

OPEX-netvolu- menmål

Teknisk beregning af omkostningsækvivalenterne til benchmarking af spildevands-selskaber

Version 2 - Juni 2019



OPEX-netvolumenmål

Konkurrence- og Forbrugerstyrelsen

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

Online ISBN 978-87-7029-669-4

OPEX-netvolumenmål er udarbejdet af
Konkurrence- og Forbrugerstyrelsen.

Juni 2019

Indhold

Kapitel 1	4
Indledning	4
1.1 Revidering af costdriver for renseanlæg.....	4
1.2 Opbygning af papiret.....	5
Kapitel 2	6
Metodebeskrivelse	6
2.1 Indledning.....	6
2.2 Mindste kvadraters metode	6
Kapitel 3	13
Bestemmelse af omkostningsækvivalenter	13
3.1 Indledning.....	13
3.2 Ledninger	13
3.3 Pumpestationer.....	17
3.4 Regnvandsbassiner.....	21
3.5 Spildevandsbassiner.....	25
3.6 Renseanlæg.....	28
3.7 Slambehandling.....	36
3.8 Slamdisponering.....	40
3.9 Målere og kunder.....	44
3.10 Generel administration.....	48
Kapitel 4	60
Samlet netvolumenmål	60
4.1 Indledning.....	60

Kapitel 1

Indledning

Dette papir er en teknisk beskrivelse af de statistiske metoder, der ligger til grund for udregningen af omkostningsækvivalenterne til brug for benchmarkingen af spildevandsselskaber med en debiteret vandmængde over 800.000 m³. og resultaterne for hver enkelt costdriver. Omkostningsækvivalenterne udtrykker den gennemsnitlige sammenhæng mellem selskabernes driftsomkostninger og de relevante costdrivere. Omkostningsækvivalenterne bruges til at beregne selskabernes OPEX-netvolumenmål, der indgår i benchmarkingen. Omkostningsækvivalenterne har været anvendt i benchmarkingen siden revisionen i 2017 og har indgået i fastsættelsen af selskabernes økonomiske rammer for 2018 og frem.

Omkostningsækvivalenterne er baseret på data fra selskaberne. Det indberettede data indeholder oplysninger om 85 spildevandsselskabers costdrivere og driftsomkostninger fra 2015. Data er indberettet på baggrund af vores konterings- og costdrivervejledning fra 2014.

1.1 Revidering af costdriver for renseanlæg

Siden revisionen i 2017 er vi blevet opmærksomme på uhensigtsmæssigheder ved den tidligere omkostningsækvivalent for renseanlæg. Det knytter sig til:

- » Datagrundlaget fra OPEX-revisionen, der er angivet på selskabsniveau og ikke for de enkelte renseanlæg
- » Forholdet mellem belastningstyperne BOD, COD og kvælstof

Den tidligere omkostningsækvivalent blev beregnet ud fra et datagrundlag, hvor det ikke var muligt at henhøre et selskabs samlede omkostninger ved at drive renseaktiviteter til hvert af de pågældende renseanlæg. Det resulterede i et datagrundlag med relativt få observationer, hvorfor vi eksempelvis fandt det nødvendigt at pulje flere af rensetyperne samt land- og byzonerne.

Vi har de seneste år oplevet et stort antal ansøgninger om særlige forhold, der vedrører forholdet mellem COD og BOD. I den nuværende model er et selskabs belastningsgrad alene givet ved BOD. De mange ansøgninger om særlige forhold vidner om, at et selskabs belastningsgrad ikke alene er givet ved BOD, men også COD og kvælstof, og at forholdet mellem de tre belastningstyper varierer på tværs af renseanlæg. Af den årsag ønsker vi at inkorporere COD og kvælstof i beregningen af omkostningsækvivalenten for renseanlæg. Dels for at lette selskabernes arbejde med ansøgninger om særlige forhold og dels for at opnå en så retvisende model som muligt.

I december 2018 bad vi derfor de selskaber, der havde indberettet til revisionen i 2017, om ekstraordinært at indberette data til genberegning af omkostningsækvivalenten for renseanlæg. Formålet er, at modellen bruger et mål for belastningsgrad, der er baseret på BOD, COD og kvælstof, samt en kontering af driftsomkostninger for hvert renseanlæg, således at modellen angiver en mere retvisende sammenhæng mellem et renseanlægs omkostninger, rensetype og zoneplacering til brug i benchmarkingen.

I dette papir er det således kun afsnit 3.6 om renseanlæg, der er nyt i forhold til den tidligere version af dette papir fra august 2017.

1.2 Opbygning af papiret

Kapitel 2 beskriver metoderne bag beregning af omkostningsækvivalenterne. Kapitlet beskriver de væsentligste teoretiske elementer for regressionsanalyserne samt vores anvendte kriterier for outliers. Kapitel 3 gennemgår de enkelte costdrivere, herunder hvordan de tilsvarende omkostningsækvivalenter er fundet. For hver costdriver bliver modellens baggrund først beskrevet, hvorefter den valgte model bliver gennemgået. Kapitel 4 undersøger afslutningsvist, hvor godt den samlede model beskriver de enkelte selskabers faktiske driftsomkostninger.

Kapitel 2

Metodebeskrivelse

2.1 Indledning

I dette kapitel vil vi gennemgå de generelle statistiske overvejelser, som 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. Når der bruges mindste kvadraters metode antages det, at der er en lineær sammenhæng mellem omkostninger og de underliggende forhold. Dette kan præsenteres i modellen nedenfor:

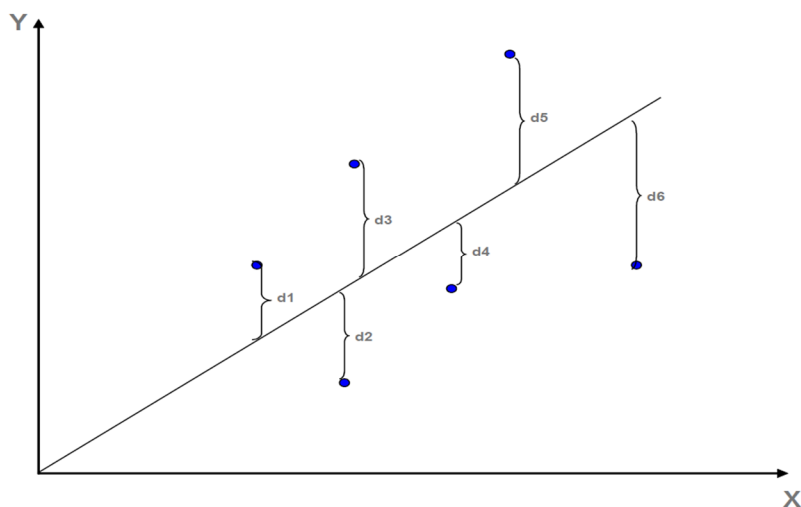
$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n + u \quad (1)$$

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 forstå, at regressionen i (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 sammenhæng 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 uden for 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.

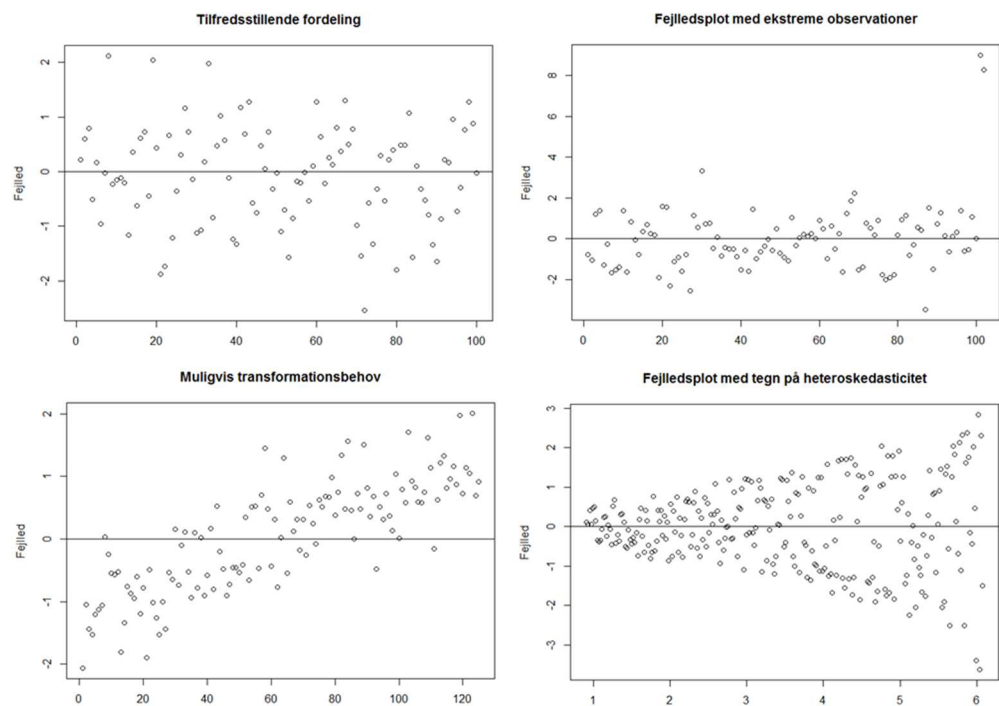
Signifikansniveauet sammenholdes med den økonomiske teori og logik omkring de enkelte costdrivere. Det betyder, at underliggende forhold, som viser sig ikke at være statistisk signifikante på et 5 pct.-niveau, kan overvejes at inddrages i modellen, hvis teorien kræver, at forholdet skal bruges til at 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 fejleddene¹ er normalfordelte med en middelværdi på nul.

I Figur 2.2 er vist forskellige plot af fejled mod en uafhængig variable. 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



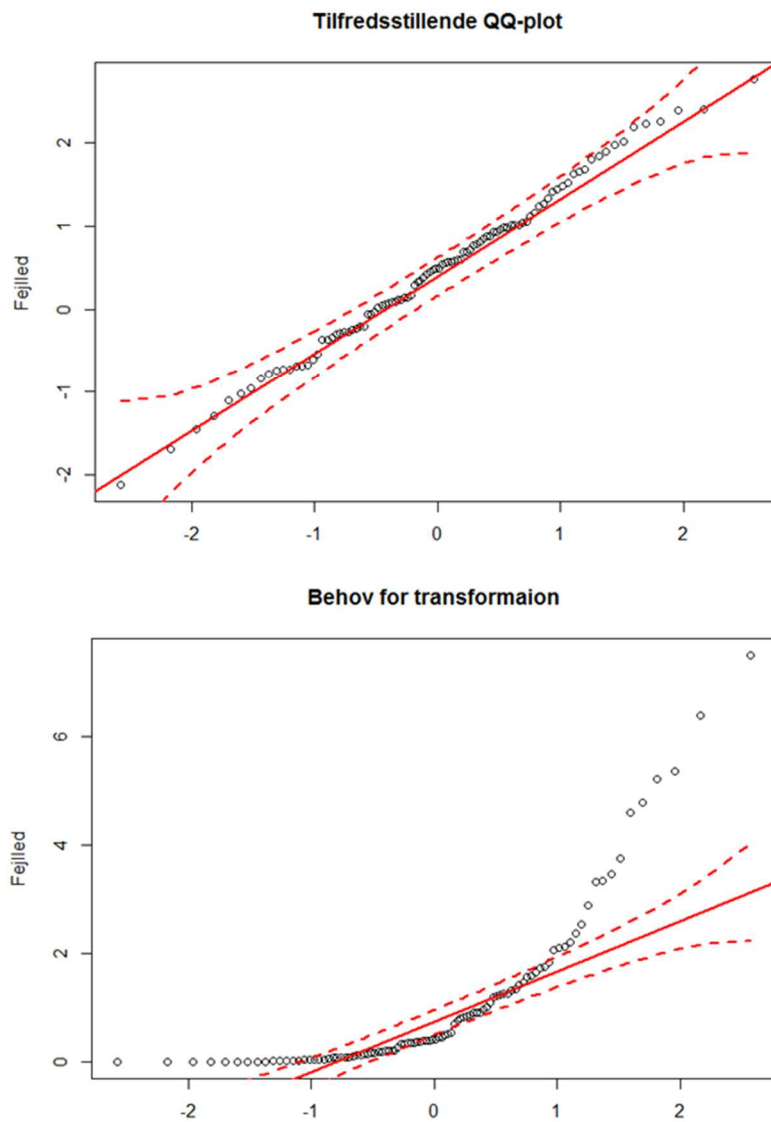
Ud fra fejleddplotsene er det muligt at se, om der er problemer med normalfordelingen af fejleddene. Der er problemer, hvis fejleddene ikke er tilfældigt fordelt omkring 0. Det er vigtigt, at antagelsen er opfyldt for at sikre, at den fundne $\hat{\beta}$ er middelfret. Hvis mere end 5 pct. af fejleddene ligger uden for 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

¹ Bemærk, at fejled også kaldes for standardresidualer

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 middelfret 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.

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 også 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 uhensigtsmæssig 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 variable 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 variablene 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 variable (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. fejlleddplotsene 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 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 variablene 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 fejlleddene, hvorfor fejlleddene 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 fejlleddet. Korrektionsfaktoren ganges på den fundne tilbagetransformerede model.

Hvis der er foretaget en transformation af data, således at det er opløftet i en potens, lægges følgende korrektionsfaktor (KF) til højresiden:

$$KF = \sigma^2 \quad (6)$$

Her betegner σ igen spredningen af fejlleddet. I dette tilfælde lægges korrektionsfaktoren til den aktuelle 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. Intuitiv og teoretisk baggrund
2. Statistiske kriterier
3. Antallet af selskaber

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

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 cost-driver. 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.

3) Til sidst er der også lagt vægt på antallet af selskaber i den endelige model. Som tidligere beskrevet er outliers blevet fjernet, hvis de har haft en uhensigtsmæssig 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 indledningsvis 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. 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 medtaget som en outlier ved den givne costdriver.

Kapitel 3

Bestemmelse af omkostningsækvivalenter

3.1 Indledning

Vi har beregnet omkostningsækvivalenter for de 9 costdrivere for spildevandsselskaberne:

- » Ledninger
- » Pumpestationer
- » Regnvandsbassiner
- » Spildevandsbassiner
- » Renseanlæg
- » Slambehandling
- » Slamdisponering
- » Målere og kunder
- » Generel administration

Beregningen af de enkelte omkostningsækvivalenter bliver beskrevet i de følgende afsnit. Først beskrives modellens 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 Ledninger

Modellens baggrund

Driftsomkostninger ved at drive ledninger forventes at være påvirket af ledningernes længde, type samt i hvilken zone de er beliggende. Ledninger kan være af typen regnvandsledning, spildevandsledning, fællesvandsledning, trykledning eller rørbassin og kan fordeles på zonerne sommerhus, land, by, city og indre city.

Vi har set det nødvendigt at pulje flere underliggende forhold for at kunne danne signifikante modeller. Ledningstyperne spildevandsledninger, fællesvandsledninger og trykledninger er samlet i én pulje, da vi efter dialog med branchen har vurderet, at omkostningerne til disse tre typer er nogenlunde ens. Yderligere er der observeret en høj korrelation mellem ledningstyperne, som medfører, at vi ikke mister vigtig information ved at pulje dem. Vi har ligeledes vurderet, at regnvandsledninger skal være i samme pulje som de resterende ledningstyper eksklusiv rørbassiner, da de er højt korrelerede med disse. Det er således kun rørbassiner, som indgår i modellen for sig selv.

