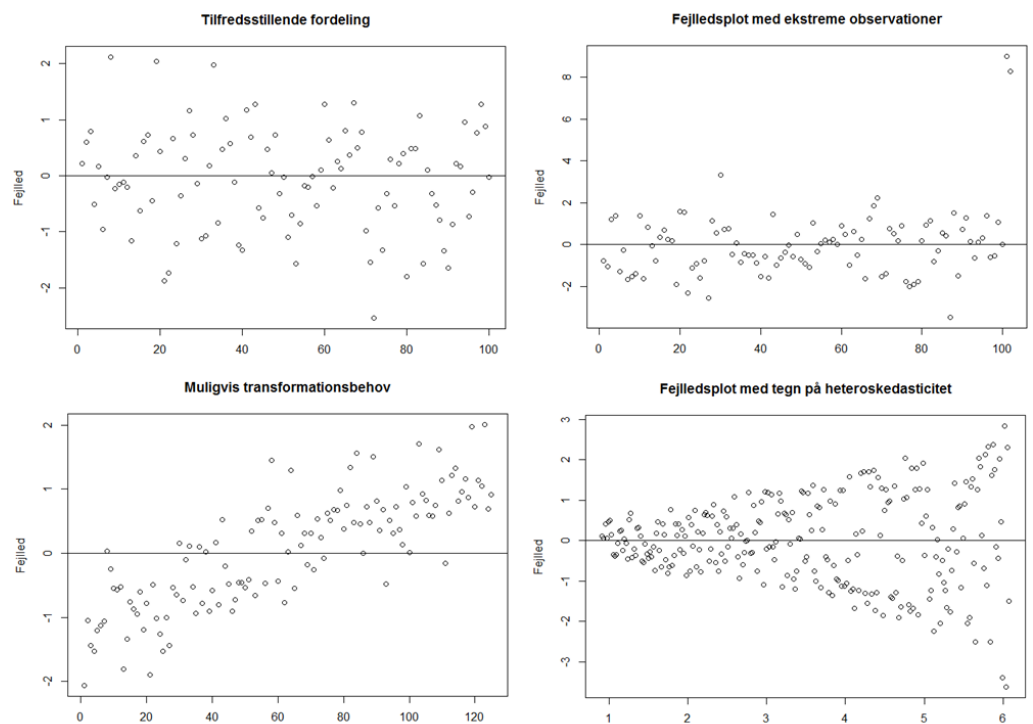
Signifikansniveauet sammenholdes med den økonomiske teori og intuition ved de enkelte costdrivere. Det betyder, at underliggende forhold, som viser sig ikke at være statistisk signifikante på et 5 pct.-niveau, kan overvejes inddraget i modellen, hvis det økonomisk intuitivt giver mening, at forholdet skal bruges til at forklare variationen i omkostningerne. Samtidigt kan signifikante underliggende forhold fjernes, hvis der er teoretiske forhold, som kræver det.

Normalfordeling af fejled

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

I Figur 2.2 er vist forskellige plots af fejled mod en uafhængig variabel. Fejleddene skal vise en tilfældig fordeling omkring 0 for at være normalfordelte. For alle costdrivere vil der i kapitel 3 blive præsenteret tilsvarende fejleddplot.

Figur 2.2 Fejleddplots



Ud fra fejleddplotsene er det muligt at se, om der er problemer med normalfordelingen af fejleddene. Der er problemer, hvis fejleddene ikke er tilfældigt fordelt omkring 0. Det er vigtigt, at antagelsen er opfyldt for at sikre, at den fundne $\hat{\beta}$ er middeleret³. Hvis mere end 5 pct. af fejleddene ligger udenfor intervallet $[-1,96; 1,96]$, kan der være problemer med antagelsen. Fejleddene påvirker den beregnede værdi af $\hat{\beta}$ således, at en højere værdi af fejleddene, givet

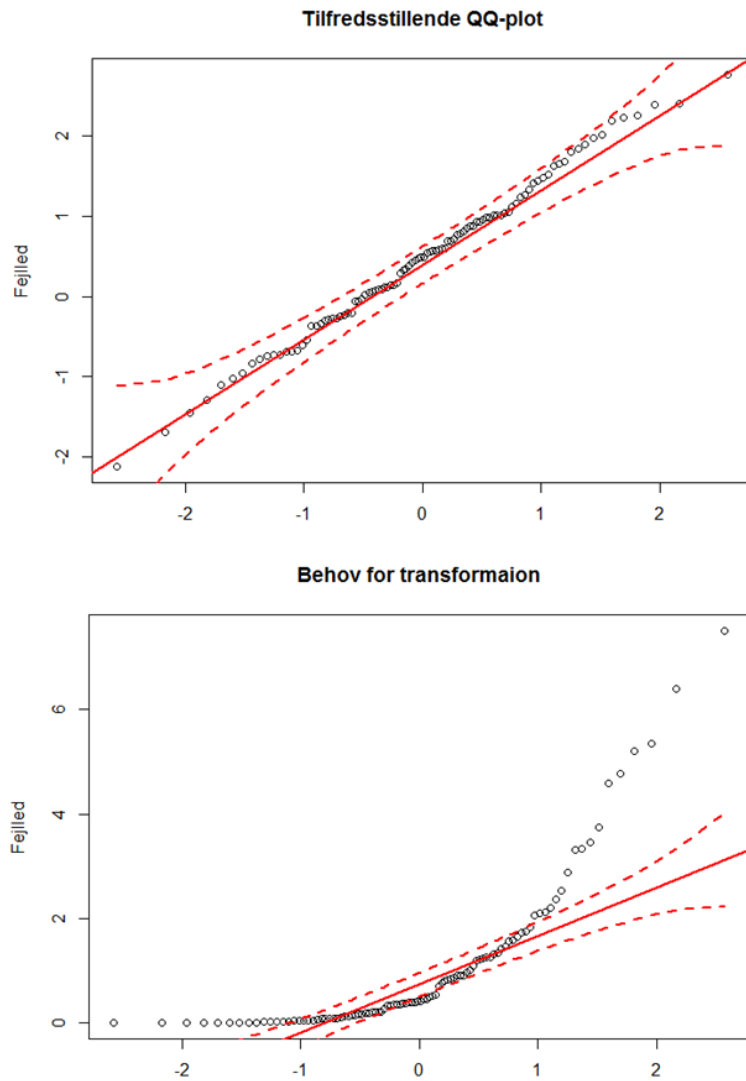
³ Med middeleret menes, at der er fundet frem til en værdi for $\hat{\beta}$, som i gennemsnit vil være lig den sande værdi. At $\hat{\beta}$ er middeleret betyder dermed, at vi har fundet den bedst mulige værdi.

værdien af det underliggende forhold, vil øge forskellen mellem $\hat{\beta}$ og β , dvs. øge forskellen mellem den estimerede værdi og den reelle værdi. I dette tilfælde vil $\hat{\beta}$ dermed blive skæv.

Yderligere kan der ud fra fejleddplotsene observeres, om der er et muligt problem med heteroskedasticitet. Heteroskedasticitet fremgår som et systematisk mønster i fejleddenes placering. Heteroskedasticitet påvirker ikke værdien af $\hat{\beta}$ ($\hat{\beta}$ er altså middelret på trods af heteroskedasticitet). Derimod påvirkes variansen af de beregnede $\hat{\beta}$. Dermed medfører heteroskedasticitet, at signifikansniveauet for $\hat{\beta}$ ændres. Det vil betyde, at vi risikerer at afvise eller acceptere forhold, som ikke burde være afvist eller accepteret. Det er også muligt at lave en statistisk test for heteroskedasticitet. Dette gøres ved hjælp af en såkaldt Breusch-Pagan test, som tester, hvorvidt der er en sammenhæng mellem fejleddene og hvert af de underliggende forhold. Hvis p-værdien i denne test er større end 0,05 afviser vi, at der er et problem med heteroskedasticitet i modellen. Yderligere er det muligt at beregne $\hat{\beta}$, som er robuste for heteroskedasticitet.

Til at undersøge fejleddenes fordeling anvendes også såkaldte QQ-plots, som er vist i Figur 2.3.

Figur 2.3 QQ-plots



QQ-plottet giver en anden visuel præsentation af fejleddenes fordeling i modellen. QQ-plottet viser den teoretiske værdi som findes, givet at modellen følger en normalfordeling. Hvis punkterne ligger på en lige linje, er det et udtryk for, at data følger normalfordelingen. I kapitel 3 er der for hver costdriver præsenteret et QQ-plot. Der er angivet et 95pct.-konfidensinterval for alle costdriverne. Dette ses som en guideline for, hvor tæt fordelingen er på en normalfordeling.

Ud fra et QQ-plot er det ligeledes muligt at se, hvorvidt det er nødvendigt at lave en transformation af data for at finde den model, som bedst beskriver data. Dette belyses i senere afsnit om transformation af data.

Outlieranalyse

Som udgangspunkt ønsker vi at inkludere så meget data som muligt til estimeringen af modellen for en costdriver. En større datamængde øger præcisionen af den estimerede model samt gør modellen mere repræsentativ for branchen. Til tider kan det dog være nødvendigt at fjerne enkelte observationer, hvis de har en u hensigtsmæssig stor påvirkning på $\hat{\beta}$.

Til at finde outliers anvender vi Cook's distance. Med Cook's distance beregnes et estimat for, hvor meget $\hat{\beta}$ vil ændre sig, hvis en enkelt observation fjernes, og $\hat{\beta}$ genberegnes. Cook's distance beregnes for alle observationerne og giver dermed et udtryk for, hvor stor en indflydelse den enkelte observation har på resultatet.