Vi har herudover vurderet, at cityzonen og indre city-zonen skal danne en fælles zone, da der er for få selskaber med ledninger i indre city-zonen til, at vi kan udarbejde en regression herpå. Ligeledes er sommerhuszonen og landzonen samlet i én zone, da det ikke har været muligt at danne en signifikant model, hvor sommerhuszonen indgår separat. Slutteligt har vi vurderet, at rørbassiner ikke skal deles op i zoner, da der i de enkelte zoner er for få selskaber, som har rørbassiner til at danne en signifikant model.

Der er ud over 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.

Vi har vurderet, at den mest velegnede model til at beskrive de samlede driftsomkostninger ved at drive ledninger er angivet i ligning (7). Ledningstyperne regnvandsledninger, spildevandsledninger, fællesvandsledninger og trykledninger er slået sammen til en samlet mængde.

Y angiver de årlige driftsomkostninger ved at drive ledninger, X_1 angiver antal kilometer ledninger eksklusiv rørbassiner i sommerhuszonen samt landzonen, X_2 angiver antal kilometer ledninger eksklusiv rørbassiner i byzonen, X_3 angiver antal kilometer ledninger eksklusiv rørbassiner i cityzonen samt indre city-zonen og X_4 angiver antal kilometer rørbassin i alle zoner. β -værdierne angiver enhedsomkostningerne for hvert af de underliggende forhold.

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 \quad (7)$$

I Tabel1 præsenteres resultaterne af modellen fra ligning (7).

Tabel1: **Regressionsanalysens resultater for ledninger²**

Variabel	Koefficient	Spredning	t-værdi	p-værdi
Sommerhus + Land eksklusiv rørbassin	2327,9856	1.052,1	2,213	0,0303*
By eksklusiv rørbassin	3.666,4429	755,2	4,855	0,00000755***
City + Indre city eksklusiv rørbassin	15.645,4437	2.438,2	6,417	1,65E-08***
Rørbassin	279.770,2922	125.082,4	2,237	0,0286*

Antal observationer: 71
Justeret $R^2 = 0,8906$

Omkostningsækvivalenten for ledninger

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

$$Y = 2.327,9865 X_1 + 3.666,4290 X_2 + 15.645,4437 X_3 + 279.770,2922 X_4 \quad (8)$$

Modellen angiver, at det koster 2.327,9865 kr. at drive en kilometer ledning eksklusiv rørbassiner i sommerhus- og landzonen, 3.666,4290 kr. i byzonen og 15.645,4437 kr. i city- og indre city-zonen. Herudover koster det 279.770,2922 kr. at drive en kilometer rørbassin i alle zonerne.

Kvalitetskontrol af omkostningsækvivalenten

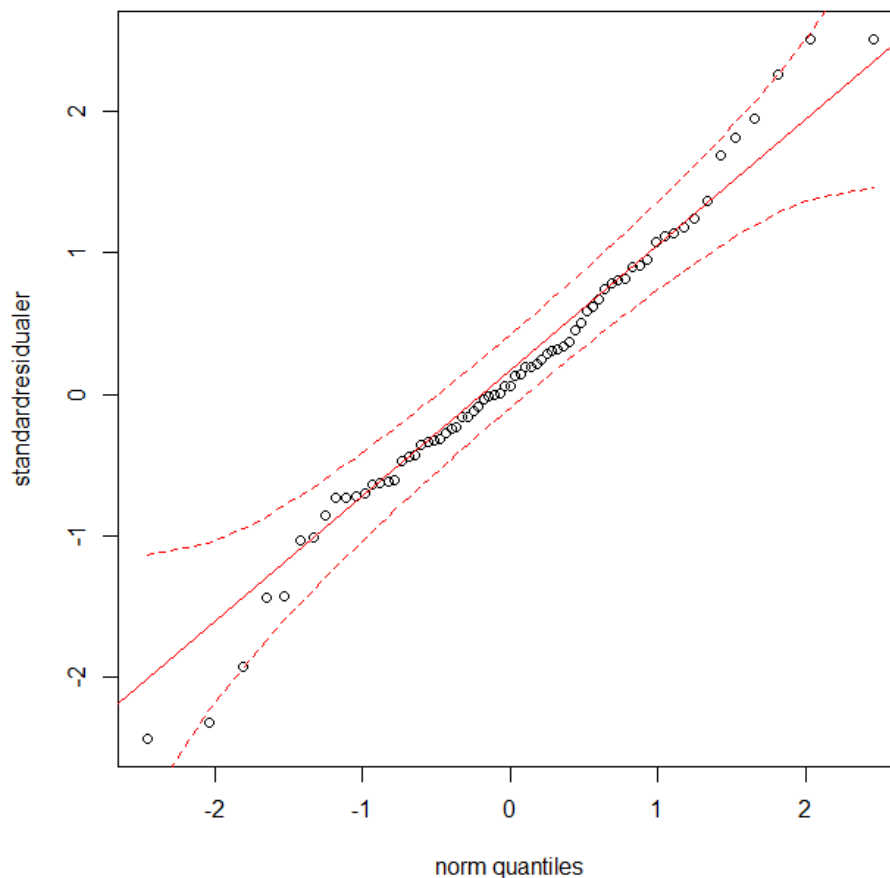
Tabel1 angiver, at alle variable er signifikante ved et 5 pct. signifikansniveau. Det betyder, at der er stor statistisk sikkerhed i modellens resultater. Herudover ses det, at det er dyrere at drive ledninger i tætbefolkede områder, og at alle β -værdier er positive.

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

² Stjernerne i kolonnen med p-værdier angiver, hvor signifikante de enkelte variable er. De variable som er mest signifikante har 3 stjerner, hvorimod variable som ikke er signifikante har 0 stjerner.

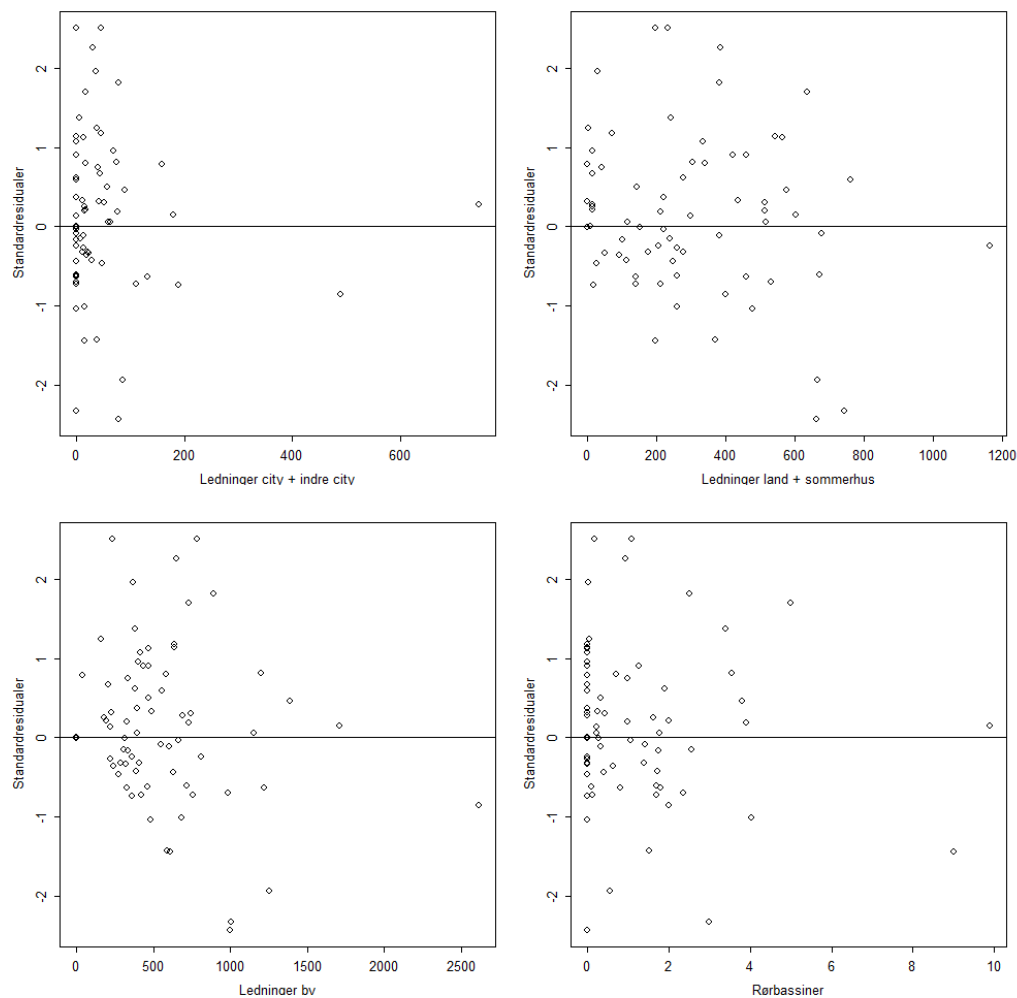
I Figur 3.1 ses det, at modellens fejlede er pænt normalfordelte, og ligger indenfor konfidensintervallet.

Figur 3.1 **QQ-plot for ledninger**

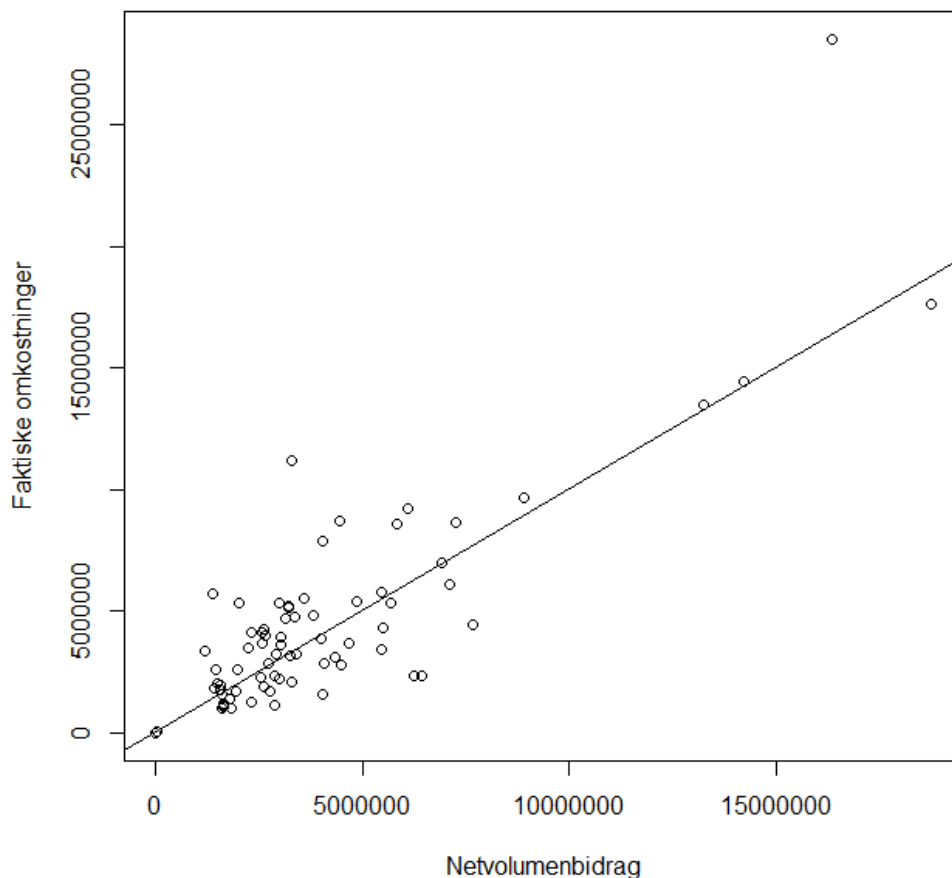


Herudover viser en Breusch-Pagan test, at der ikke er problemer med heteroskedasticitet. Dette kan ligeledes ses i Figur 3.2, hvor de standardiserede fejlede er plottet mod de observerede værdier.

Figur 3.2 Residualplot for hver variabel



Slutteligt viser Figur 3.3, at modellens beregnede værdier stemmer godt overens med de observerede omkostninger. Der er få observationer, som har et stort fejld. To af observationerne er outliers og vi har vurderet, at der er tilstrækkelig få af disse observationer til, at vi kan fortsætte med den valgte model.

Figur 3.3 **Observerede omkostninger mod beregnede**

3.3 Pumpestationer

Modellens baggrund

Driftsomkostninger ved at drive pumpestationer forventes at være påvirket af pumpestationernes maksimale kapacitet, pumpestationernes type og antallet af pumpestationer. Pumpestationerne kan være af typen fællespumpestationer, spildevandspumpestationer og regnvandspumpestationer. Herudover er pumpestationerne inddelt i intervaller baseret på deres maksimale kapacitet. Intervallerne er givet ved husstandspumper, 0-10 l/s, 11-50 l/s, 51-100 l/s, 101-200 l/s, 201-400 l/s, 401-600 l/s, 601-800 l/s, 801-1000 l/s og >1.000 l/s. I hvert interval indgår antallet af pumpestationer samt summen af deres maksimale kapacitet.

Vi har set det nødvendigt at pulje flere underliggende forhold for at kunne danne signifikante modeller. Pumpestationstyperne er samlet i én pulje, da vi efter dialog med branchen har vurderet, at omkostningerne til disse tre typer er nogenlunde ens. Yderligere er der observeret en høj korrelation mellem typerne, som medfører, at vi ikke mister vigtig information ved at pulje dem.

Grundet høj korrelation mellem antal pumpestationer og samlet kapacitet i de enkelte intervaller har det været nødvendigt kun at se på et af de to underliggende forhold i hvert interval.

Efter dialog med branchen har vi vurderet, at den bedste model findes ved at se på antallet af pumpestationer i de små intervaller og kapaciteten af pumpestationerne i de store intervaller.

Slutteligt har vi vurderet, at den bedste model findes, når vi puljer en del af intervallerne. Vi anvender dermed fem intervaller; husstandspumper, 0-10 l/s, 11-100 l/s, 101-600 l/s, >601 l/s.

Der er ud over 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.

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

Y angiver de årlige driftsomkostninger ved at drive pumpestationer, X_1 angiver antal husstandspumper, X_2 angiver antal pumpestationer i intervallet 0-10 l/s, X_3 angiver antal pumpestationer i intervallet 11-100 l/s, X_4 angiver den samlede maksimumskapacitet af pumpestationerne i intervallet 101-600 l/s og X_5 angiver den samlede maksimumskapacitet af pumpestationerne i intervallet >601 l/s. β -værdierne angiver enhedsomkostningerne for hvert af de underliggende forhold.

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 \quad (9)$$

I Tabel2 præsenteres resultaterne af modellen fra ligning (9).

Tabel2: Regressionsanalysens resultater for pumpestationer

Variabel	Koefficient	Spredning	t-værdi	p-værdi
Antal husstandspumper	2.855,5628	822,56	3,472	0,00094***
Antal 0-10 l/s	11.103,2371	2.059,28	5,392	1,11e-06***
Antal 11-100 l/s	23.253,3823	3.757,83	6,188	5,07e-08***
Kapacitet 101-600 l/s	315,9553	99,92	3,162	0,00241**
Kapacitet >601 l/s	214,5233	111,69	1,921	0,05929

Antal observationer: 68
Justeret R2= 0,8696

Omkostningsækvivalenten for pumpestationer

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

$$Y = 2.855,5628X_1 + 11.103,2371X_2 + 23.253,3823X_3 + 315,9553X_4 + 214,5233X_5 \quad (10)$$

Modellen angiver, at det koster 2.855,5628 kr. at drive en husstandspumpe, 11.103,2371 kr. at drive en pumpestation i intervallet 0-10 l/s, 23.253,3823 kr. at drive en pumpestation i intervallet 11-100 l/s, 315,95536 kr. per l/s at drive en pumpestation i intervallet 101-600 l/s og 214,5233 kr. per l/s at drive en pumpestation i intervallet >600 l/s.

Kvalitetskontrol af omkostningsækvivalenten

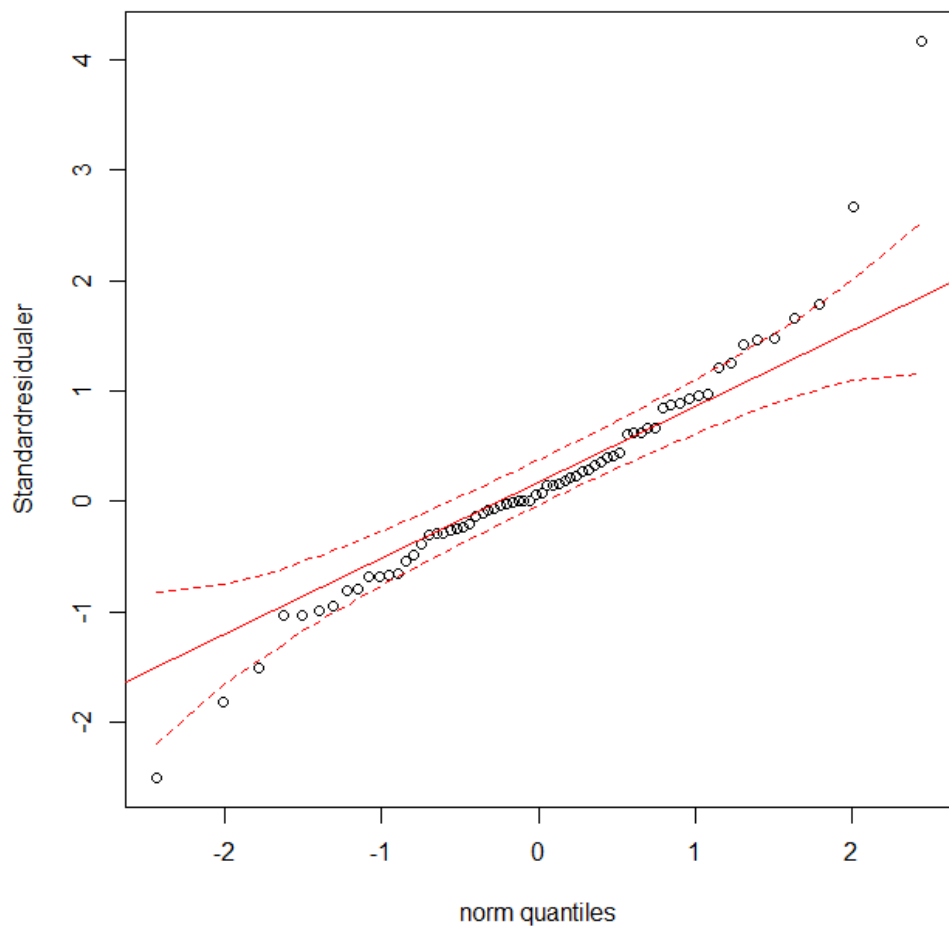
Tabel2 angiver, at alle variabler undtagen kapacitet >601 l/s er signifikante ved et 5 pct. signifikansniveau. Det betyder, at der er stor statistisk sikkerhed i modellens resultater. Den sidste

variabel er ikke signifikant, men har en lav p-værdi. Da denne model har gode statistiske egenskaber og giver intuitiv mening, har vi vurderet, at variabelens p-værdi kan accepteres. Det kan ses, at det bliver dyrere at drifte på store pumpestationer, og at alle β -værdier er positive.

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

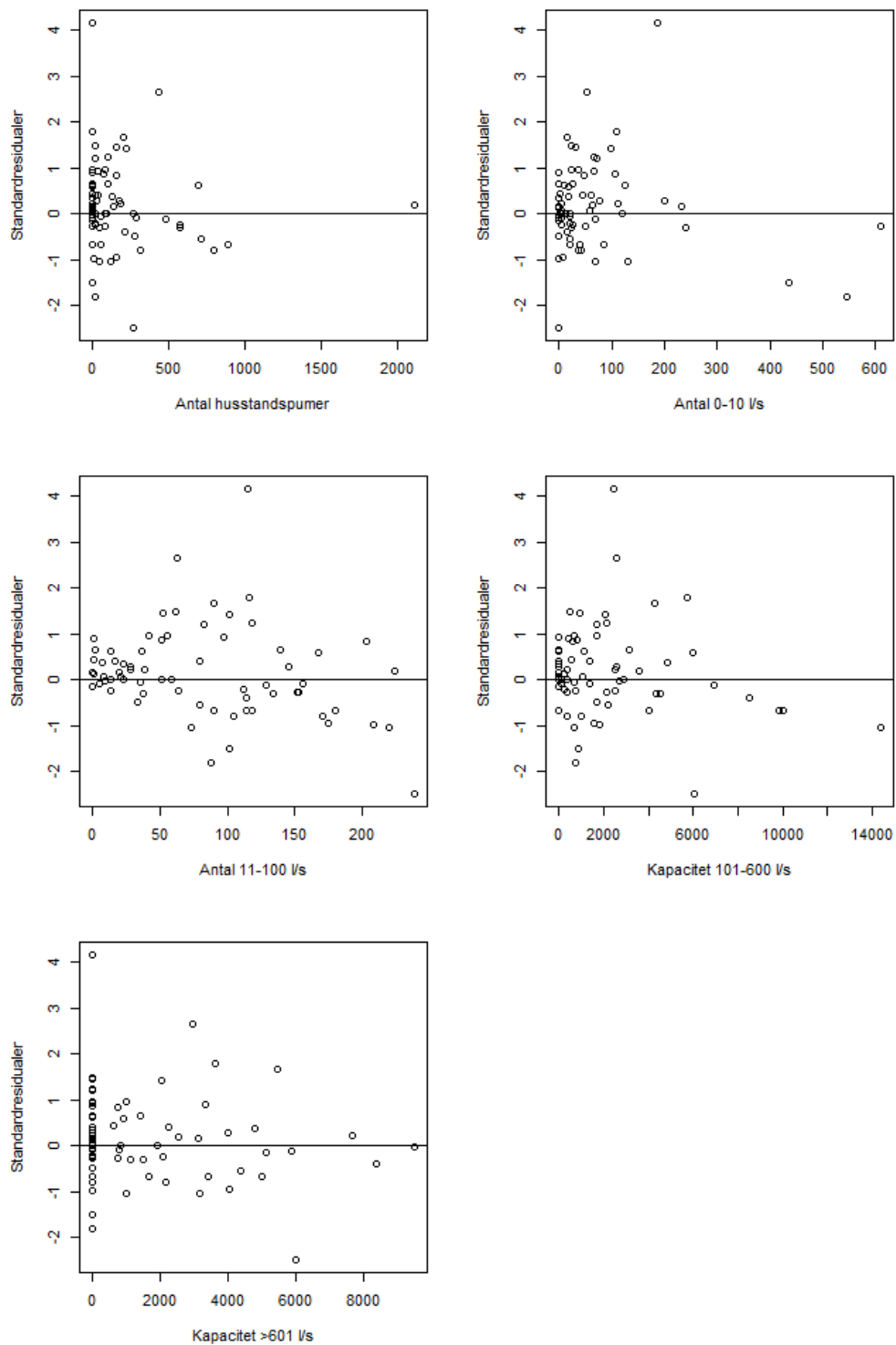
I Figur 3.4 ses det, at modellens residualer er pænt normalfordelte og at de fleste observationer ligger indenfor konfidensintervallet. Det er forventeligt, at der er få observationer som ligger udenfor konfidensintervallet og antagelsen om, at modellens residualer er normalfordelte, er dermed opfyldt.

Figur 3.4 **QQ-plot for pumpestationer**



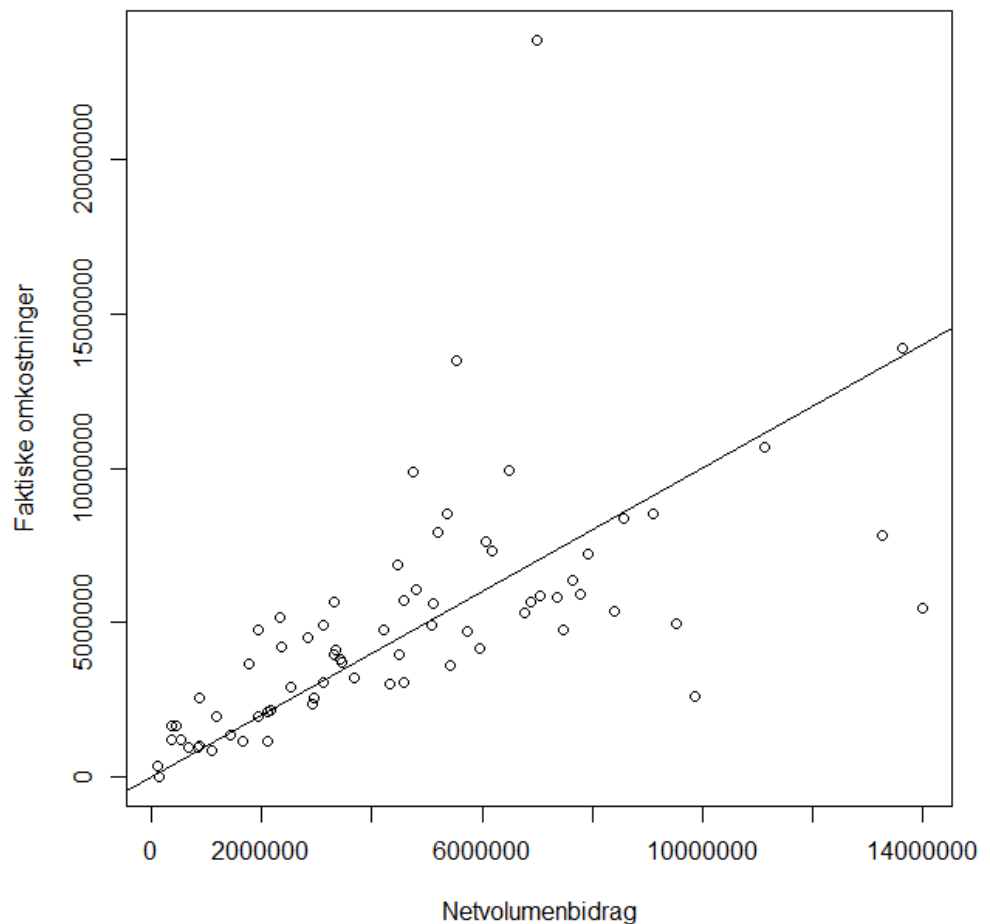
Herudover viser en Breusch-Pagan test, at der ikke er problemer med heteroskedasticitet. Dette kan ligeledes ses i Figur 3.5, hvor de standardiserede fejlede er plottet mod de observerede værdier.

Figur 3.5 Residualplot for hver variabel



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 vi har vurderet, at der er tilstrækkelig få af disse observationer til, at vi kan fortsætte med den valgte model.

Figur 3.6 **Observerede omkostninger mod beregnede**



3.4 Regnvandsbassiner

Modellens baggrund

Driftsomkostningerne ved at drive regnvandsbassiner forventes at være påvirket af antallet af regnvandsbassiner, det samlede vedligeholdelsesareal til regnvandsbassiner samt om regnvandsbassinerne ligger i land- eller byzonen.

Grundet høj korrelation mellem antal regnvandsbassiner og det samlede vedligeholdelsesareal har det været nødvendigt kun at se på et af de to underliggende forhold. Vi har vurderet, at den bedste model er den, hvor vi kun ser på det samlede vedligeholdelsesareal og dermed ikke antallet af regnvandsbassiner.

Der er ud over 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.

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

Y angiver de årlige driftsomkostninger ved at drive regnvandsbassiner, X_1 angiver vedligeholdelsesarealet i landzonen og X_2 angiver vedligeholdelsesarealet i byzonen. β -værdierne angiver enhedsomkostningerne for hvert af de underliggende forhold.

$$Y = \beta_1 X_1 + \beta_2 X_2 \quad (11)$$

I Tabel3 præsenteres resultaterne af modellen fra ligning (11).

Tabel3: Regressionsanalysens resultater for regnvandsbassiner

Variabel	Koefficient	Spredning	t-værdi	p-værdi
Vedligeholdelsesareal, landzonen	1,4890	0,7005	2,1260	0,037590*
Vedligeholdelsesareal, byzonen	2,5387	0,6457	3,9320	0,000218***

Antal observationer: 63
Justeret $R^2 = 0,4209$

Omkostningsækvivalenten for regnvandsbassiner

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

$$Y = 1,4890X_1 + 2,5387X_2 \quad (12)$$

Modellen angiver, at det koster 1,4890 kr. per m^2 at drive regnvandsbassiner i landzonen og 2,5387 kr. per m^2 at drive regnvandsbassiner i byzonen.

Kvalitetskontrol af omkostningsækvivalenten

Tabel3 angiver, at begge 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 regnvandsbassiner i byzonen i forhold til landzonen, hvilket opfylder vores forventninger i en zoneopdeling.

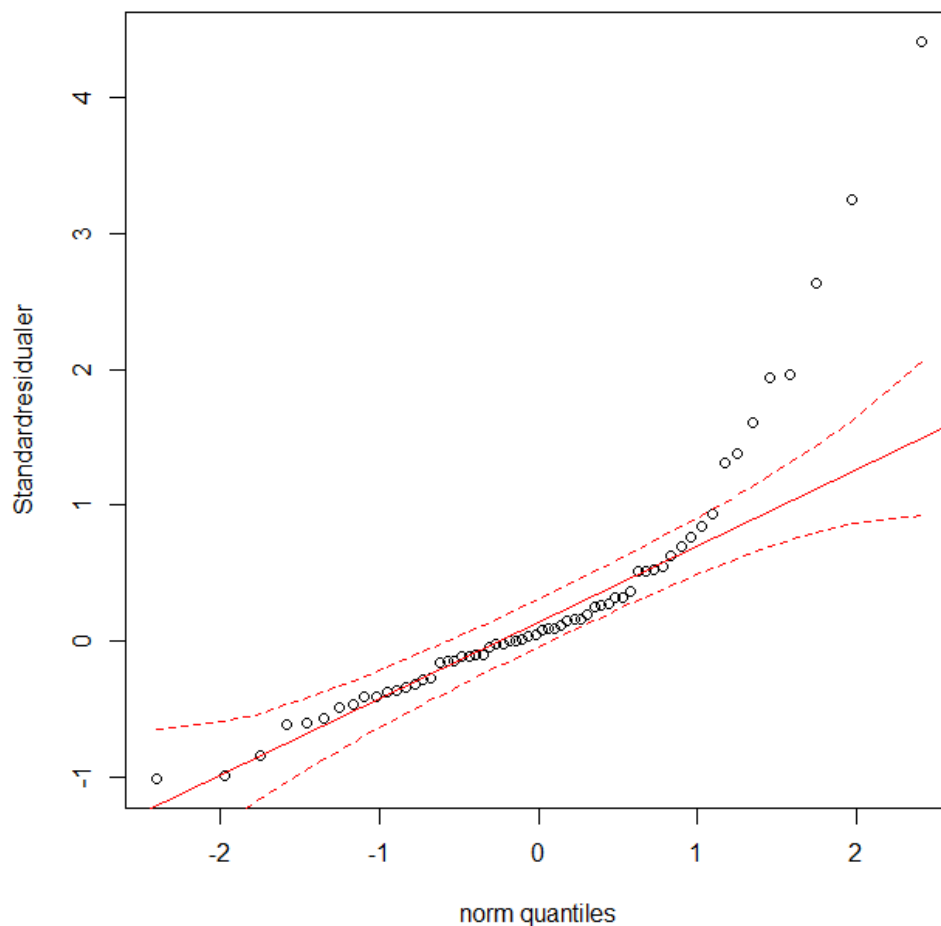
Ved brug af Cook's distance er der fjernet fire selskaber, som ellers ville have haft for stor indflydelse på modellens resultater.