Der er ingen fast statistisk regel for, hvor grænseværdien for Cook's distance skal ligge. Vi har for alle costdrivere valgt en grænseværdi for Cook's distance på 0,5. Det betyder, at hvis Cook's distance for en enkelt observation er større end 0,5, så fjernes den og modellen genberegnes. Der bliver dog altid kun fjernet en enkelt observation af gangen. Det vil sige, at hvis der er flere observationer med en Cook's distance over 0,5, fjernes kun den observation med højest Cook's distance, hvorefter modellen genberegnes. Proceduren gentages indtil der ikke er observationer med en Cook's distance over 0,5.

Grænsen på 0,5 er sat ud fra et forsigtighedshensyn, da vi ikke ønsker at fjerne alt for mange observationer for dermed at risikere at finde en model, som ikke er repræsentativ for branchen. Mere restriktive grænser er undersøgt, men vi har vurderet, at disse vil have for stor påvirkning på et i forvejen begrænset datagrundlag.

Transformation af data

For at kunne bruge mindste kvadraters metode er det nødvendigt, at der findes en lineær sammenhæng mellem den afhængige variabel og de underliggende forhold. Det sker dog ofte, at data ikke umiddelbart viser dette. Der vil eksempelvis ikke være en lineær sammenhæng i data, hvis der er stordriftsfordele ved en costdriver. I disse tilfælde er det nødvendigt at lave en transformation af data. De mest udbredte transformationer er at tage logaritmen, kvadratroden eller andre potenser til variablerne i modellen.

Et eksempel på en transformation er en "log-log" transformation. Det betyder, at der laves en logaritmisk transformation af data for både den afhængige variabel (driftsomkostningerne) og de uafhængige variabel (de underliggende forhold). En log-log transformation kan bruges, når en sammenhæng kan beskrives ved en Cobb-Douglas funktion:

$$Y = \beta_0 X_1^{\beta_1} X_2^{\beta_2} \quad (3)$$

I dette tilfælde er det nødvendigt at lave en log-log transformation af data for at kunne estimere modellen ved mindste kvadraters metode. Det giver følgende lineære form:

$$\log(Y) = \log(\beta_0) + \beta_1 \log(X_1) + \beta_2 \log(X_2) \quad (4)$$

Med log-log transformationen er det dermed muligt at beregne β -værdierne, selvom sammenhængen i data ikke er lineær.

Om en transformation er nødvendig, og hvilken der er relevant, kan til dels ses ud fra plots af resultaterne fra en lineær model, dvs. fejleddplotsene og QQ-plottet – især QQ-plottet giver et overskueligt billede af dette. Derudover har vi brugt en deduktiv tilgang til transformationer, hvor forskellige transformationer er afprøvet på den enkelte costdriver for at finde den mest passende transformation.

For at undgå at estimere modeller, som mangler økonomisk intuition, men blot har et godt statistisk fit, har vores udgangspunkt været en lineær model. Samtidig er data kun blevet transformeret, såfremt det giver økonomisk intuitiv mening.

For at kunne anvende den transformerede model (4) til at beregne sammenhængen mellem driftsomkostningerne og de underliggende forhold, er det nødvendigt at tilbagetransformere variablerne i (4). Alt efter hvilken transformation af data der er anvendt, er der forskellige metoder.

For nogle transformationer gælder det dog, at det er nødvendigt at lave en korrektion af resultaterne. Dette skyldes, at de transformerede data skaber en normalfordeling af fejleddene, hvorfor fejleddene ikke er normalfordelte i datas oprindelige form. For at tage højde for den skævhed som det giver, er det nødvendigt at lave en korrektion i form af en korrektionsfaktor. Hvilken korrektionsfaktor der skal anvendes afhænger af, hvilken transformation der er anvendt på data.

Hvis der er lavet en logaritmisk transformation af data, ganges følgende korrektionsfaktor (KF) på hele højresiden af den valgte model:

$$KF = \exp\left(\frac{\sigma^2}{2}\right) \quad (5)$$

Exp betegner den naturlige eksponentialfunktion, og σ betegner spredningen af fejleddet. Korrektionsfaktoren ganges på den fundne tilbagetransformerede model.

Såfremt der er valgt en transformation for en given costdriver, er dette forklaret under de enkelte costdrivere i kapitel 3.

Valg af den bedste model

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

Overordnet set er modellerne vurderet ud fra nedestående tre kriterier:

1. Økonomisk intuition
2. Statistiske kriterier
3. Antallet af selskaber/observationer

Ad 1. Som nævnt tidligere, skal der være en økonomisk intuitiv baggrund for at antage den sammenhæng, som undersøges. Dette vil ofte føre til transformationer af data, men også fra valg af transformationer, hvis det ikke giver teoretisk mening.

Ad 2. Der er lavet en vurdering af modellernes statistiske performance for at vurdere modellen. Først er signifikansen for de enkelte underliggende forhold vurderet, som tidligere beskrevet. Hvis der ikke har været en betydelig signifikans af et underliggende forhold, har det

været nødvendigt at slå underliggende forhold sammen. Dette er gjort under hensynstagen til kriterie 1.

Til at vurdere modellen som helhed er anvendt forskellige statistiske redskaber.

R^2 fortæller hvor meget af variationen i den afhængige variabel, som er forklaret af de uafhængige variable. R^2 kan altså bruges til at sammenligne forskellige modeller af samme costdriver. Det er dog nødvendigt at være påpasselig med tolkningen af R^2 , når der laves transformationer af data, samt når skæringen i en model fjernes.

Normalfordelingen af fejleddene samt fit af data analyseres ud fra fejleddplots og QQ-plots, som forklaret tidligere, og sammenlignes modellerne imellem.

Ad 3. Til sidst er der lagt vægt på antallet af selskaber i den endelige model. Som tidligere beskrevet er outliers blevet fjernet, hvis de har haft en u hensigtsmæssig stor indflydelse på modellen (ved hjælp af Cook's distance). I vurderingen af den enkelte model er der taget hensyn til, at der ikke er blevet fjernet for mange selskaber ved outlieranalysen. Således er flest mulige selskaber med til at estimere modellen.

Datakvalitet

For at kunne lave de mest retvisende regressionsanalyser er det nødvendigt, at datakvaliteten er høj. Derfor er der indledningsvist lagt et stort arbejde i at kvalitetssikre de data, som selskaberne har indberettet.

I kvalitetssikringen er indberetningerne blevet sammenlignet med selskabernes indberetning til de økonomiske rammer for 2017. Såfremt der har været store afvigelser, er selskaberne blevet kontaktet og bedt om at redegøre for afvigelse rne. For costdrivere, som ikke har været indberettet i forbindelse med de økonomiske rammer 2017, har vi ikke anvendt en systematisk kvalitetssikring, da vi ikke har haft mulighed for at sammenligne data med tidligere indberetninger. I de tilfælde, hvor vi har fundet data, som vi ikke har vurderet var retvisende, er selskaberne blevet kontaktet eller udtaget som en outlier ved den givne costdriver. Desuden er selskaber, der ikke har indberettet til en given costdriver, ikke medtaget i beregningen af omkostningsækvivalenten for costdriveren.

Kapitel 3

Bestemmelse af omkostningsækvivalenter

3.1 Indledning

Vi har beregnet omkostningsækvivalenter for de seks costdrivere for drikkevandsselskaberne:

- » Boringer
- » Vandværker
- » Trykforøger- og trykreduktionsstationer
- » Rentvandsledninger og stik
- » Målere og kunder
- » Generel administration

Beregningen af de enkelte omkostningsækvivalenter bliver beskrevet i de følgende afsnit. Først beskrives costdriverens baggrund, som ligger til grund for valg af model. Derefter beskrives den valgte omkostningsækvivalent. Til slut beskrives den kvalitetskontrol, der er foretaget i forbindelse med beregningen af omkostningsækvivalenten.

3.2 Boringer

Modellens baggrund

Driftsomkostninger ved at drive boringer forventes umiddelbart at være påvirket af den samlede oppumpede vandmængde samt antallet af boringer. Vi har været nødsaget til samle den oppumpede vandmængde fra indvindingsboringer og den oppumpede vandmængde fra afværgeboringer, da det ikke har været muligt at finde en signifikant forskel på de to. Derudover er korrelationen mellem den samlede oppumpede vandmængde, det samlede antal boringer og længden af råvandsledninger så høj, at det kun er nødvendigt at medtage en af variablerne.

Den bedste model beskriver dermed de samlede omkostninger til boringer ud fra den samlede oppumpede vandmængde. Det er fundet, at den mest velegnede model til at beskrive sammenhæng mellem omkostninger og boringer er en udgave af en Cobb-Douglas-model, som beskrevet i ligning (3). Det er forsøgt både med og uden en konstantfaktor β_0 og det er vurderet, at den bedste model er uden konstantfaktor.

Y angiver de årlige driftsomkostninger til boringer, og X_1 angiver den samlede oppumpede vandmængde fra både indvindingsboringer og afværgeboringer.

$$Y = X_1^{\beta_1} \quad (6)$$

For at kunne estimere en model som udviser eksponentiel sammenhæng, er det nødvendigt at lave en logaritmisk transformation af modellen.

$$\log(Y) = \beta_1 \log(X_1) \quad (7)$$

De estimerede værdier for modellen er præsenteret i Tabel 1.

Tabel 1 Regressionsanalysens resultater for boringer

Variabel	Koefficient	Spredning	t-værdi	p-værdi
Log(Samlet oppumpet vandmængde)	0,9121	0,009280	98,2876	< 2 e-16 ***

Antal observationer: 58
Justeret $R^2 = 0,9940$

Anm.: Stjerneerne i kolonnen med p-værdier angiver, hvor signifikante de enkelte variable er. De variable, som er mest signifikante, har tre stjerner, hvorimod variable som ikke er signifikante, har nul stjerner.