I Figur 3.7 ses det, at modellens residualer ikke er pænt normalfordelt i den højre hale af et QQ-plot. Dette skyldes hovedsageligt, at selskaber med periodevise driftsomkostninger til regnvandsbassiner indgår i modellen. Disse selskaber har i nogle år høje omkostninger, imens de i andre år har lave omkostninger.

De fem selskaber, som ligger længst til højre, har alle periodevise driftsomkostninger i det år, som vi anvender tal fra. De har derfor specielt høje omkostninger i forhold til resten af selskaberne. Årsagen til at de alligevel indgår i modellen er, at der tilsvarende er en række selskaber med specielt lave omkostninger, fordi de først skal oprense deres bassiner i et kommende år. Der er dermed et flertal af selskaber, der trækker regressionslinjen ned, imens andre selskaber trækker den op. Da der er flere selskaber, der trækker regressionslinjen ned end op, er det forventeligt, at det kun er de sidste, som vil komme til udtryk i et QQ-plot.

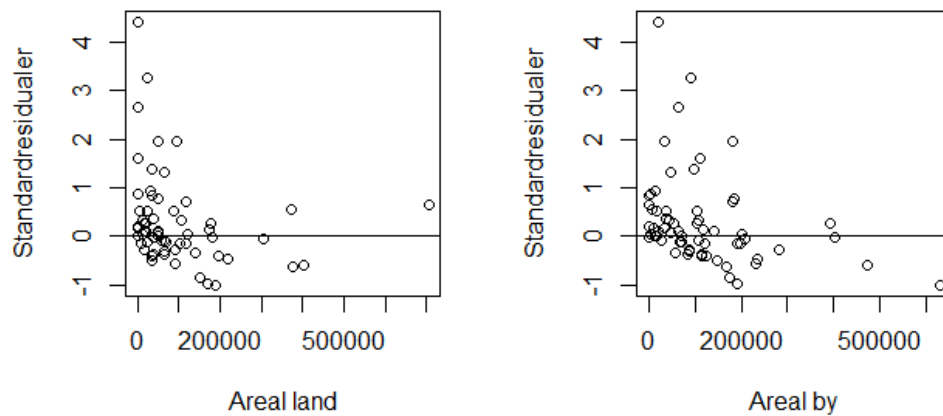
Hvis vi ser bort fra de fem selskaber med periodevise driftsomkostninger, er der kun tre selskaber, som ligger udenfor konfidensintervallet, som er markeret med de to røde striplet linjer. Dette er et forventeligt antal og et tegn på en god model.

Figur 3.7 QQ-plot for regnvandsbassiner

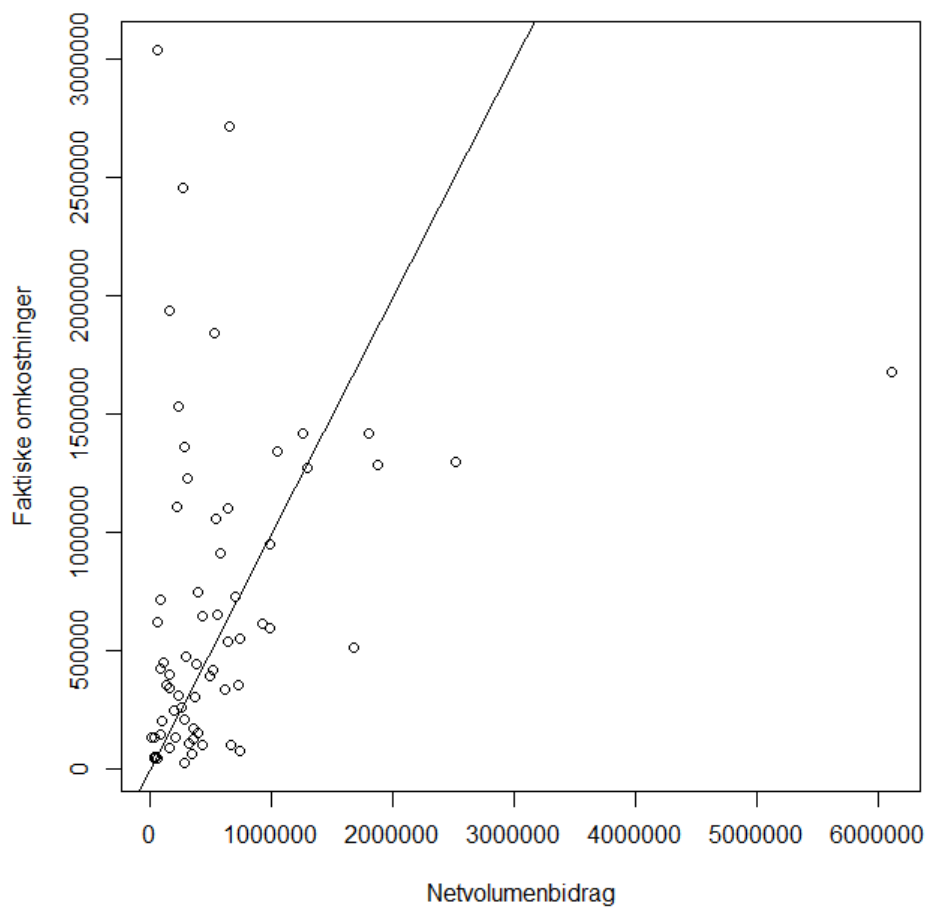


Herudover viser en Breusch-Pagan test, at der ikke er problemer med heteroskedasticitet. Dette kan ligeledes ses i Figur 3.8, hvor de standardiserede fejllid er plottet mod de observerede værdier.

Figur 3.8 Residualplot for hver variabel



Figur 3.9 Observerede omkostninger mod beregnede



3.5 Spildevandsbassiner

Modellens baggrund

Driftsomkostninger ved at drive spildevandsbassiner forventes at være påvirket af antallet af et selskabs spildevandsbassiner målt i stk. og den samlede volumen af spildevandsbassinerne målt i m³.

Vi er i arbejdet med at finde en model for costdriveren for spildevandsbassiner kommet frem til, at en model med både antal og samlet volumen ikke giver et statistisk godt resultat. Da det yderligere må formodes, at antal og volumen indeholder en del af den samme information, har vi valgt at lave en lineær model udelukkende med volumen som underliggende forhold. Andre modeller er undersøgt, både i forhold til forskellige kombinationer af underliggende forhold og forskellige transformationer af de underliggende forhold, men ingen andre modeller har de samme intuitive og statistiske fordele som denne.

Den valgte model er beskrevet i ligning (13). Y angiver de årlige driftsomkostninger ved at drive spildevandsbassiner og X_1 angiver den samlede volumen af et selskabs spildevandsbassiner. β -værdien angiver enhedsomkostningerne for det underliggende forhold.

$$Y = \beta_1 X_1 \quad (13)$$

I Tabel4 præsenteres resultaterne af modellen fra ligning (13).

Tabel4: **Regressionsanalysens resultater for spildevandsbassiner**

Variabel	Koefficient	Spredning	t-værdi	p-værdi
Volumen	6,1210	0,951	6,436	1,38e-08 ***

Antal observationer: 70
Justeret R²= 0,3661

Omkostningsækvivalenten for spildevandsbassiner

Omkostningsækvivalenten ved at drive renseanlæg kan ses i ligning (14).

$$Y = 6,121X_1 \quad (14)$$

Modellen angiver, at det koster 6,121 kr. at drive en kubikmeter spildevandsbassin.

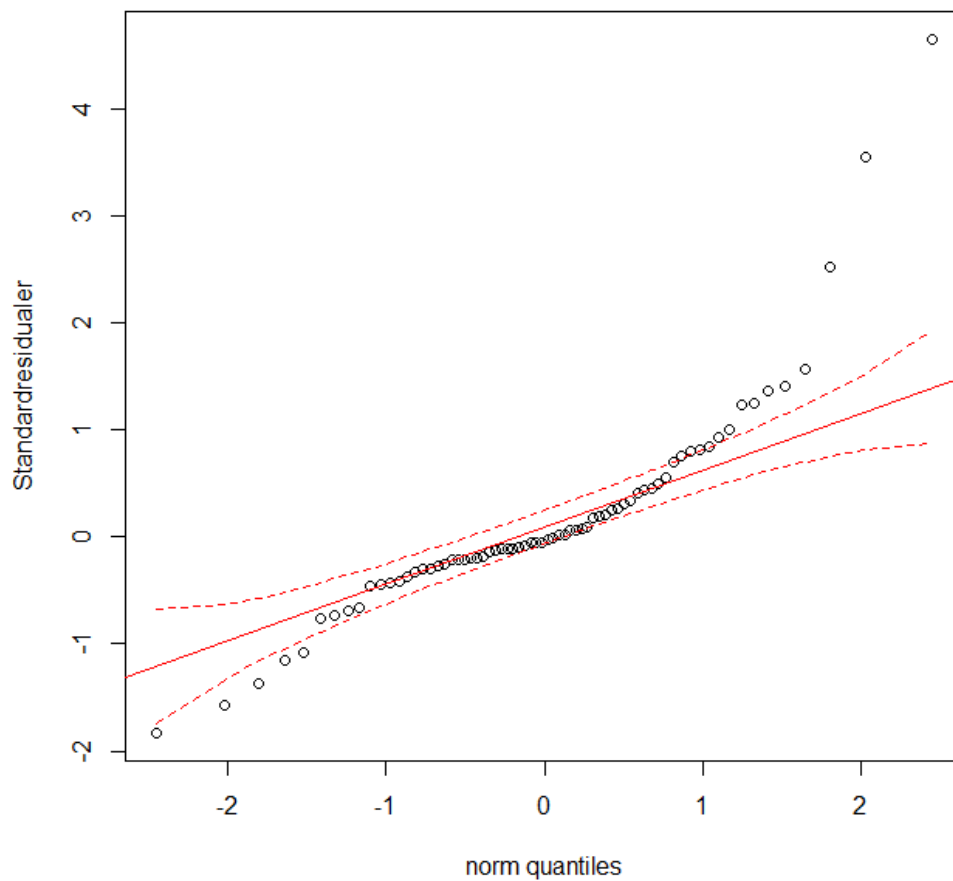
Kvalitetskontrol af omkostningsækvivalenten

Tabel4 angiver, at den samlede volumen af spildevandsbassiner er signifikant 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 tre selskaber, som ellers ville have haft for stor indflydelse på modellens resultater.

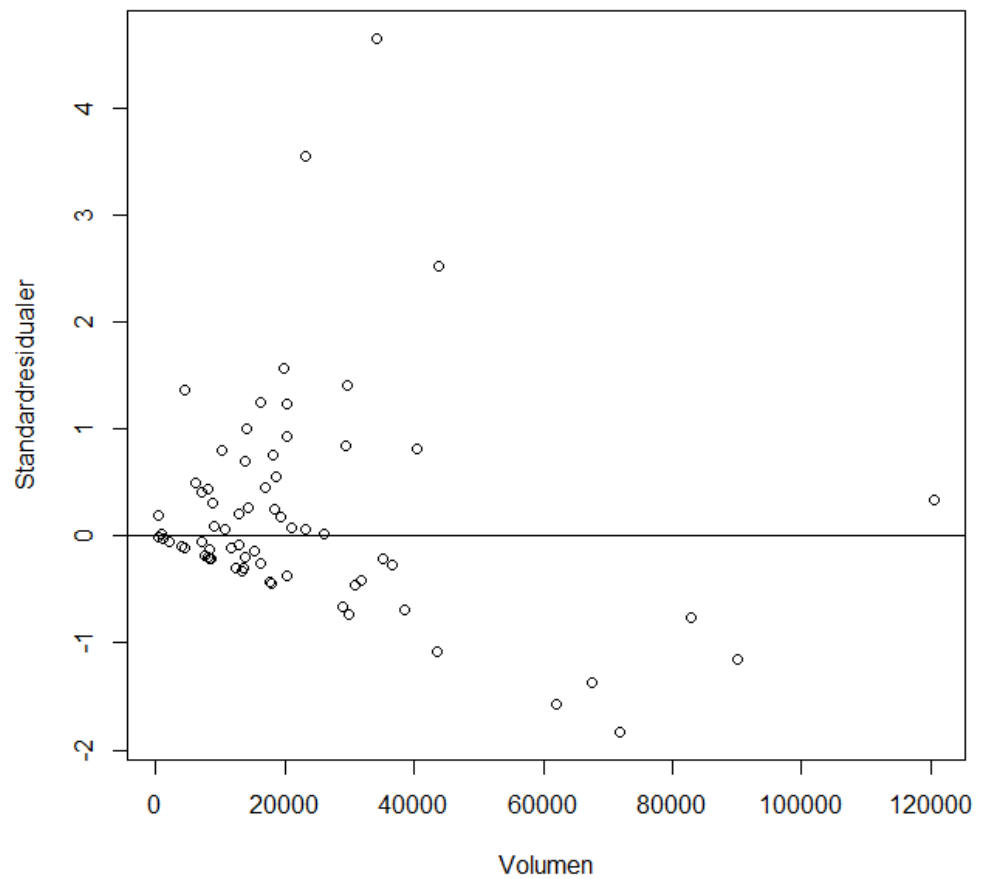
I Figur 3.10 ses det, at der er et problem med antagelsen om normalfordelte residualer i modellen. Det ser dog også ud til, at det er et fåtal af observationerne, der er problemer med. Der er forsøgt med en model, hvor disse selskaber er udeladt, men forskellen på de to modeller er minimal, og det er derfor valgt, at disse selskaber skal have mulighed for at påvirke omkostningsækvivalenten.