Korrektionsfaktor

Når data transformeres, er det nødvendigt at lave en korrektion i forbindelse med tilbage-transformationen af den estimerede model.

Standardafvigelsen for fejleddene er $\sigma = 1,0254$. Dermed kan det udledes, at korrektionsfaktoren er 1,6917, hvilket skal ganges på højre side af ligningen.

Omkostningsækvivalenten for boringer

Omkostningsækvivalenten ved at drive boringer kan ses i ligning (8).

$$Y = 1,6917X_1^{0,9121} \quad (8)$$

Når parameteren β_1 er under 1, betyder det, at der er stordriftsfordele forbundet med omkostninger til boringer. Når mængden af den oppumpede vandmængde øges med 1 pct. stiger omkostningerne med 0,9121 pct.

Kvalitetskontrol af omkostningsækvivalenten

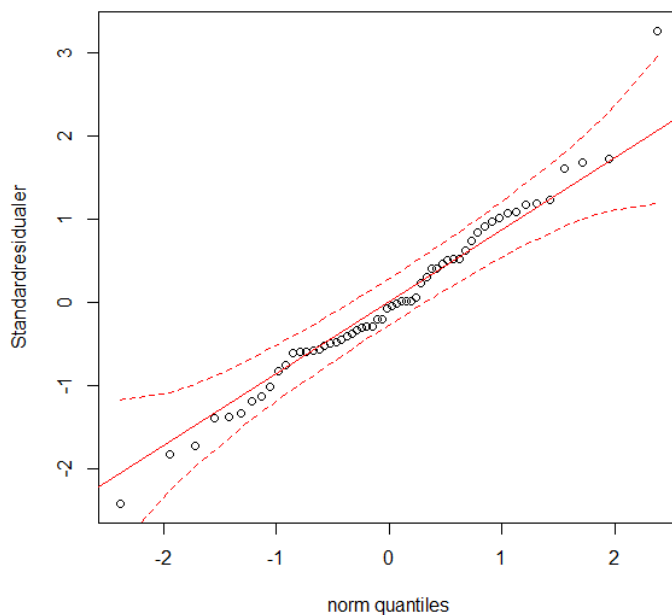
Tabel 1 angiver, at variabelen er signifikant ved et 5 pct. signifikansniveau. Det betyder, at der er stor statistisk sikkerhed i modellens resultater. Herudover ses det, at β -værdien er mindre en 1, hvilket betyder, at der er stordriftsfordele ved at oppumpe vand, hvilket vi vurderer stemmer overens med den økonomiske intuition.

HOFOR Vand Herlev A/S og Hørsholm Vand ApS er ikke med til at bestemme omkostningsækvivalenten for denne costdriver, da de ikke har indberettet underliggende forhold hertil.

Der er ikke fjernet nogen outliers på baggrund af Cook's distance. Det betyder, at alle selskaber har indflydelse på estimeringen af parameteren β_1 , men at ingen har alt for stor indflydelse.

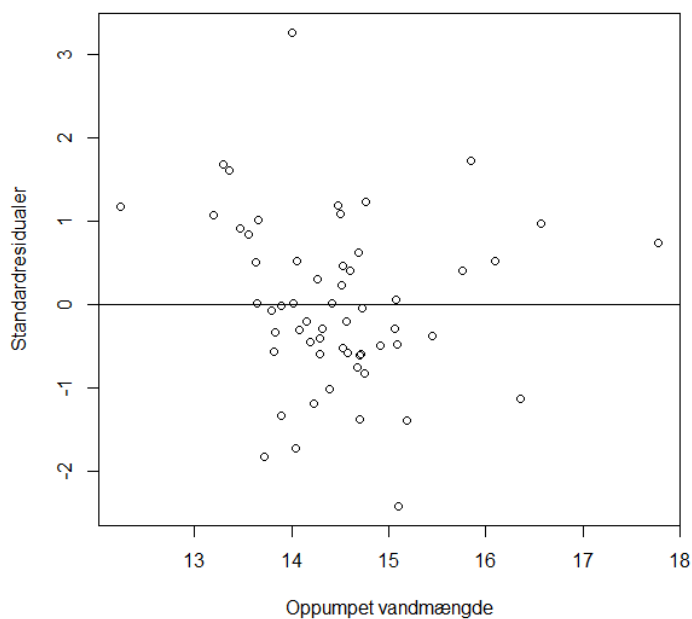
I Figur 3.1 ses det, at modellens fejlede er pænt normalfordelte, og ligger indenfor konfidensintervallet.

Figur 3.1 **QQ-plot for boringer**



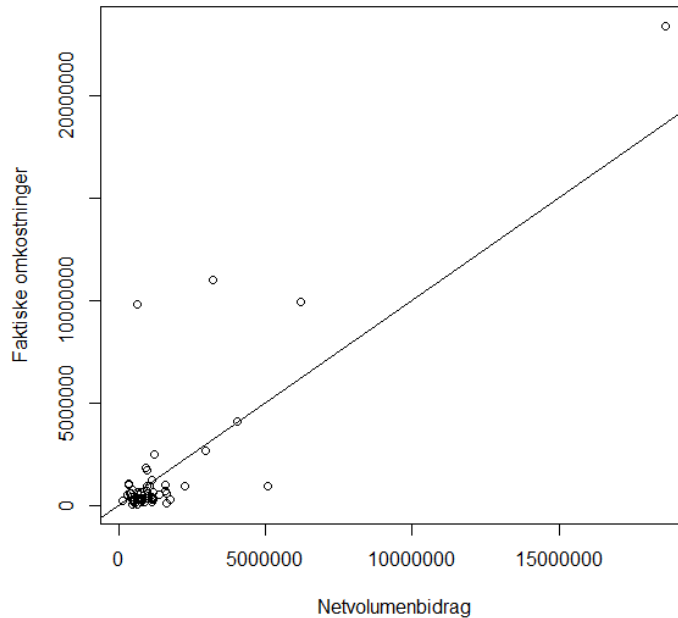
Dette kan ligeledes ses i Figur 3.2, hvor fejleddene er plottet mod de observerede værdier.

Figur 3.2 **Fejleddsplot for boringer**



Slutteligt viser Figur 3.3 modellens beregnede værdier sammenlignet med de observerede omkostninger. Der er få observationer, der ikke estimeres optimalt, men da alle selskaber har haft indflydelse på estimeringen, vurderer vi, at vi kan fortsætte med den valgte model.

Figur 3.3 Faktiske omkostninger mod netvolumenbidraget for boringer



3.3 Vandværker

Modellens baggrund

Driftsomkostninger ved at drive vandværker forventes at være påvirket af den udpumpede vandmængde og af typen af behandling, som vandet modtager inden den udpumpes. Typen af behandling er inddelt i tre typer. Type 1 - ingen behandling, type 2 - normal behandling og type 3 - avanceret behandling.

Det har været muligt, i indberetningen til den reviderede OPEX-model, at specificere hvilket avanceret behandling vandet har modtaget, inden det udpumpes. Der har dog kun været indberetninger nok af én enkelt type avanceret behandling, til at kunne udskille den fra de andre avancerede behandlinger, og det er regulering af pH. De andre typer af avanceret behandling er slået sammen til en samlet avanceret behandling.

Derudover har det vist sig, at der ikke har været tilstrækkeligt med indberetninger af type 1 til at lade den optræde alene i modellen. Derfor er type 1 (ingen behandling) og type 2 (normal behandling) slået sammen i modellen.

Der er, udover den valgte model, blevet testet en række alternative modeller. I disse modeller er der blandt andet blevet testet for andre sammensætninger af de underliggende forhold og behovet for at transformere data. Den valgte model har vist de stærkeste statistiske egenskaber og finder samtidig koefficienter, som giver økonomisk intuitiv mening i forhold til hinanden.

Vi har vurderet, at den mest velegnede model til at beskrive de samlede driftsomkostninger ved at drive vandværker er angivet i ligning (9).

Y angiver de årlige driftsomkostninger ved at drive vandværker, X_1 angiver den samlede udpumpede vandmængde, der har modtaget enten ingen behandling eller normal behandling, X_2 angiver den samlede udpumpede vandmængde, der har modtaget anden avanceret behandling end regulering af pH og X_3 angiver den samlede udpumpede vandmængde, der har modtaget regulering af pH. β -værdierne angiver enhedsomkostningerne for hvert af de underliggende forhold.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \quad (9)$$

I Tabel 2 præsenteres resultaterne af modellen fra ligning (9).

Tabel 2 Regressionsanalysens resultater for vandværker

Variabel	Koefficient	Spredning	t-værdi	p-værdi
Konstantled	285.172,4638	194.029,4612	1,470	0,148
Type1 og Type 2	1,2346	0,06521	18,933	< 2 e-16 ***
Type 3	1,9115	0,2547	7,505	1,24 e-09 ***
Regulering af pH	1,4292	0,2459	5,811	4,87 e-07 ***

Antal observationer: 52
Justeret $R^2 = 0,8842$

Anm.: Stjernerne i kolonnen med p-værdier angiver, hvor signifikante de enkelte variable er. De variable, som er mest signifikante, har tre stjerner, hvorimod variable som ikke er signifikante, har nul stjerner.

Omkostningsækvivalenten for vandværker

Omkostningsækvivalenten ved at drive vandværker kan ses i ligning (10).

$$Y = 285.172,4638 + 1,2346X_1 + 1,9115X_2 + 1,4292X_3 \quad (10)$$

Modellen angiver, at der er faste omkostninger på 285.172,4638 kr. ved at drive vandværker, og derefter koster det 1,2346 kr. per kubikmeter udpumpet vand der modtager ingen eller normal behandling, 1,9115 kr. per kubikmeter udpumpet vand der modtager en avanceret behandling andet end regulering af pH og 1,4292 kr. per kubikmeter udpumpet vand der bliver reguleret for pH.

Kvalitetskontrol af omkostningsækvivalenten

Tabel 2 angiver, at alle variable undtagen konstantleddet er signifikante ved et 5 pct. signifikansniveau. Det betyder, at der er stor statistisk sikkerhed i modellens resultater. Konstantleddet er ikke signifikant, men har en lav p-værdi. Vi har vurderet, at det giver intuitiv mening med et konstantled, og da modellens statistiske egenskaber er gode, har vi valgt at acceptere signifikansniveauet på konstantleddet. Det ses, at det er billigst per kubikmeter vand at udføre ingen eller normal behandling. Derefter er det lidt dyrere per kubikmeter vand at udføre regulering af pH og dyrest per kubikmeter vand at udføre en anden type af avanceret behandling.

HOFOR Vand Herlev A/S er ikke med til at bestemme omkostningsækvivalenten for denne costdriver, da de ikke har indberettet underliggende forhold hertil.

Ved brug af Cook's distance er der fjernet fire selskaber, som ellers ville have haft for stor indflydelse på modellens resultater. Disse selskaber er:

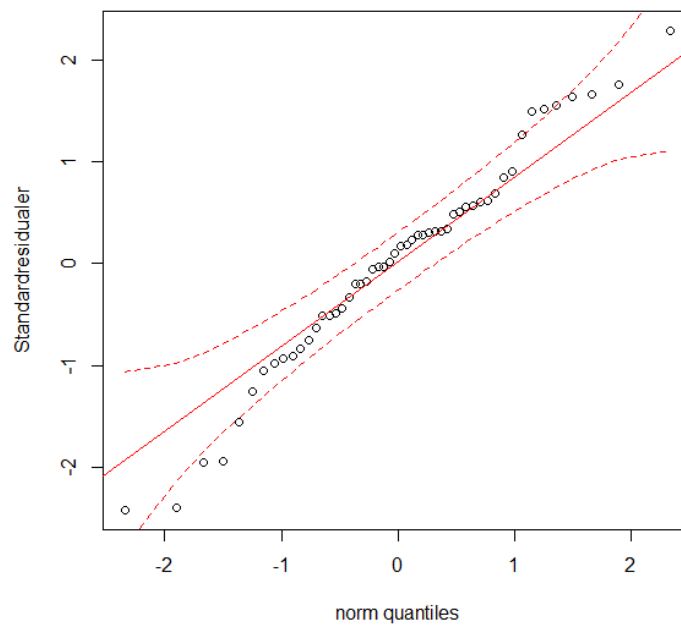
- » Forsyning Helsingør A/S
- » HOFOR Vand København A/S
- » Aarhus Vand A/S
- » Vandcenter Syd A/S

Derudover er der fjernet to selskaber ved manuelt at vurdere deres indflydelse på resultaterne. Disse selskaber er:

- » Frederikshavn Vand A/S
- » Køge Vand A/S

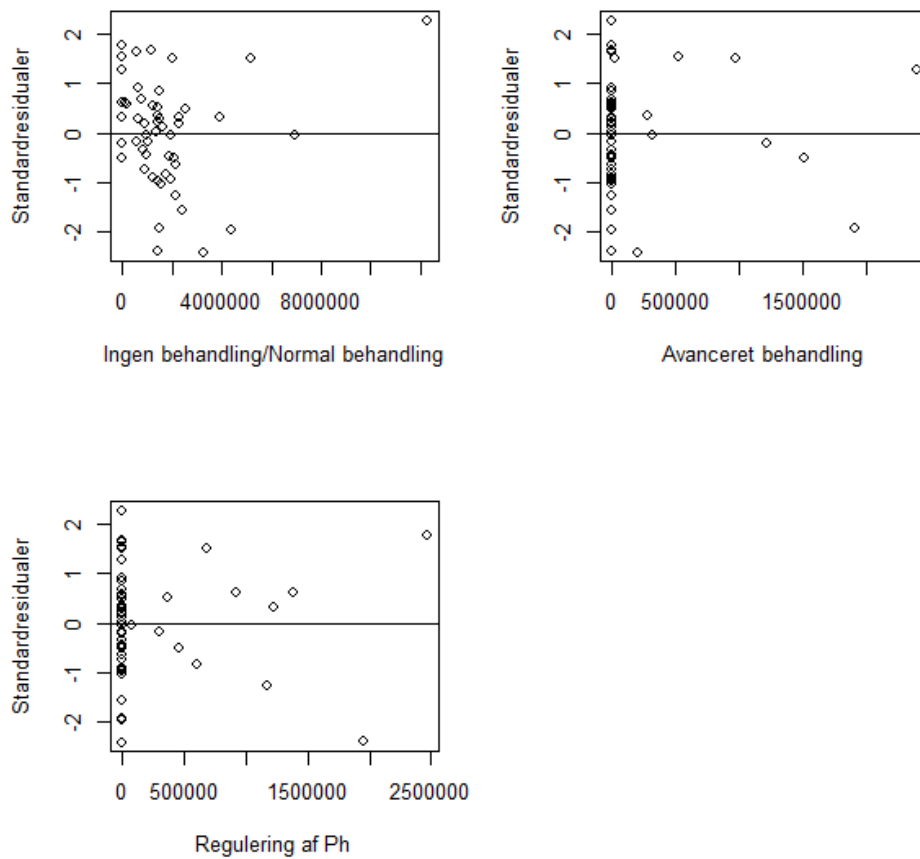
I Figur 3.4 ses det, at modellens fejlede er pænt normalfordelte og at de fleste observationer ligger indenfor konfidensintervallet. Det er forventeligt, at der er få observationer, som ligger udenfor konfidensintervallet og antagelsen om, at modellens fejlede er normalfordelte, er dermed opfyldt.

Figur 3.4 **QQ-plot for vandværker**



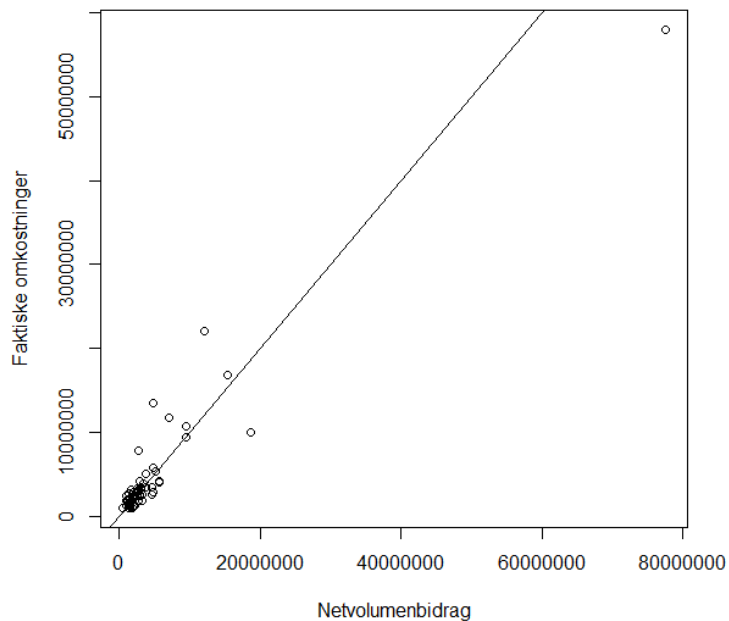
Dette kan ligeledes ses i Figur 3.5, hvor fejleddene er plottet mod de observerede værdier.

Figur 3.5 Fejleddsplot for vandværker



Slutteligt viser Figur 3.6, at modellens beregnede værdier stemmer godt overens med de observerede omkostninger. Der er få observationer, som har et stort fejld. Fire af observationerne er outliers og to er fjernet manuelt, og vi har vurderet, at der er tilstrækkelig få af disse observationer til, at vi kan fortsætte med den valgte model.

Figur 3.6 Faktiske omkostninger mod netvolumenbidraget for vandværker



3.4 Trykforøger- og trykreduktionsstationer

Modellens baggrund

Driftsomkostningerne ved at drive trykforøger- og trykreduktionsstationer forventes at være påvirket antallet af trykreduktionsstationer, antallet af trykforøgerstationer og samlet kapacitet på trykforøgerstationer. Endvidere forventes driftsomkostningerne at være påvirket af intervaller af den maksimale kapacitet på trykforøgerstationerne. Intervallerne er opdelt i 0-20 m³/t, 21-50 m³/t, 51-100 m³/t, 101-200 m³/t, 201-400 m³/t, 401-600 m³/t, 601-800 og over 800 m³/t.

Grundet høj korrelation mellem antal trykforøgerstationer i de enkelte intervaller og den tilsvarende maksimale kapacitet i intervallerne, har det været nødvendigt kun at se på et af de to underliggende forhold per interval.

Der er udover den valgte model blevet testet en række alternative modeller. I disse modeller er der blandt andet testet for andre sammensætninger af de underliggende forhold, og om der er behov for at transformere data. Den valgte model har vist de stærkeste statistiske egenskaber og finder samtidig koefficienter, som giver økonomisk intuitiv mening i forhold til hinanden.

Vi har vurderet, at den mest velegnede model til at beskrive de samlede driftsomkostninger ved at drive t trykforøger- og trykreduktionsstationer, er angivet i ligning (11).