Figur 3.10 QQ-plot for spildevandsbassiner

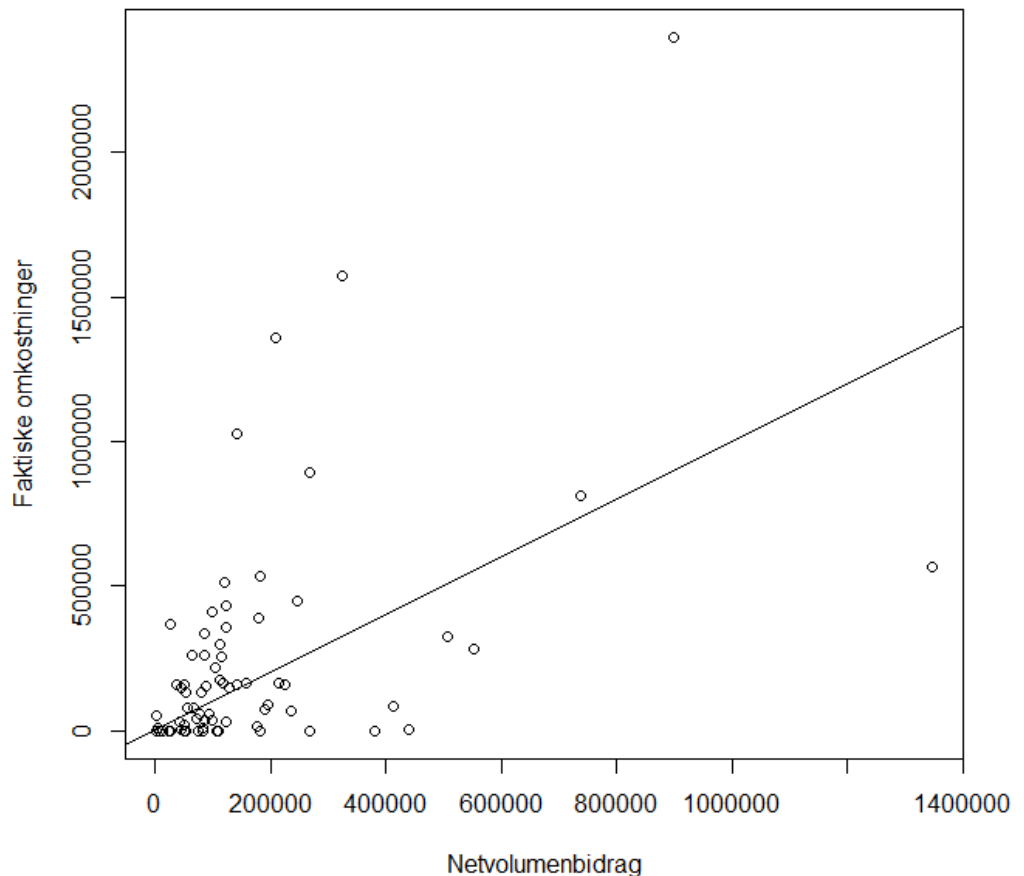


Figur 3.11 viser ligesom Figur 3.10, at det er et fåtal at selskaberne, der ikke passer helt ind i modellen.

Figur 3.11 Residualplot for hver variabel



Slutteligt sammenligner Figur 3.12 modellens beregnede omkostninger med de observerede omkostninger. Det ses, at der er en rimelig stor spredning på observationerne. Det stemmer godt overens med resultaterne af Figur 3.10 og Figur 3.11, at der er flere selskaber, der falder udenfor, og som modellen ikke passer så godt til.

Figur 3.12 **Observerede omkostninger mod beregnede**

3.6 Renseanlæg

Modellens baggrund

Et renselanlægs godkendte kapacitet indgår i modellen som udtryk for et renselanlægs størrelse. Tankegangen bag er, at den godkendte kapacitet udtrykker den maksimale belastning, et renselanlæg kan håndtere, mens den faktiske belastning er betinget af eksterne forhold, der påvirker belastningen af vandet, som renselanlægget skal rense. Vi forventer, at der eksempelvis er højere faste omkostninger forbundet med større renselanlæg, hvorfor vi ønsker både at medtage den godkendte kapacitet og den faktiske belastning i målet for et selskabs belastningsgrad.

Et renselanlægs godkendte kapacitet er angivet i personækvivalenter (PE) og er fastsat på baggrund af anlæggets gældende udledningskapacitet. Den faktiske belastning er ligeledes angivet i PE, men er opgjort på grundlag af gennemsnittet af de månedlige indløbsprøver for henholdsvis BOD, COD og kvælstof. 1 PE i BOD svarer til 60 g/dag, mens det svarer til henholdsvis 125

g/dag og 12 g/dag for COD og kvælstof. Standardværdierne for PE skal sikre en ensartet indberetning.³ Belastningstyperne er konverteret til PE af hensyn til sammenligneligheden.

I overensstemmelse med tidligere metode har vi valgt fortsat at benytte et gennemsnit mellem den godkendte kapacitet og den faktiske belastning som parameter for forholdet, der driver et renseanlægs driftsomkostninger. Vi anvender det niveau for den faktiske belastning, der har den højeste værdi for henholdsvis BOD, COD eller kvælstof. I forhold til tidligere tager modellen forbehold for, at der er selskaber, hvis niveau af COD eller kvælstof er højere end det målte BOD. Dermed undgår selskaberne fremadrettet at skulle søge om særlige forhold for dette. Samtidig opnår vi en mere retvisende model baseret på observationer for de højest målte belastningsgrader. I det følgende henviser betegnelsen PE til selskabets individuelle PE-niveau baseret på gennemsnitsmetoden beskrevet ovenfor. PE kan være baseret på BOD, COD eller kvælstof.

Modelspecifikation

Formålet med omkostningsækvivalenten er at finde en sammenhæng mellem driftsomkostninger og tilhørende underliggende forhold som kapacitet, belastning, rensetype og zoneplacering. Sagt på en anden måde ønsker vi at forklare, hvordan driftsomkostningerne påvirkes, når eksempelvis belastningsgraden ændrer sig. For en nærmere beskrivelse af den anvendte økonomiske metode samt de statistiske overvejelser se metodepapiret *OPEX-netvolumenmål* (august 2017)⁴.

I forbindelse med beregningen af omkostningsækvivalenten for minirenselanlæg har selskaberne indberettet antal minirenselanlæg og de tilhørende driftsomkostninger per selskab. Det betyder, at vi ikke kan foretage samme kontering som for renseanlæg, hvorfor renseanlæg og minirenselanlæg er estimeret hver for sig.

Det er fortsat vores forventning, at driftsomkostningerne ved at drive renseanlæg er påvirket af renseanlæggets kapacitet, belastningsgrad, rensetype og hvorvidt renseanlægget er placeret i land- eller byzonen. Rensetypen er angivet som en kombination af mekanisk (M), kemisk (K) og biologisk (B) rensning samt nitrifikation (N) og denitrifikation (D) per renseanlæg fordelt efter land- og byzone. Tankegangen bag opdelingen efter land- og byzone er, at det er dyrere at udføre renseaktivitet i byzonen eventuelt på grund af højere omkostninger til reduktion af lugtgener.

Modellen indeholder to input: Gennemsnittet mellem godkendt kapacitet og faktisk belastning målt i PE og zoneplacering. Zoneplacering er en såkaldt binær variabel. Det betyder, at den antager værdien 1 for renseanlæg placeret i landzonen, og 0 for renseanlæg i byzonen. Den valgte model er beskrevet i ligning 15. Y angiver de årlige driftsomkostninger ved at drive et renseanlæg, β_0 angiver skæringspunktet, X_1 er et gennemsnit mellem den godkendte kapacitet og den maksimale belastningsgrad målt i BOD, COD og N, mens X_2 markerer, om det pågældende renseanlæg er placeret i land- eller byzonen.

$$Y = \beta_0 X_1^{\beta_1} e^{\beta_2 X_2} \quad (15)$$

Det indebærer at estimere modellen

$$\log(Y) = \log(\beta_0) + \beta_1 \log(X_1) + \beta_2 X_2 \quad (16)$$

³ Oplysningerne stammer fra et rådgivende konsulenthus for sektoren, i forbindelse med ansøgninger om særlige forhold

⁴ <https://www.kfst.dk/media/46749/opex-netvolumenmaal.pdf>

Indledningsvist har vi testet, om der er signifikant forskel i de fundne effekter på tværs af de forskellige rensetyper. Resultaterne viser, at der ikke er en signifikant forskel på rensetyperne MBNDK og MBNK/MBND. Derudover har vi modtaget hørings svar til modellen, der angiver, at forskellen mellem MBNDK og MBNK/MBND kun har lille indflydelse på driftsomkostningerne, da begge typer anvendes til fjernelse af fosfor. Vi har på den baggrund slået de to rensetyper sammen til MBNDK/MBNK/MBND. Derudover har vi slået rensetyperne MB, MK, MBK og MBN sammen som følge af insignifikans mellem rensetyperne eller for få observationer. Det betyder, at rensetyperne er inddelt i følgende tre puljer: M, MB/MK/MBK/MBN og MBNDK/MBNK/MBND.

Efterfølgende har vi testet betydningen af zoneplaceringen. Resultaterne viser, at zoneplaceringen ikke har en signifikant betydning i nogen af modellerne. I modellen for MBNDK/MBNK/MBND er p-værdien dog så tæt på at være signifikant, at vi har besluttet at inkludere parameteren i denne model alligevel. Der er dermed lagt vægt på den intuitive vurdering i denne model. I de andre modeller er zoneplaceringen så insignifikant, at der efter vores vurdering ikke er grund til at inkludere den. Modellen for MBNDK/MBNK/MBND adskiller sig dermed fra de øvrige rensetyper, da dennes model som den eneste indeholder zoneplacering.

Modellen for M og MB/MK/MBK/MBN er derfor, i modsætning til ligning 15, givet ved:

$$Y = \beta_0 X_1^{\beta_1} \quad (17)$$

Det indebærer at estimere modellen

$$\log(Y) = \log(\beta_0) + \beta_1 \log(X_1) \quad (18)$$

Det gælder for alle kombinationer af rensetyper samt for minirens anlæg, at vi har undersøgt data for outliers. Det er vores vurdering, at datasættet ikke indeholder outliers, hvorfor samtlige rens anlæg indgår i beregningerne.

Model for MBNDK/MBNK/MBND

Omkostningsækvivalenten ved at drive rens anlæg af rens typen MBNDK/MBNK/MBND er givet af Tabel 3.5.

Tabel 3.5: Regressionsanalysens resultater for rens typen MBNDK/MBNK/MBND

Variabel	Koefficient	Spredning	t-værdi	p-værdi
Skæringspunkt	7,3705	0,2313	31,862	< 2e-16 ***
Log(PE)	0,7360	0,02225	33,073	< 2e-16 ***
Landzone	-0,1354	0,07523	-1,799	0,0731

Antal observationer: 268

Justeret R²= 0,8451

Kappa = 1,1412

Omkostningsækvivalenten for et rens anlæg af rens typen MBNDK/MBNK/MBND fås ved indsættelse af resultaterne fra tabel 1 i ligning 2 og er givet af ligning 19:

$$Y = e^{7,3705} \cdot PE^{0,7360} \cdot e^{-0,1354 \cdot Landzone} \cdot 1,1412 \quad (19)$$

Det tredje led på højre side indeholder den binære variabel *landzone*. Det tilhørende parameterestimat på -0,1354, jf. tabel 5, angiver forskellen i omkostningsniveauet mellem rens anlæg

i land- og byzonen. Det negative parameterestimat indikerer, at der er lavere omkostninger forbundet med at drive renseaktiviteter i landzonen end i byzonen. Det fremgår ved at indsætte henholdsvis 0 og 1 på *landzones* plads i ligningen. For renseanlæg i byzonen svarer det til, at tredje led bliver $e^{-0,1354 \cdot 0} = 1$, mens det for renseanlæg i landzonen svarer til $e^{-0,1354 \cdot 1} = 0,8734$. Fjerde led er en konstant, der udgør korrektionsfaktoren. Den er beregnet på baggrund af standardafvigelsen og er betegnet ved $kappa^5$. Det giver to modeller for rensetyper MBNDK/MBNK/MBND:

$$\text{Land: } Y = 1588,4278 \cdot PE^{0,7360} \cdot 0,8734 \cdot 1,1412 = 1583,1635 \cdot PE^{0,7360} \quad (20)$$

$$\text{By: } Y = 1588,4278 \cdot PE^{0,7360} \cdot 1,1412 = 1812,7138 \cdot PE^{0,7360} \quad (21)$$

Model for MB/MK/MBK/MBN

Omkostningsækvivalenten ved at drive renseanlæg af rensetyper MB/MK/MBK/MBN er givet af tabel 6:

Tabel 6: **Regressionsanalysens resultater for rensetyper MB/MK/MBK/MBN**

Variabel	Koefficient	Spredning	t-værdi	p-værdi
Skæringspunkt	7,8289	0,4433	17,662	< 2e-16 ***
Log(PE)	0,6289	0,0694	9,068	4,95e-13 ***

Antal observationer: 65
Justeret R²= 0,711
Kappa = 1,2877

Omkostningsækvivalenten for et renseanlæg af rensetyper MB/MK/MBK/MBN fås ved indsættelse af resultaterne fra tabel 6 i ligning 18 og er givet af ligning 22:

$$Y = e^{7,8289} \cdot PE^{0,6289} \cdot 1,2877 = 3234,9142 \cdot PE^{0,6289} \quad (22)$$

Model for M

Omkostningsækvivalenten ved at drive renseanlæg af rensetyper M er givet af tabel 7:

Tabel 7: **Regressionsanalysens resultater for rensetyper M**

Variabel	Koefficient	Spredning	t-værdi	p-værdi
Skæringspunkt	7,7786	1,0997	7,074	1,87e-06 ***
Log(PE)	0,5190	0,2393	2,169	0,0446 *

⁵ Kappa beregnes ved at indsætte standardafvigelsen (σ) fra estimeringen i formelen: $\exp(\frac{\sigma^2}{2})$

Antal observationer: 19
 Justeret R²= 0,1706
 Kappa = 1,3411

Omkostningsækvivalenten for et renseanlæg af rensetypen M fås ved indsættelse af resultaterne fra tabel 7 i ligning 4 og er givet af ligning 23:

$$Y = e^{7,7786} \cdot PE^{0,5190} \cdot 1,3411 = 3203,7913 \cdot PE^{0,5190} \quad (23)$$

Modellen for minirenselanlæg

Minirenselanlæg er renseanlæg med en kapacitet under 30 PE. Selskaberne har i forbindelse med den ekstraordinære indberetning til brug for genberegning af omkostningsækvivalenten indberettet deres samlede omkostninger ved at drive minirenselanlæg samt antallet af minirenselanlæg. Vi har vurderet, at et indberettet omkostningsniveau på 0 kroner ikke er repræsentativt, hvorfor vi har valgt at fjerne disse observationer fra analysen.

Den valgte model er beskrevet af ligning 24. Y angiver de årlige driftsomkostninger ved at drive minirenselanlæg, og X_1 er antallet af minirenselanlæg.

$$Y = \beta_1 X_1 \quad (24)$$

Modellen inkluderer ikke et konstantled, da vi vurderer, at der kun er omkostninger forbundet med minirenselanlæg, såfremt et selskab faktisk driver et eller flere minirenselanlæg. Derfor er konstantleddet overflødigt. Omkostningsækvivalenten ved at drive minirenselanlæg kan ses af tabel 8:

Tabel 8: **Regressionsanalysens resultater for minirenselanlæg**

Variabel	Koefficient	Spredning	t-værdi	p-værdi
Antal minirenselanlæg	2863,2070	315,7	9,068	4,06e-05 ***

Antal observationer: 8
 Justeret R²= 0,9103

De endelige modeller

Omkostningsækvivalenten for renseanlæg dækker over fem modeller: Fire for renseanlæg og en for minirenselanlæg. De er givet ved:

Rensetype M: $Y = 3203,7913 \cdot PE^{0,5190}$

Rensetype MB/MK/MBK/MBN: $Y = 3234,9142 \cdot PE^{0,6289}$

Rensetype MBNDK/MBNK/MBND i landzone: $Y = 1583,1635 \cdot PE^{0,7360}$

Rensetype MBNDK/MBNK/MBND i byzone: $Y = 1812,7138 \cdot PE^{0,7360}$

Antal minirenselanlæg: $Y = 2863,2070 \cdot \text{Antal Minirenselanlæg}$

Selskaberne kan beregne det samlede netvolumenmål for costdriveren *renseanlæg* ved indsætte belastningsgraden for de respektive rensetyper samt eventuelle minirenselanlæg i de endelige modeller ovenfor.