Y angiver de årlige driftsomkostninger ved at drive trykforøgerstationer, X_1 angiver antallet af trykreduktionsstationer samt trykforøgerstationer i intervallet 0-100 m³/t, X_2 angiver den maksimale kapacitet i intervallet 101-400 m³/t, og X_3 angiver den maksimale kapacitet i intervallet over 400 m³/t. β -værdierne angiver enhedsomkostningerne for hvert af de underliggende forhold, og β_0 angiver modellens konstantled.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 \quad (11)$$

I Tabel 3 præsenteres resultaterne af modellen fra ligning (11).

Tabel 3 Regressionsanalysens resultater for trykforøger- og trykreduktionsstationer

Variabel	Koefficient	Spredning	t-værdi	p-værdi
Konstantled	72.371,35	32.361,82	2,236	0,03012*
Antal trykreduktionsstationer samt trykforøgerstationer 0-100 m ³ /t	15.076,40	3.303,27	4,564	0,0000361***
Maksimal kapacitet 101-400 m ³ /t	162,32	54,08	3,001	0,00429**
Maksimal kapacitet over 400 m ³ /t	65,66	26,83	2,447	0,01820*
Antal observationer: 51				
Justeret R ² = 0,4459				

Anm.: Stjerne i kolonnen med p-værdier angiver, hvor signifikante de enkelte variable er. De variable, som er mest signifikante, har tre stjerner, hvorimod variable som ikke er signifikante, har nul stjerner.

Omkostningsækvivalenten for trykforøger- og trykreduktionsstationer

Omkostningsækvivalenten ved at drive trykforøgerstationer kan ses i ligning (12).

$$Y = 72.371,35 + 15.076,40X_1 + 162,32X_2 + 65,66X_3 \quad (12)$$

Modellen angiver, at det koster 72.371,35 kr. at have trykforøger- og trykreduktionsstationer. Herudover angiver den, at det koster 15.076,40 kr. per trykreduktionsstation og trykforøgerstationer i intervallet 0-100 m³/t, 162,32 kr. per m³/t kapacitet i intervallet 101-400 m³/t samt 65,66 kr. per m³/t kapacitet i intervallet over 400 m³/t.

Kvalitetskontrol af omkostningsækvivalenten

Tabel 3 angiver, at alle variable er signifikante ved et 5 pct. signifikansniveau. Det betyder, at der er stor statistisk sikkerhed i modellens resultater. Det kan ses, at det er dyrere at drifte større pumper end mindre, hvilket opfylder vores forventninger.

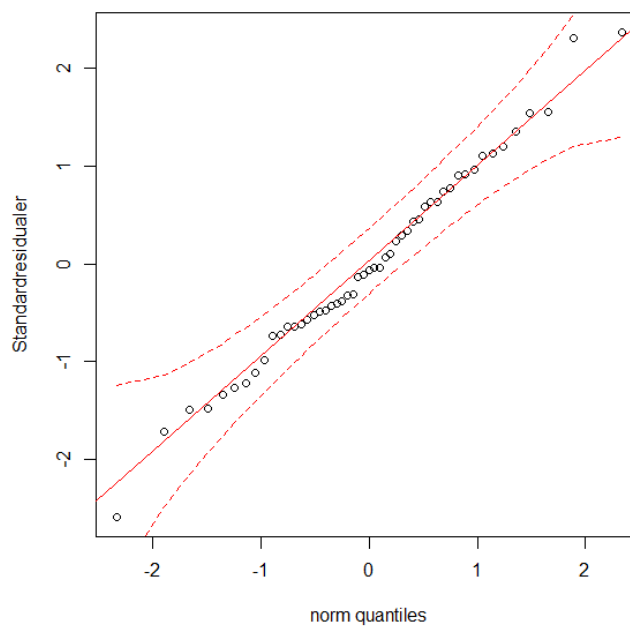
Hørsholm Vand ApS er ikke med til at bestemme omkostningsækvivalenten for denne costdriver.

Ved brug af Cook's distance er der fjernet fem selskaber, som ellers ville have haft for stor indflydelse på modellens resultater. Disse selskaber er:

- » Frederikshavn Vand A/S
- » TREFOR Vand A/S
- » Vand Ballerup A/S
- » Verdo Vand A/S
- » Aarhus Vand A/S

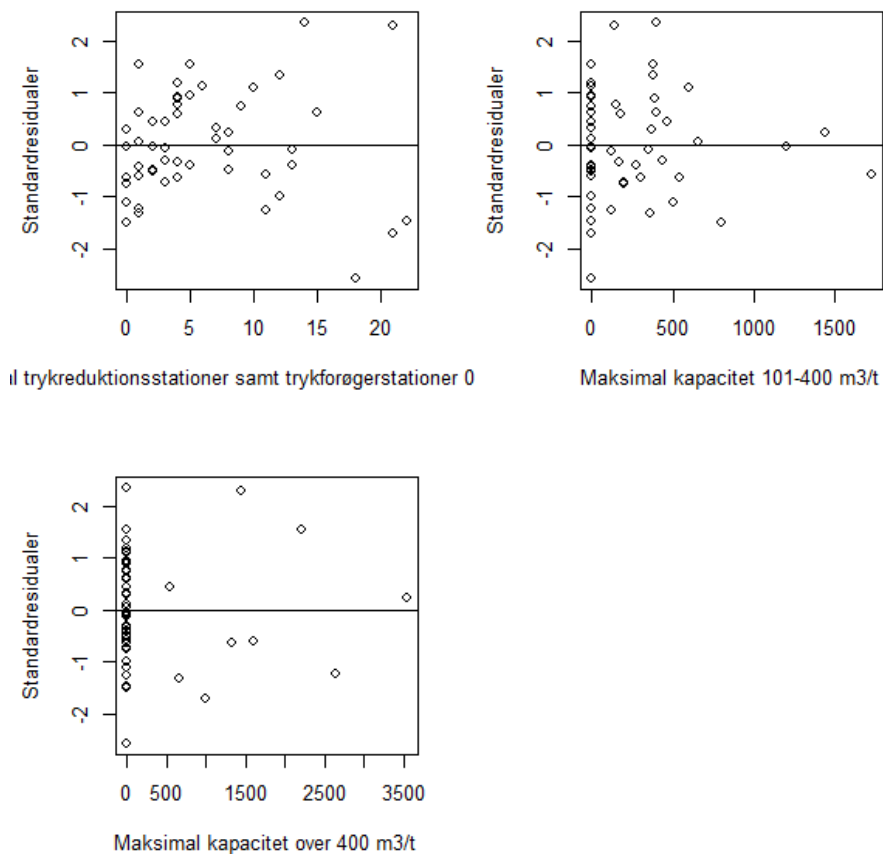
I Figur 3.7 ses det, at modellens fejlede er pænt normalfordelte og ligger indenfor konfidensintervallet.

Figur 3.7 **QQ-plot for trykforøger- og trykreduktionsstationer**



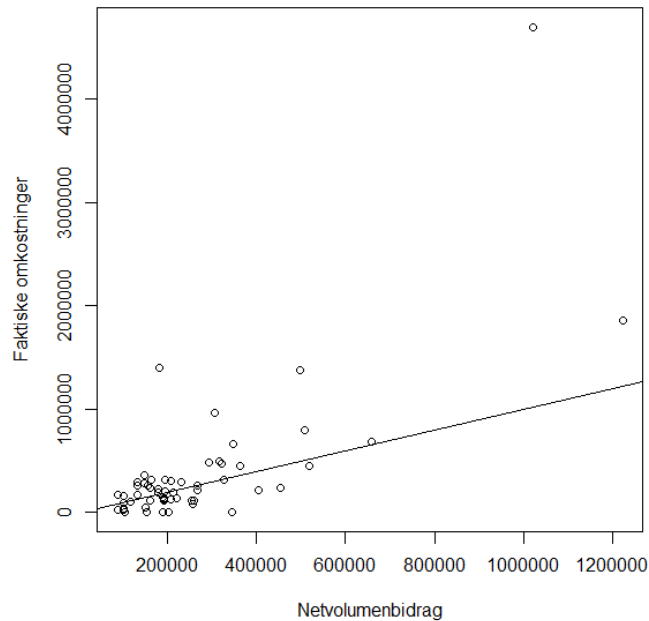
Herudover viser en Breusch-Pagan test, at der ikke er problemer med heteroskedasticitet. Dette kan ligeledes ses i Figur 3.8, hvor de standardiserede fejllid er plottet mod de observerede værdier.

Figur 3.8 Fejllidplot for trykforøger- og trykreduktionsstationer



Slutteligt viser Figur 3.9 et plot af de indberettede omkostninger mod de beregnede omkostninger. Den markerede linje er et udtryk for det punkt, hvor de beregnede og faktiske er ens. Som det ses, er der få selskaber med store afvigelser, og vi har derfor valgt at gå videre med modellen.

Figur 3.9 Faktiske omkostninger mod netvolumenbidraget for trykforøger- og trykreduktionsstationer



3.5 Rentvandsledninger og stik

Modellens baggrund

Driftsomkostninger ved at drive rentvandsledninger og stik forventes at være påvirket af længden på rentvandsledningerne og antallet af stik fordelt på zonerne land, by, sommerhus, city og indre city.

Længden på rentvandsledningerne og antallet af stik fordelt på de enkelte zoner er højt korrelerede. Det medfører, at vi ikke har kunnet finde en model, hvor begge de underliggende forhold er indeholdt i en zone. Herudover har vi valgt at slå zonerne land og sommerhus sammen samt city og indre city. Dette giver bedre statiske modeller og har vist sig nødvendigt for at få en solid model. Den valgte model er en lineær model uden konstantled og med parametrene antal stik i land- og sommerhuszonen, kilometer rentvandsledning i byzonen og kilometer rentvandsledning i city- og indre cityzonen.

Der er udover den valgte model blevet testet en række alternative modeller. I disse modeller er der blandt andet testet for andre sammensætninger af de underliggende forhold, og om der er behov for at transformere data. Den valgte model har vist de stærkeste statistiske egenskaber og finder samtidig koefficienter, som giver intuitiv mening i forhold til hinanden.

Den valgte model er beskrevet i ligning (13). Y angiver de årlige driftsomkostninger ved at drive rentvandsledninger og stik. X_1 angiver antallet af stik i land- og sommerhuszonen, X_2 angiver kilometer rentvandsledning i byzonen og X_3 angiver kilometer rentvandsledning i city- og indre cityzonen. β -værdierne angiver enhedsomkostningerne for de underliggende forhold.

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \quad (13)$$

I Tabel 4 præsenteres resultaterne af modellen fra ligning (13).

Tabel 4 Regressionsanalysens resultater for rentvandsledninger og stik

Variabel	Koefficient	Spredning	t-værdi	p-værdi
Antal stik i land og sommerhus-zonen	248,44	94,15	2,639	0,0108*
Kilometer rentvandsledning i byzonen	7.395,11	1.085,27	6,814	7,60E-09***
Kilometer rentvandsledning i city- og indre cityzonen	36.417,47	4.218,61	8,633	8,22E-12***
Antal observationer: 58				
Justeret R ² = 0,9332				

Anm.: Stjernerne i kolonnen med p-værdier angiver, hvor signifikante de enkelte variable er. De variable, som er mest signifikante, har tre stjerner, hvorimod variable som ikke er signifikante, har nul stjerner.

Omkostningsækvivalenten for rentvandsledninger og stik

Omkostningsækvivalenten ved at drive rentvandsledninger og stik kan ses i ligning (14).

$$Y = 248,44X_1 + 7.395,11X_2 + 36.417,47X_3 \quad (14)$$

Modellen angiver, at det koster 248,44 kr. at drive et stik i land- og sommerhuszonen, 7.395,11 kr. at drive en kilometer rentvandsledning i byzonen og 36.417,47 kr. at drive en kilometer rentvandsledning i city- og indre cityzonen.

Kvalitetskontrol af omkostningsækvivalenten

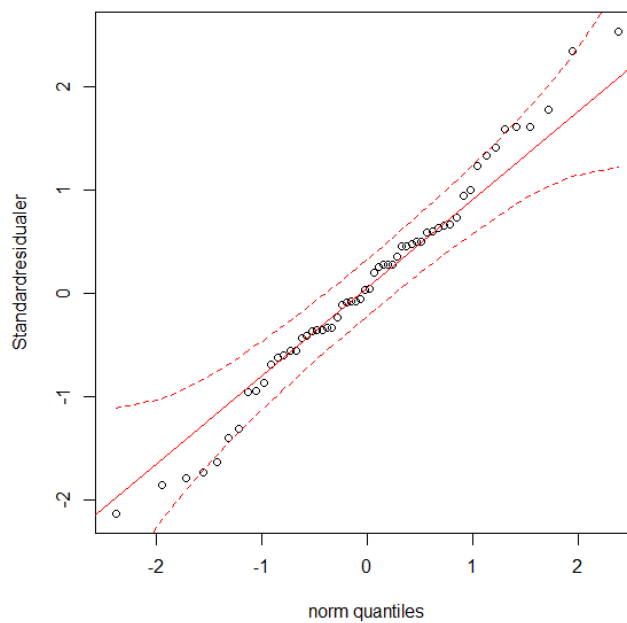
Tabel 4 angiver, at alle de underliggende forhold er signifikante ved et 5 pct. signifikansniveau. Det betyder, at der er stor statistisk sikkerhed for den valgte model.

Ved brug af Cook's Distance er der fjernet to selskaber, som ellers ville have haft for stor indflydelse på modellens resultater. Disse selskaber er:

- » HOFOR Vand København A/S
- » Ringkøbing-Skjern Vand A/S

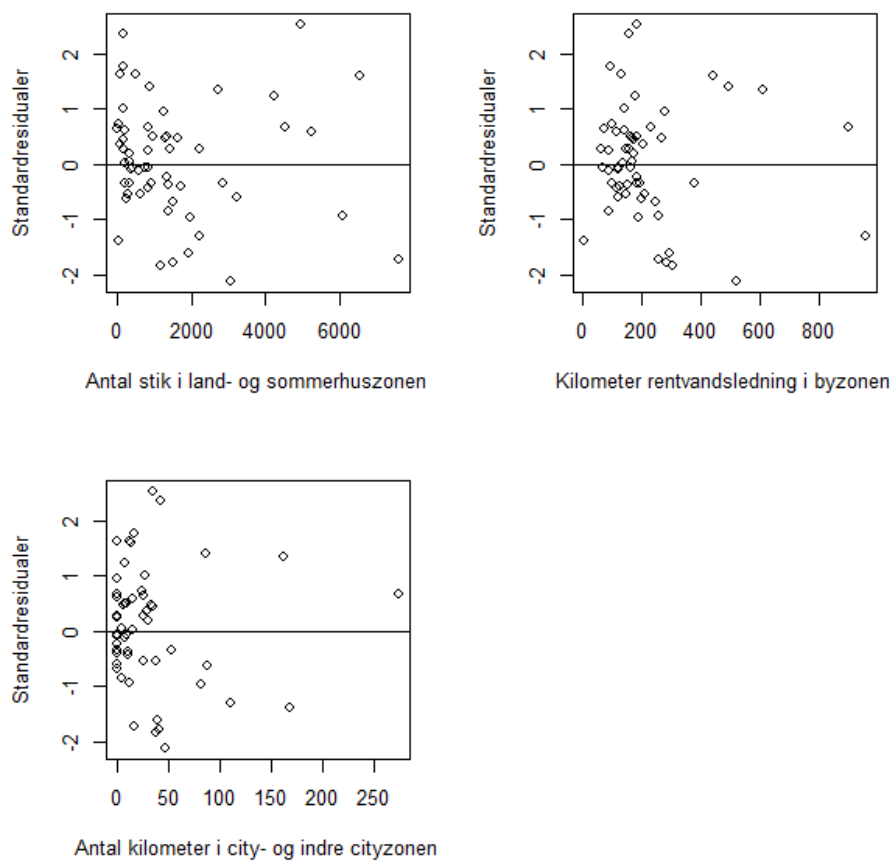
I Figur 3.10 ses det, at modellens fejlede er pænt normalfordelte og ligger indenfor konfidensintervallet.

Figur 3.10 QQ-plot for rentvandsledninger og stik



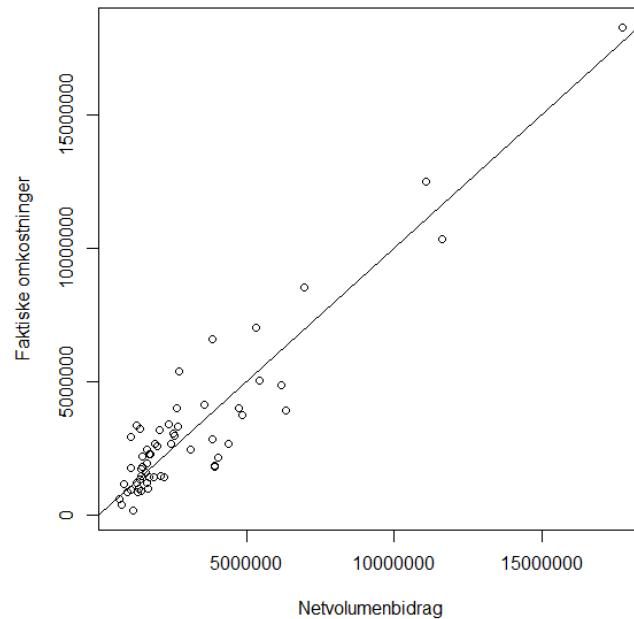
Herudover viser en Breusch-Pagan test, at der ikke er problemer med heteroskedasticitet. Dette kan ligeledes ses i Figur 3.11, hvor de standardiserede fejlede er plottet mod de observerede værdier.

Figur 3.11 Fejledeplot for rentvandsledninger og stik



Slutteligt viser Figur 3.12 et plot af de indberettede omkostninger mod de beregnede omkostninger. Den markerede linje er et udtryk for det punkt, hvor de beregnede og faktiske er ens. Som det ses, har de fleste observationer et meget lille fejld. Vi har derfor vurderet, at vi kan fortsætte med modellen.

Figur 3.12 **Faktiske omkostninger mod netvolumenbidraget for rentvandsledninger og stik**



3.6 Målere og kunder

Modellens baggrund

Driftsomkostningerne, som vedrører målere og kunder, forventes at være påvirket af antallet af kunder, som selskabet transporterer vand til. Antallet af målere er i indberetningen blevet inddelt i intervaller opdelt efter årsforbruget på den enkelte måler, således at der er antal målere i intervallerne 1-200 m³ og 201-10.000 m³ samt antal målere over 10.000 m³.

Vi har set det nødvendigt at pulje alle målerintervallerne sammen til ét underliggende forhold. Der er udført flere forskellige regressioner med forskellige transformationer af data samt inddeling af intervallerne. Det har dog ikke været muligt at finde en statistisk signifikant sammenhæng mellem de to store intervaller og de indberettede omkostninger, hvorfor det er valgt at slå intervallerne sammen til ét underliggende forhold.

Den bedste model beskriver dermed de samlede omkostninger til kundeforvaltning ud fra det samlede antal af målere. Det er fundet, at den mest velegnede model til at beskrive sammenhæng mellem omkostninger og målere er en udgave af en Cobb-Douglas-model, som beskrevet i ligning (15).