Eksempel

Et selskab har tre renselanlæg og otte minirenselanlæg. De tre renselanlæg udfører rens af typen M samt MBNDK i henholdsvis land- og byzonen.

Det første renselanlæg af typen M har en godkendt kapacitet på 300 PE og har derudover en faktisk belastning på 200 PE målt i BOD, 250 PE målt i COD og 170 PE målt i N.

Det andet renselanlæg af typen MBNDK i landzonen har en godkendt kapacitet på 10.000 PE og har derudover en faktisk belastning på 9.000 PE målt i BOD, 8.000 PE målt i COD og 7.500 PE målt i N.

Det tredje renselanlæg af typen MBNDK i byzonen har en godkendt kapacitet på 20.000 PE og har derudover en faktisk belastning på 16.000 PE målt i BOD, 14.000 PE målt i COD og 19.000 PE målt i N.

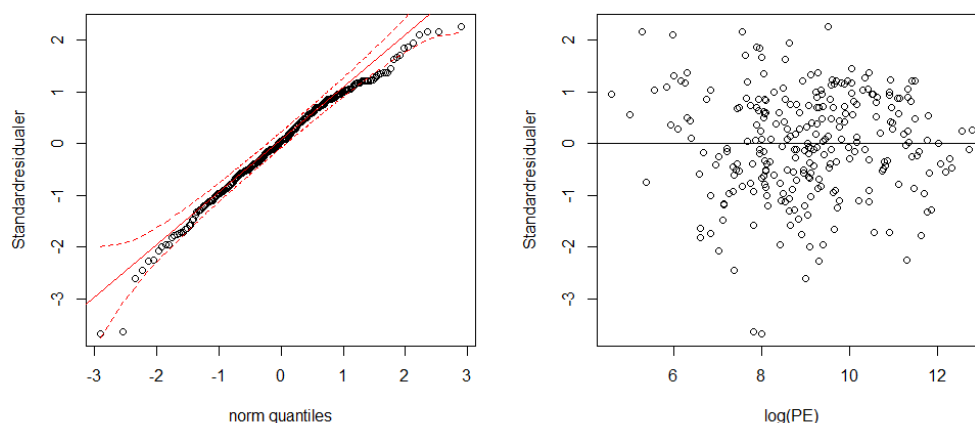
Det giver selskabet et samlet netvolumenmål for costdriveren *renselanlæg* givet ved:

$$\begin{aligned}
 Y &= 3203,7913 \cdot \left(\frac{300 + \max\{200, 250, 170\}}{2} \right)^{0,5190} + 1583,7138 \\
 &\quad \cdot \left(\frac{10.000 + \max\{9.000, 8.000, 7.500\}}{2} \right)^{0,7360} + 1812,7138 \\
 &\quad \cdot \left(\frac{20.000 + \max\{16.000, 14.000, 19.000\}}{2} \right)^{0,7360} + 2863,2070 \cdot 8 \\
 &= 4.027.018,7245
 \end{aligned}$$

Nedenfor fremgår QQ- og residualplots for hver af de fire modeller. For en nærmere beskrivelse af QQ- og residualplots' betydning for en statistisk model se kapitel 2.

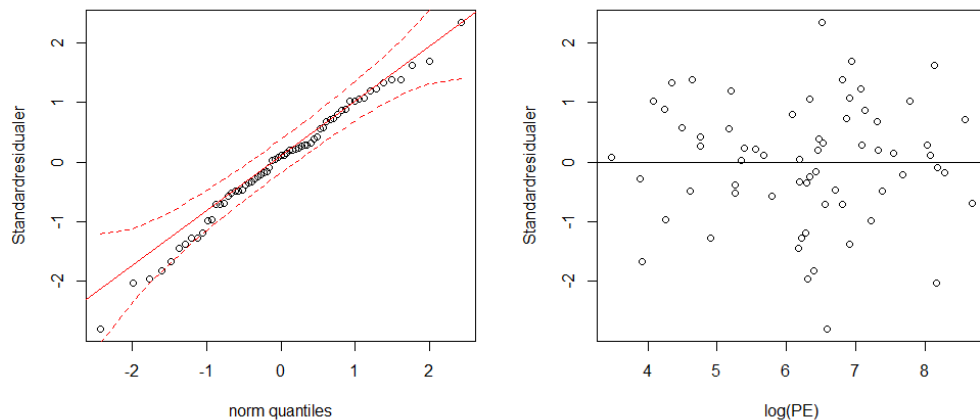
Det ses af Figur 3.13-3.16, at modellens residualer er pænt normalfordelte og ligger inden for 95 pct. konfidensintervallet samt, at der ikke er tegn på at transformationer af data vil give en bedre model. Det støtter os i overbevisningen om, at det er en god model.

Figur 3.13: QQ- og residualplot for rensetyper MBNDK/MBNK/MBND



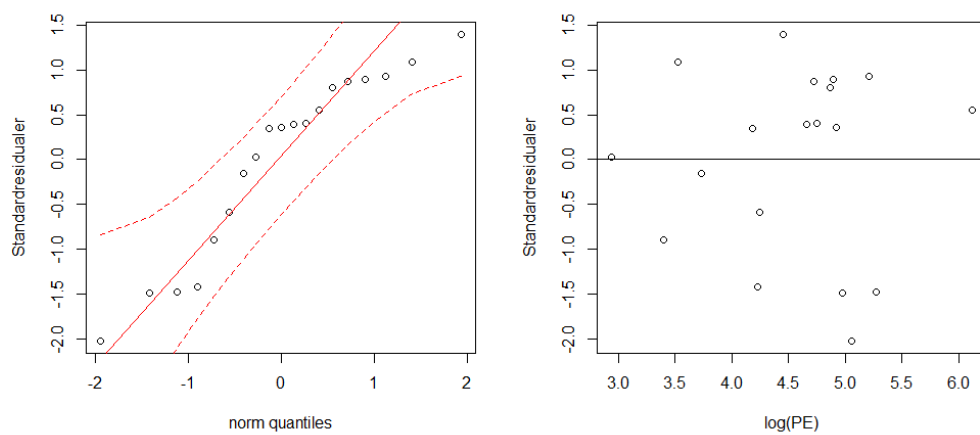
Kilde: Forsyningssekretariatet

Figur 3.14: QQ- og residualplot for rensetypen MB/MK/MBK/MBN



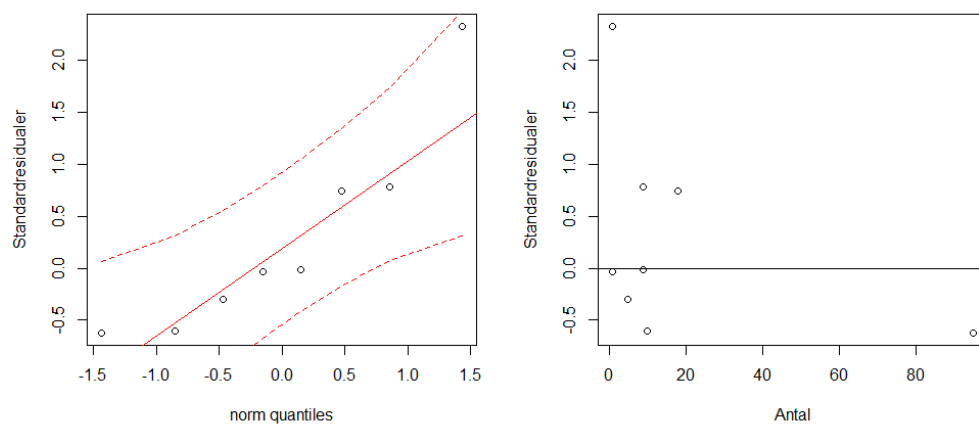
Kilde: Forsyningssekretariatet

Figur 3.15: QQ- og residualplot for rensetypen M



Kilde: Forsyningssekretariatet

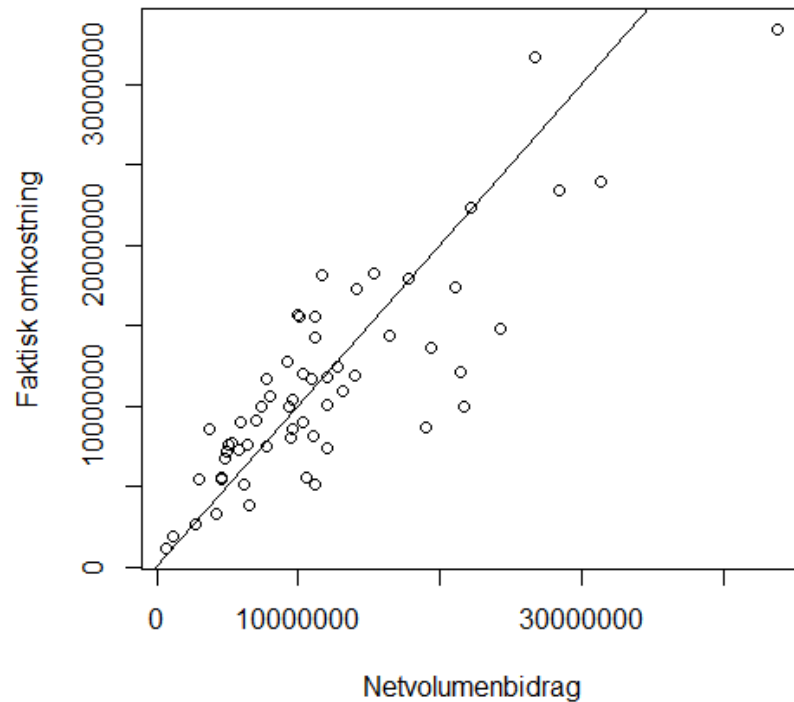
Figur 3.16: QQ- og residualplot for minirenselanlæg



Kilde: Forsyningssekretariatet

Slutteligt sammenligner Figur 3.17 modellens beregnede omkostninger med de observerede omkostninger. De beregnede værdier følger de faktiske omkostninger i overvejende grad og med kun få egentlige outliere.

Figur 3.137: Faktiske omkostninger mod beregnede



Kilde: Forsyningssekretariatet

3.7 Slambehandling

Modellens baggrund

Driftsomkostninger ved slambehandling forventes at være påvirket af mængden af slam målt i tons tørstof (tTS) og metoden, hvorpå slammet bliver behandlet.

I indberetningen til arbejdet med den reviderede OPEX-benchmarkingmodel var det muligt for selskaberne at angive omkostninger, der specifikt var forbundet til oprensning af slambede. Da der har været meget få indberetninger, har vi været nødsaget til at slå disse omkostninger sammen med de øvrige driftsomkostninger forbundet med slambehandling. Derudover har vi vurderet, at der er en række faste omkostninger i forbindelse med slambehandling, og derfor har vi inkluderet et konstantled i modellen.

Vi har i forbindelse med valget af modellen testet en række andre modeller. Den valgte model er den, som vi har vurderet til intuitivt og statistisk at være bedst.

Den valgte model er beskrevet i ligning (25). Y angiver de årlige driftsomkostninger forbundet med slambehandling, X_1 angiver tons tørstof slam behandlet ved normal slambehandling, X_2 angiver tons tørstof slam behandlet ved anaerob udrådning og X_3 angiver tons tørstof slam behandlet ved tilførsel til slammineraliseringsanlæg. β -værdierne angiver enhedsomkostningerne for hvert af de underliggende forhold.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \quad (25)$$

I Tabel 9 præsenteres resultaterne af modellen fra ligning (25).

Tabel 9: **Regressionsanalysens resultater for slambehandling**

Variabel	Koefficient	Spredning	t-værdi	p-værdi
Konstantled	738.847,7915	314.746,6	2,347	0,0223 *
Normal behandling	277,6197	262,8	1,056	0,2951
Anaerob udrådning	767,2944	110,1	6,967	3,07e-09 ***
Slammineraliseringsanlæg	1.660,4795	730,5	2,273	0,0267 *

Antal observationer: 63

Justeret $R^2 = 0,449$

Omkostningsækvivalenten for slambehandling

Omkostningsækvivalenten ved at udføre slambehandling kan ses i ligning (26).

$$Y = 738.847,7915 + 277,6197X_1 + 767,2944X_2 + 1.660,4795X_3 \quad (26)$$

Modellen angiver, at der i gennemsnit er en fast omkostning på 738.847,7915 kr. ved at udføre slambehandling. Derefter koster det 277,6197kr. pr. tons tørstof at udføre normal slambehandling, 767,2944 kr. pr. tons tørstof at udføre anaerob udrådning og 1.660,4795 kr. pr. tons tørstof ved tilførsel til slammineraliseringsanlæg.

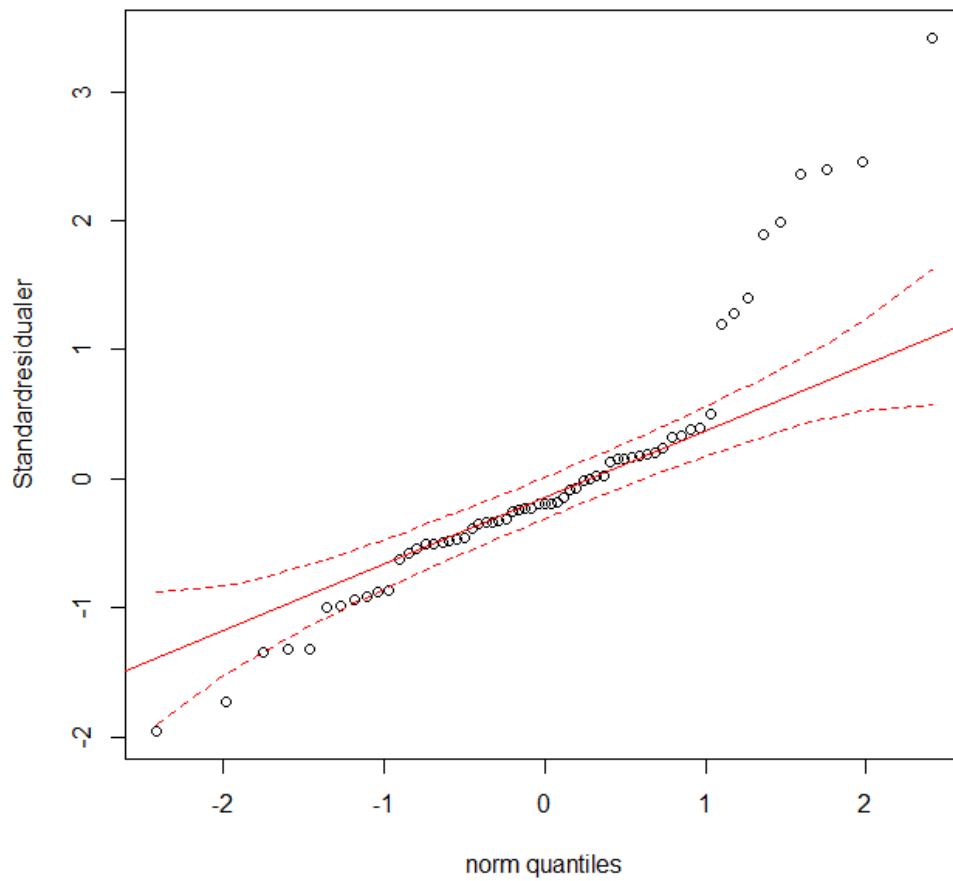
Kvalitetskontrol af omkostningsækvivalenten

Tabel 6 angiver, at ikke alle variable er signifikante ved et 5 pct. signifikansniveau. Vi har ikke kunnet finde en model, der kan kombinere det intuitive i, at der er forskel i driftsomkostningerne ved at udføre slambehandling på forskellige måder og samtidig opretholde en signifikans i alle variable. Vi har valgt, at en intuitiv model, hvor der er forskel på de forskellige typer af slambehandling, vejer højere end en model, hvor alle variable er signifikante på et 5 pct. signifikansniveau.