Y angiver de årlige driftsomkostninger til målere og kunder, og X_1 angiver det samlede antal målere i de tre intervaller.

$$Y = \beta_0 X_1^{\beta_1} \quad (15)$$

For at kunne estimere en model som udviser eksponentiel sammenhæng, er det nødvendigt at lave en logaritmisk transformation af modellen.

$$\log(Y) = \log(\beta_0) + \beta_1 \log(X_1) \quad (16)$$

De estimerede værdier for modellen er præsenteret i Tabel 5.

Tabel 5 Regressionsanalysens resultater for målere og kunder

Variabel	Koefficient	Spredning	t-værdi	p-værdi
Log (konstant)	5,5171	1,2564	4,391	4,84 e-05 ***
Log (målere)	0,9052	0,1356	6,676	1,03 e-08 ***

Antal observationer: 60
Justeret $R^2 = 0,4247$

Anm.: Stjernerne i kolonnen med p-værdier angiver, hvor signifikante de enkelte variable er. De variable, som er mest signifikante, har tre stjerner, hvorimod variable som ikke er signifikante, har nul stjerner.

Korrektionsfaktor

Når data transformeres, er det nødvendigt at lave en korrektion i forbindelse med tilbage-transformationen af den estimerede model.

Standardafvigelsen for fejleddene er $\sigma = 0,6622$. Dermed kan det udledes at korrektionsfaktoren er 1,2451, hvilket skal ganges på højre side af ligningen.

Omkostningsækvivalenten for kunder

Omkostningsækvivalenten for målere og kunder kan beskrives ud fra ligning (17)

$$Y = 1,2451 \cdot 248,9238 = 309,9350 X_1^{0,9052} \quad (17)$$

Konstantfaktoren 248,9238 i (17) er beregnet ved $\exp(5,5171)^4$, og derefter er korrektionsfaktoren ganget på. Når parameteren β_1 er under 1, betyder det, at der er stordriftsfordele forbundet med omkostninger til målere og kunder. Når antallet af målere øges med 1 pct. stiger omkostningerne med 0,9052 pct.

Kvalitetskontrol af omkostningsækvivalenten

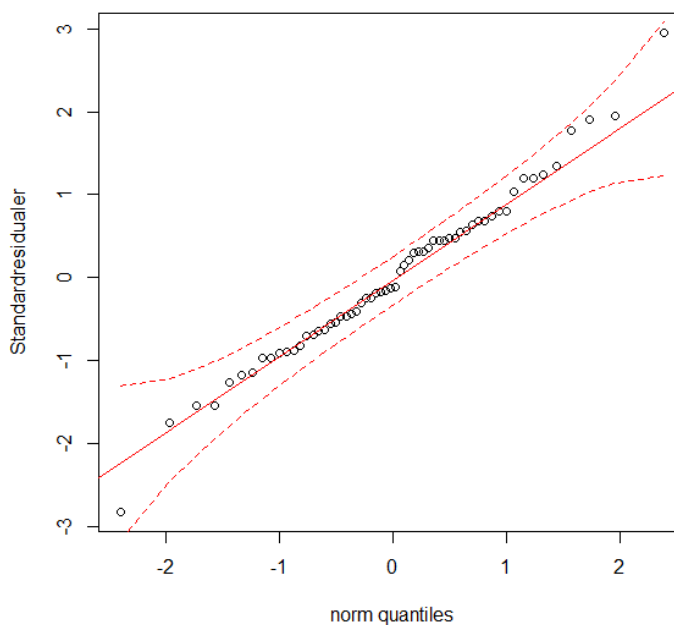
I Tabel 5 er det angivet, at koefficienten for målere er signifikant på et 5 pct. niveau.

Der er ikke fjernet nogen outliers på baggrund af Cook's distance. Det betyder, at alle selskaber har indflydelse på estimeringen af parametrene, men at ingen har alt for stor indflydelse.

⁴ Exp betegner eksponentialfunktionen.

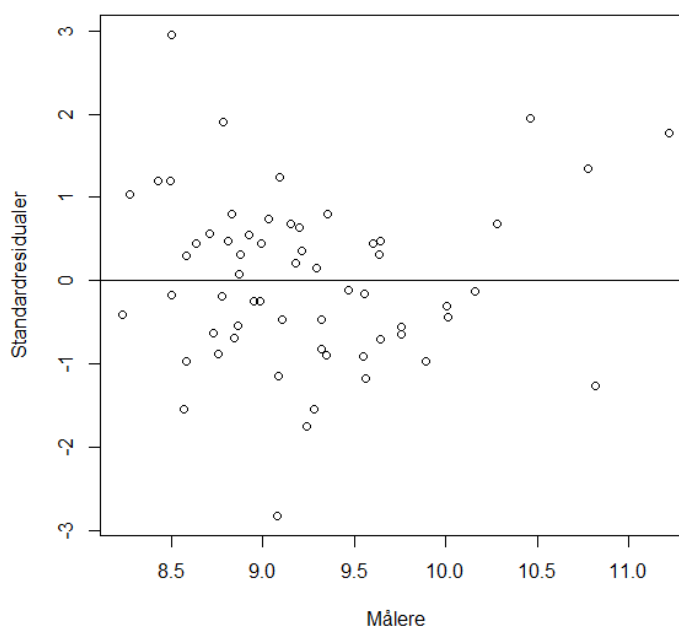
Figur 3.13 viser et QQ-plot af modellen. Det ses, at fejlleddene er pænt normalfordelte. Desuden ses det, at der ikke er tegn på, at data skal transformeres yderligere.

Figur 3.13 QQ-plot for målere og kunder



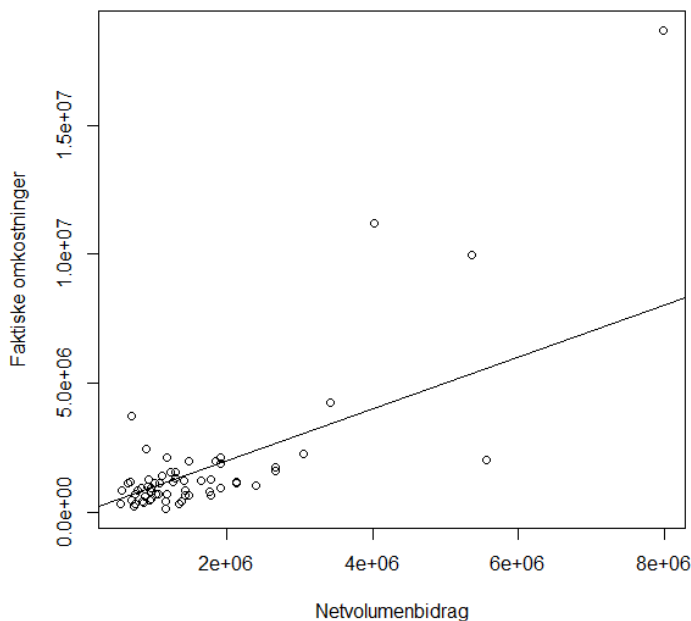
Af fejlledsplottet (Figur 3.14) ses, at fejlleddene ligger pænt fordelt omkring 0. Yderligere viser en Breusch-Pagan test, at der ikke er problemer med heteroskedasticitet.

Figur 3.14 Fejlledsplot for målere og kunder



Figur 3.15 viser et plot af de indberettede omkostninger mod de beregnede omkostninger. Den markerede linje er et udtryk for det punkt, hvor de beregnede og faktiske er ens. Som det ses, er der enkelte observationer med et meget stort fejld. Vi har vurderet, at det er tilstrækkeligt få observationer til, at vi kan fortsætte med modellen.

Figur 3.15 Faktiske omkostninger mod netvolumenbidraget for målere og kunder



3.7 Generel administration

Modellens baggrund

Driftsomkostninger til generel administration forventes at være påvirket af selskabets debiterede vandmængde. I forbindelse med genindhentningen af den debiterede vandmængde, har det været muligt at lave en underopdeling af den totale mængde. Den totale debiterede vandmængde er således indberettet i følgende delmængder:

- » Salg af ubehandlet råvand til andre vandforsyninger eller virksomheder
- » Rentvandsmængde leveret til andre vandforsyninger
- » Det samlede målte vandforbrug hos alle de tilsluttede forbrugere.

Der er foretaget forskellige beregninger og transformationer af data for at undersøge, hvordan omkostningerne til generel administration forklares af den debiterede vandmængde. Der er ikke blevet fundet nogen statistisk signifikant sammenhæng mellem omkostningerne og "Salg af ubehandlet råvand til andre vandforsyninger eller virksomheder" samt "Rentvandsmængde leveret til andre vandforsyninger". I den endelige model er mængderne derfor summeret til en total debiteret vandmængde.

Udover de omkostninger selskaberne har indberettet til generel administration, indgår også de omkostninger som er indberettet til revisorerklæringer samt til brancheforeninger.

Der er foretaget forskellige transformationer af data for at undersøge, hvilken model som bedst beskriver sammenhængen mellem omkostninger til generel administration og den debi-

terede vandmængde. Det er fundet, at den bedste model er en Cobb-Douglas funktion som beskrevet i (18).

Y angiver de årlige driftsomkostninger til generel administration. X_1 angiver selskabets totale debiterede vandmængde i m^3 .

$$Y = \beta_0 \cdot X_1^{\beta_1} \quad (18)$$

De estimerede værdier for modellerne er præsenteret i Tabel 6.

Yderligere fremgår også standardafvigelsen for fejledene (σ) i tabellerne samt den deraf udledte korrektionsfaktor. Korrektionsfaktoren til Cobb-Douglas skal ganges på højre side af ligning (18).