Ved brug af Cook's distance er der fjernet tre selskaber, som ellers ville have haft for stor indflydelse på modellens resultater. Derudover har vi vurderet, at yderligere et enkelt selskab bør fjernes, da det på baggrund af QQ- og residualplot fremstod som en ekstrem observation.

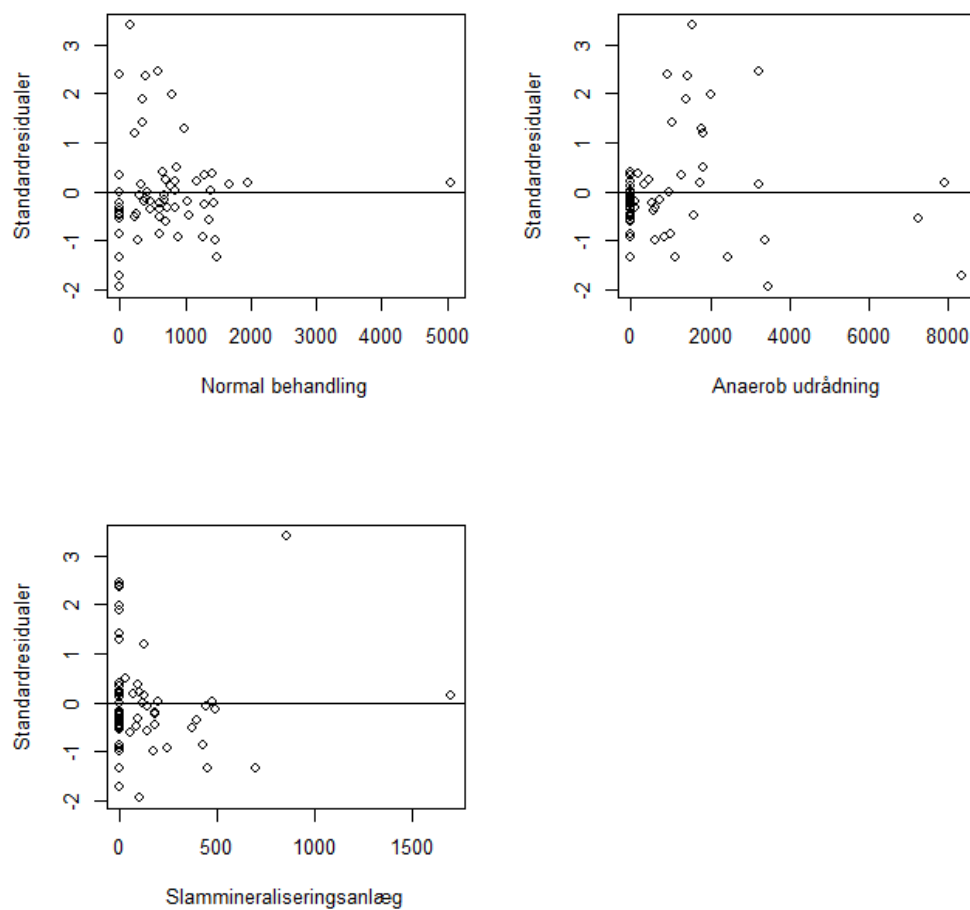
I Figur 3.18 ses det, at langt de fleste af modellens residualer ligger indenfor 95 pct. konfidensintervallet af normalfordelingen. Der er enkelte observationer, der ligger udenfor konfidensintervallet, men det er vores vurdering, at det ikke er nok til at rejse tvivl om modellen.

Figur 3.148 QQ-plot for slambehandling



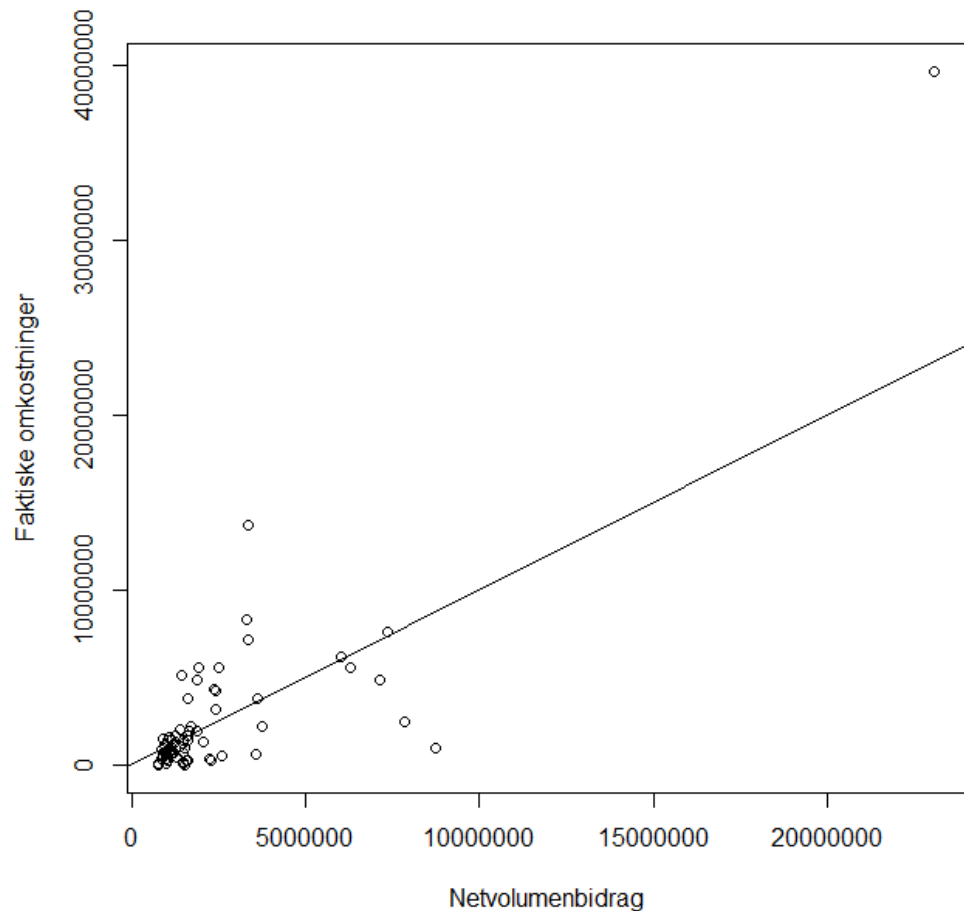
Figur 3.19 viser ingen tegn på, at en transformation af data ville give en bedre model. Ligesom med QQ-plottet antyder det, at der er enkelte observationer, der falder udenfor, men ikke nok til at betvivle valget af model. En Breusch-Pagan test viser, at der ikke er problemer med heteroskedasticitet.

Figur 3.159 Residualplot for hver variabel



Slutteligt viser Figur 3.20, at modellens beregnede omkostninger stemmer godt overens med de observerede omkostninger. Enkelte observationer skaber en del varians i modellen, men alt i alt så passer den godt på størstedelen af observationerne.

Figur 3.20 Faktiske omkostninger mod beregnede



3.8 Slamdisponering

Modellens baggrund

Driftsomkostninger ved slamdisponering forventes at være påvirket af mængden af slam målt i tons tørstof (tTS), metoden hvorpå slammet er disponeret, og om slammet er disponeret fra tømning af slambede eller ej. Derudover er det undersøgt, om det har indflydelse på driftsomkostningerne at have eget slamlager.

Vi har set os nødsaget til at slå mængden af slam disponeret fra tømning af slambede og mængden af slam disponeret andre steder fra sammen. Dette er gjort, da der har været meget få indberetninger fra tømning af slambede til at kunne vurdere dette særskilt. Det betyder, at vi ser på mængden af slam disponeret til jordbrugsformål, kompostering og forbrænding/deponering, uanset om den kommer fra tømning af slambede eller ej. Derudover har det vist sig, at det ikke har indflydelse på driftsomkostningerne, om et selskab ejer et slamlager eller ej. Vi har vurderet, at der ligeledes er en række faste omkostninger i forbindelse med slamdisponering, og derfor har vi inkluderet et konstantled i modellen.

Vi har i forbindelse med valget af modellen testet en række andre modeller. Den valgte model er den, som vi har vurderet til intuitivt og statistisk at være bedst.

Den valgte model er beskrevet i ligning (27). Y angiver de årlige driftsomkostninger forbundet med slamdisponering, X_1 angiver tons tørstof slam disponeret til jordbrugsformål, X_2 angiver tons tørstof slam disponeret til kompostering og X_3 angiver tons tørstof slam disponeret til forbrænding eller deponering. β -værdierne angiver enhedsomkostningerne for hvert af de underliggende forhold.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \quad (27)$$

I Tabel 10 præsenteres resultaterne af modellen fra ligning (27).

Tabel 10: Regressionsanalysens resultater for slamdisponering

Variabel	Koefficient	Spredning	t-værdi	p-værdi
Konstantled	336.440,7036	297.169,2	1,132	0,2623
Jordbrugsformål	997,3105	221,2	4,508	3,31e-05 ***
Kompostering	2.945,6348	246,7	11,941	< 2e-16 ***
Forbrænding/Deponering	1.952,3183	751,4	2,598	0,0119 *

Antal observationer: 61

Justeret $R^2 = 0,7071$

Omkostningsækvivalenten for slamdisponering

Omkostningsækvivalenten ved slamdisponering kan ses i ligning (28).

$$Y = 336.440,7036 + 997,3105X_1 + 2.945,6348X_2 + 1.952,3183X_3 \quad (28)$$

Modellen angiver, at der i gennemsnit er faste omkostninger på 336.440,7036 kr. forbundet med slamdisponering. Derefter koster det 997,3105 kr. pr. tons tørstof, der disponeres til jordbrugsformål, 2.945,6348 kr. pr. tons tørstof, der disponeres til kompostering og 1.952,3183 kr. pr. tons tørstof, der disponeres til forbrænding eller disponering.

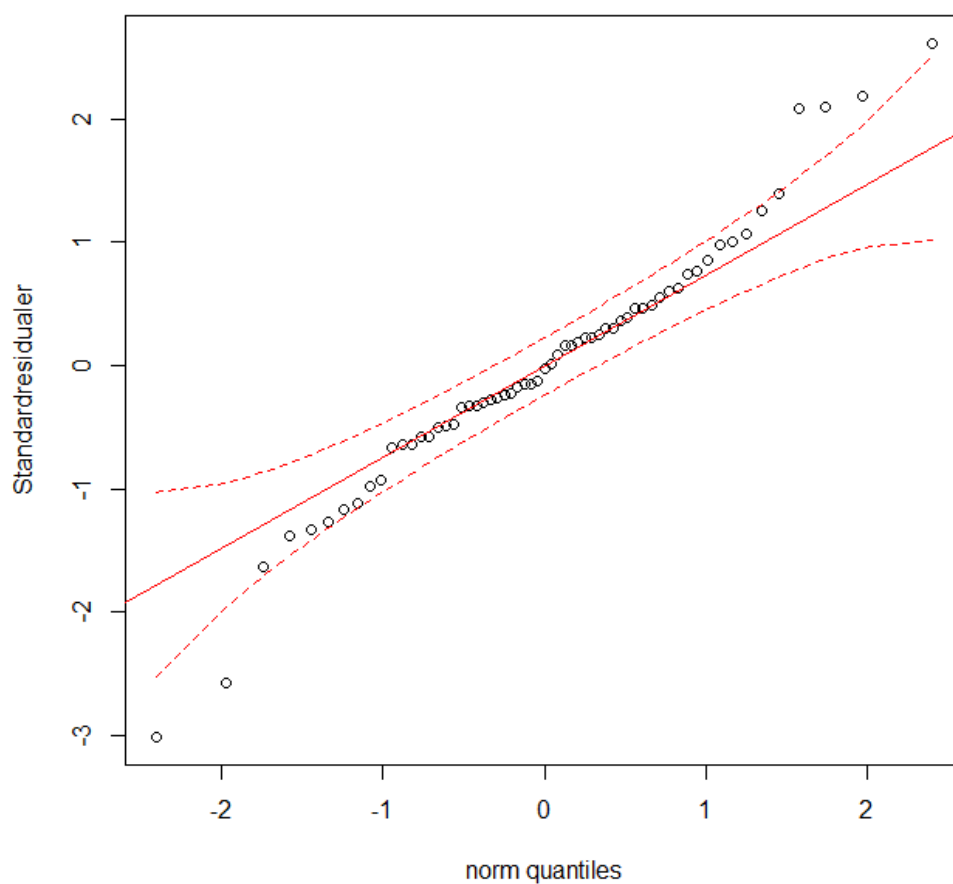
Kvalitetskontrol af omkostningsækvivalenten

Tabel 7 angiver, at alle de underliggende forhold er signifikante ved et 5 pct. signifikansniveau. Dog er konstantledet ikke signifikant ved 5 pct.-niveauet. Vi har vurderet, at det intuitivt giver mening at have et konstantled med i modellen, og at det vejer højere, end at konstantledet ikke er statistisk signifikant.

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

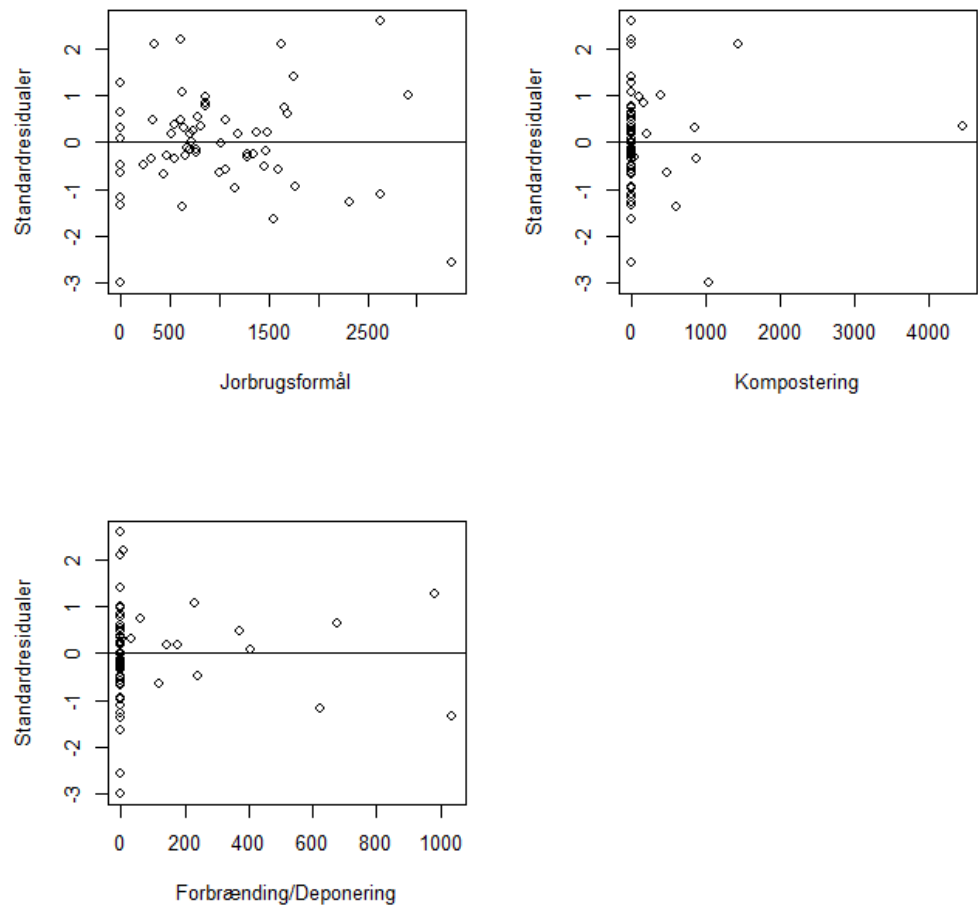
I Figur 3.21 ses det, at modellens residualer er pænt normalfordelte. Der er enkelte observationer, der ligger udenfor 95 pct. konfidensintervallet, men det er ikke nok til at rejse tvivl om normalfordelingsantagelsen af residualerne.