Tabel 6 Resultater for regressionsanalysen af generel administration

Variabel	Koefficient	Spredning	t-værdi	p-værdi
Log (konstant)	1,4511	1,4123	1,027	0,308
Log(debiteret vandmængde)	0,8980	0,0967	9,287	<0,0001***

Antal observationer: 60
Justeret $R^2 = 0,5910$

Anm.: Stjerne i kolonnen med p-værdier angiver, hvor signifikante de enkelte variable er. De variable, som er mest signifikante, har tre stjerner, hvorimod variable som ikke er signifikante, har nul stjerner.

Korrektionsfaktor

Når data transformeres, er det nødvendigt at lave en korrektion i forbindelse med tilbage-transformationen af den estimerede model.

Standardafvigelsen for fejleddene er $\sigma = 0,5662$. Dermed kan det udledes at korrektionsfaktoren er 1,1738, hvilket skal ganges på højre side af ligningen.

Omkostningsækvivalenten for generel administration

Omkostningsækvivalenten for generel administration kan beskrives ved ligning (19):

$$Y = 1,1738 \cdot 4,2676 \cdot X_1^{0,8980} \quad (19)$$

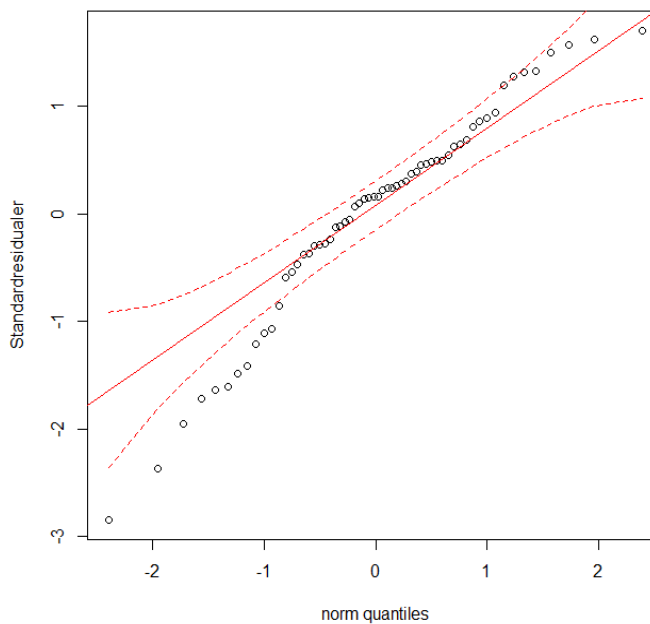
Konstanten 4,2676 er beregnet ved $\exp(1,4511)$. Når parameteren β_1 er mindre end 1, betyder det, at der er stordriftsfordele forbundet driften omkring generel administration. Når den debiterede vandmængde øges med 1 pct. stiger omkostningerne til generel administration med 0,8980 pct.

Kvalitetskontrol af omkostningsækvivalenten

Af Tabel 6 fremgår det, at koefficienten for debiteret vandmængde er statistisk signifikant på et 5 pct.-niveau. Der er foretaget outlieranalyse under estimationen, men ingen selskaber er fjernet på baggrund af dette.

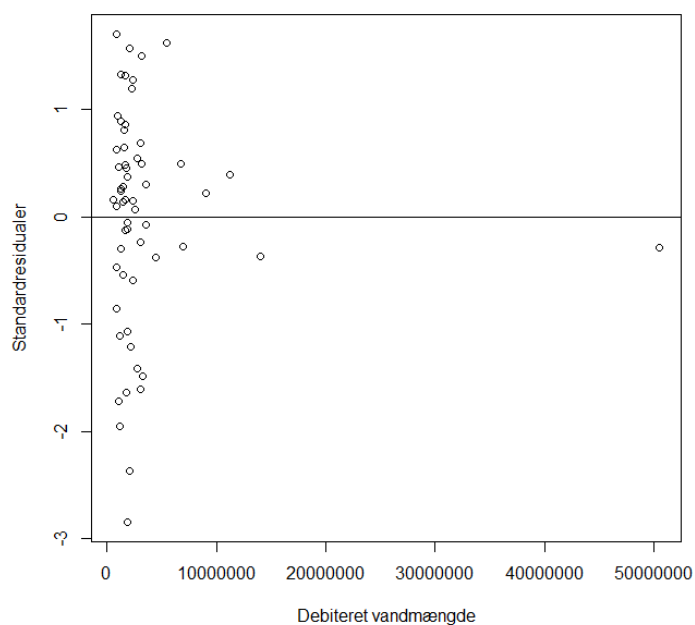
Som nævnt tidligere er det undersøgt, om transformationer af data kan skabe en bedre model. QQ-plottet for regressionen fremgår af Figur 3.16, og det ses, at nogle observationer ligger udenfor konfidensintervallet. Det er valgt ikke at ændre på modellen, da andre transformationer ikke giver bedre resultater samt fordi modellen giver økonomisk intuitiv mening og indeholder samtlige selskaber.

Figur 3.16 QQ-plot for generel administration



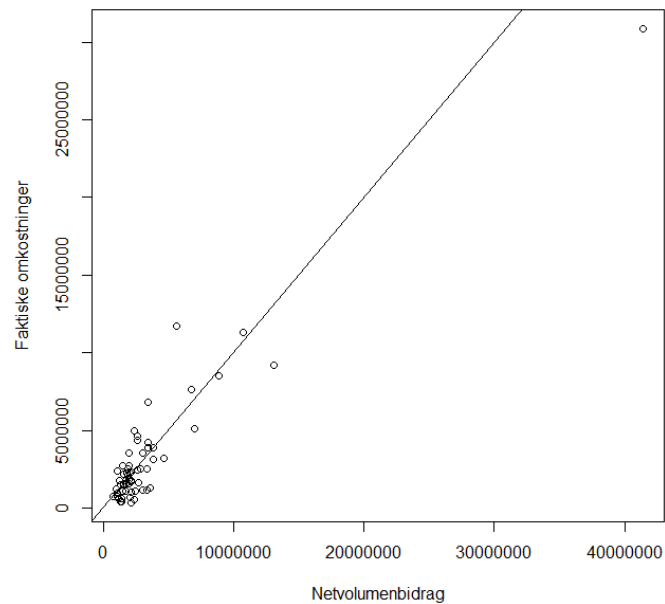
Figur 3.17 viser plot af fejleddene fra modellen mod de debiterede vandmængder. Det ses, at fejleddene ligger pænt fordelt omkring 0. Der ser ud til at være enkelte store selskaber med meget små fejllid samt enkle selskaber med større fejllid, men dog inden for hvad der er acceptabelt. En Breusch-Pagan test viser, at der ikke er problemer med heteroskedasticitet.

Figur 3.17 Fejllidplot for generel administration



Til sidst viser Figur 3.18, plottet af de beregnede værdier mod de faktisk indberettede omkostninger. Den markerede linje er et udtryk for det punkt, hvor de beregnede værdier er lig de indberettede. Generelt ligger selskaberne pænt fordelt omkring linjen.

Figur 3.18 **Faktiske omkostninger mod netvolumenbidraget for general administration**



Kapitel 4

Samlet netvolumenmål

4.1 Indledning

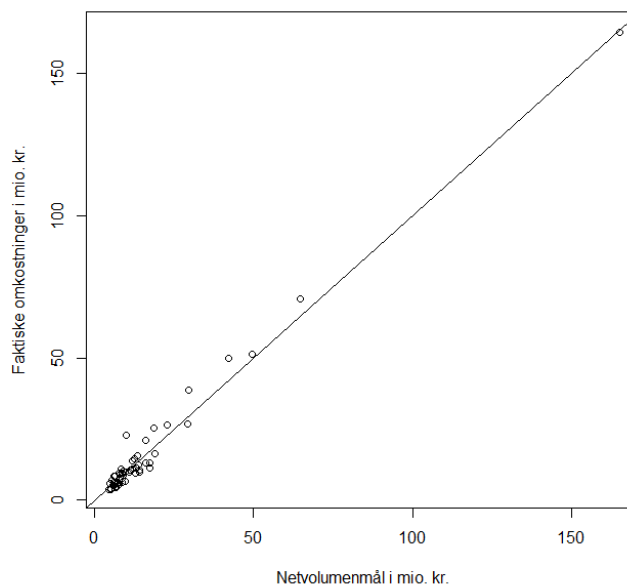
I dette kapitel beskrives sammenhængen mellem de samlede resultater for de fundne omkostningsækvivalenter og selskabernes faktiske driftsomkostninger.

4.2 Kvalitetskontrol

Vi har i de ovenstående kapitler gennemgået de enkelte costdrivere, hvor vi blandt andet har vist, hvor godt modellerne forudsiger selskabernes faktiske omkostninger. Vi vil i dette kapitel vise, hvor godt det samlede netvolumenmål er til at forudsige selskabernes samlede faktiske driftsomkostninger. Alle selskaber, som har indberettet fyldestgørende, indgår i de følgende resultater.

I Figur 4.1 ses et plot over selskabernes samlede omkostninger holdt op i mod deres samlede netvolumenmål.

Figur 4.1 Samlede **faktiske driftsomkostninger mod samlet netvolumenmål**



Figur 4.2 viser samme plot som Figur 4.1, blot uden HOFOR Vand København, for bedre at kunne se fordelingen af de mindre selskaber.

De to figurer viser, at selskaberne følger linjen, som angiver en præcis forudsigende model, meget fint. Der er få selskaber, som afviger fra linjen, hvilket er forventeligt i en statistisk model.

Figur 4.2 **Samlede faktiske driftsomkostninger mod samlet netvolumenmål Uden HOFOR Vand København**