Figur 3.21 QQ-plot for slamdisonering



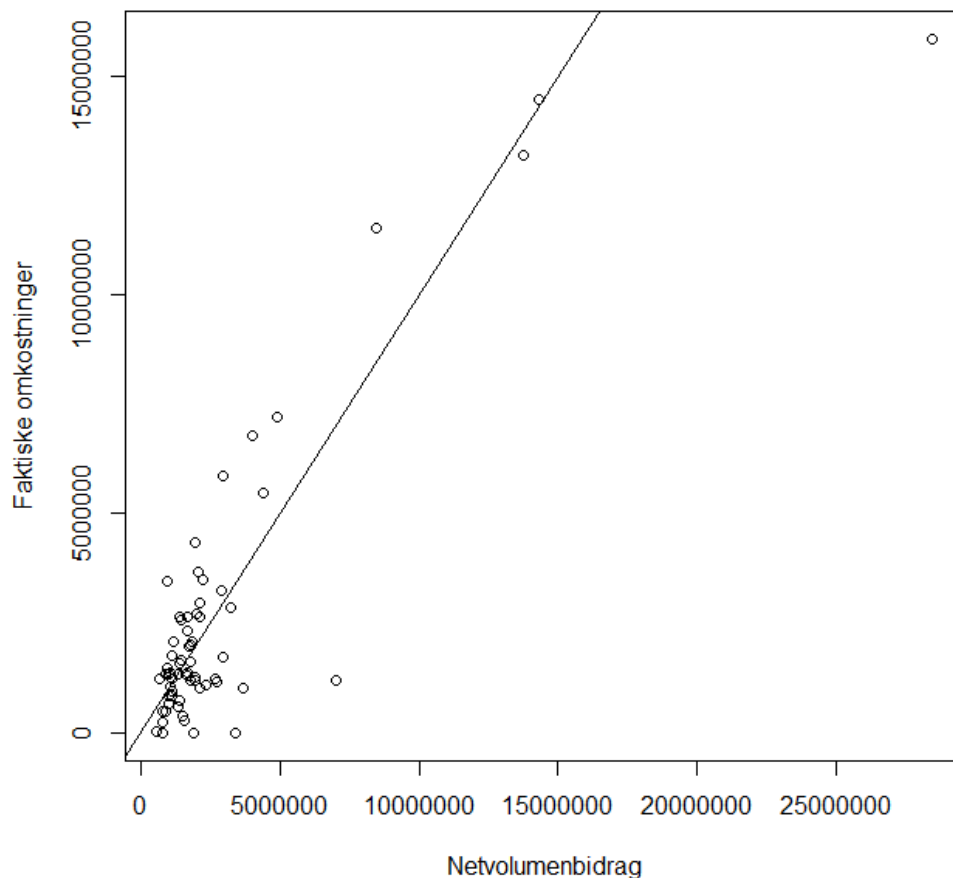
Figur 3.22 viser en fin fordeling af standardresidualerne og giver ikke anledning til at tro, at en transformation af data er nødvendig. En Breusch-Pagan test viser, at der ikke er problemer med heteroskedasticitet.

Figur 3.22 Residualplot for hver variabel



Slutteligt sammenligner Figur 3.23 modellens beregnede omkostninger med de observerede omkostninger. Det ses, at de beregnede omkostninger stemmer godt overens med de observerede omkostninger. Der er dog en enkelt observation, der falder udenfor modellen.

Figur 3.23 Faktiske omkostninger mod beregnede



3.9 Målere og kunder

Modellens baggrund

Driftsomkostningerne, som vedrører målere og kunder, forventes at være påvirket af antallet af kunder, som selskabet modtager spildevand fra. 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 kundefølgelse 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 (21).

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

$$Y = \beta_0 X_1^{\beta_1} \quad (29)$$

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

$$\log(Y) = \log(\beta_0) + \beta_1 \log(X_1) \quad (29)$$

De estimerede værdier for modellen er præsenteret i Tabel 11.

Tabel 11: **Regressionsanalysens resultater for målere og kunder**

Variabel	Koefficient	Spredning	t-værdi	p-værdi
Log (konstant)	10,1123	0,4983	20,292	< 0,0001 ***
Log (målere)	0,3942	0,05295	7,443	< 0,0001 ***

Antal observationer: 72

Justeret $R^2 = 0,4338$

Omkostningsækvivalenten for kunder

Omkostningsækvivalenten for målere og kunder kan beskrives ud fra ligning (30)

$$Y = 32.936,2474 X_1^{0,3942} \quad (30)$$

Konstantleddet i (30) er beregnet ved $\exp(10,1123)^6$, og derefter er korrektionsfaktoren ganget på. Beregningen gennemgås i næste afsnit. Den forholdsvis lave koefficient for målere viser, at der er betydelige stordriftsfordele forbundet med omkostninger til målere og kunder. Når antallet af målere øges med 1 pct. stiger omkostningerne med 0,3942 pct.

Korrektionsfaktor

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

Standardafvigelsen for fejldedene er $\sigma = 0,7617$. Dermed kan det udledes at korrektionsfaktoren er 1,3366, hvilket skal ganges på højre side af ligningen.

Den korrigerede ligning for omkostningsækvivalenten fremgår dermed af (31).

$$Y = 1,3366 \cdot 24.641,8131 X_1^{0,3942} \quad (31)$$

Kvalitetskontrol af omkostningsækvivalenten

I Tabel 11 er det angivet, at koefficienten for målere er signifikant på et 5 pct. niveau.

Efter kvalitetskontrol af data har det været nødvendigt at fjerne seks selskaber, da de ikke har kunnet redegøre for afvigelser i indberetningen af målere sammenholdt med indberetningen

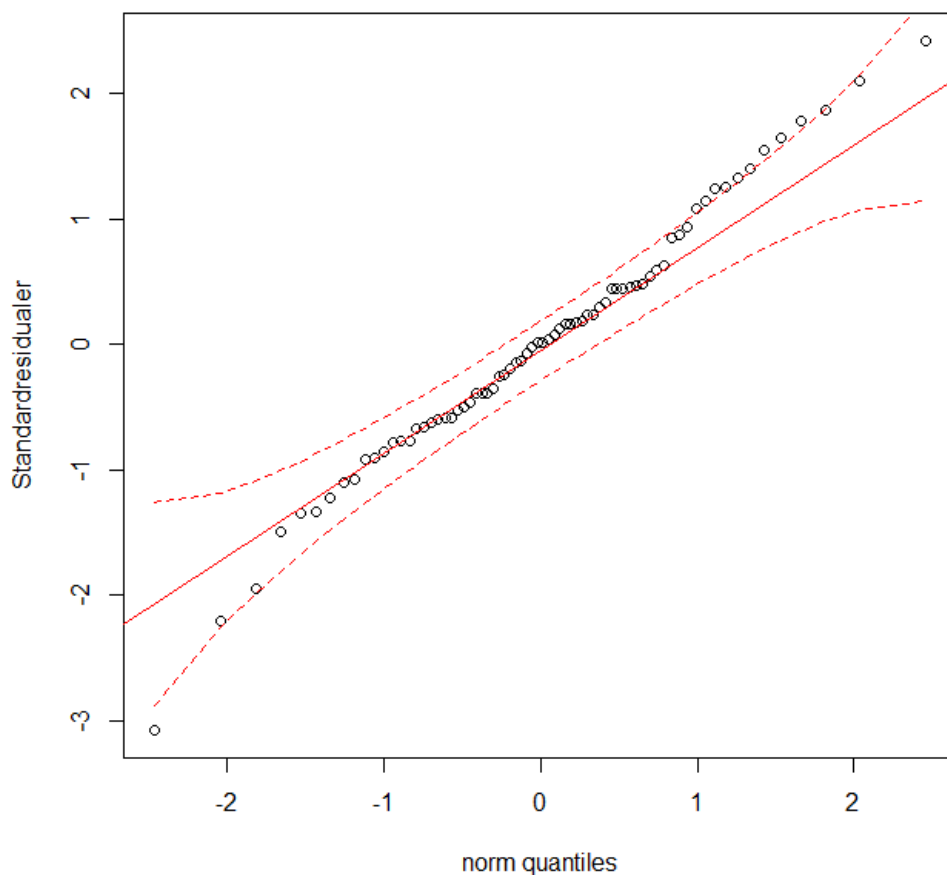
⁶ Exp betegner eksponentialfunktionen.

til de økonomiske rammer for 2017. Yderligere er der fjernet fire selskaber, som ikke har indberettet nogen omkostninger eller ingen costdrivere til målere og kunder.

Ved brug af Cook's distance er fjernet tre selskaber, som har haft uhensigtsmæssig stor indflydelse på resultatet af koefficienterne.

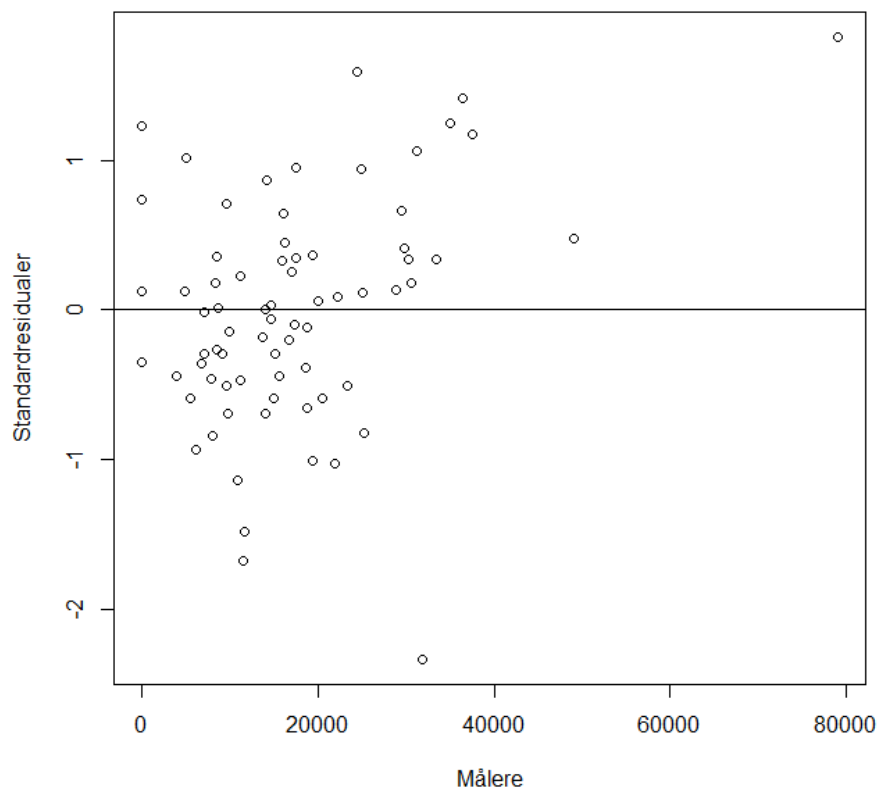
Figur 3.24 viser et QQ-plot af modellen. Det ses, at fejlleddene er pænt normalfordelte. Yderligere ses det, at der ikke er tegn på, at data skal transformeres yderligere.

Figur 3.164 QQ-plot for målere og kunder



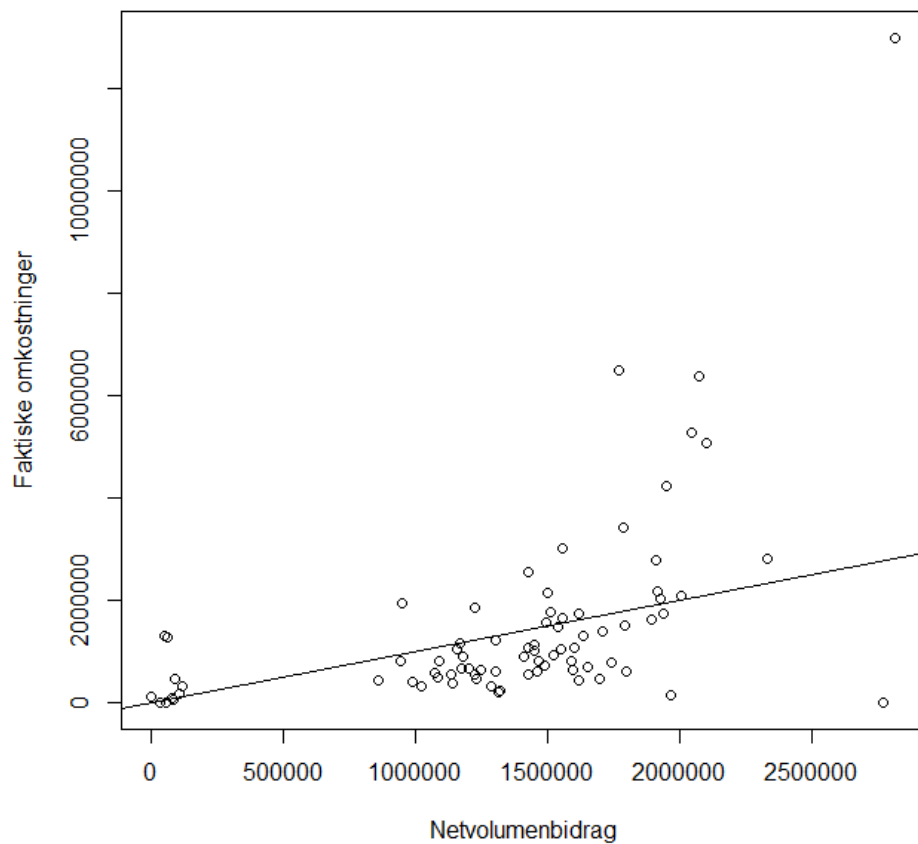
Figur 3.25 viser et plot af fejlleddene mod målere. Plottet viser, at fejlleddene ligger pænt fordelt omkring 0. Yderligere viser en Breusch-Pagan test, at der ikke er problemer med heteroskedasticitet.

Figur 3.1725 Residualplot



Figur 3.26 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 3.186 Faktiske omkostninger mod beregnede



3.10 Generel administration

Modellens baggrund

Driftsomkostninger til generel administration forventes at være påvirket af selskabets debiterede vandmængde. Den debiterede vandmængde er valgt som det eneste underliggende forhold til at bestemme omkostningerne til general administration. Udover de omkostninger selskaberne har indberettet til general administration, indgår også de omkostninger som er indberettet til revisorerklæringer samt til brancheforeninger. Der bliver opdelt efter selskabstype, hvor der udarbejdes en selvstændig regression til hver type. Altså laves der tre forskellige regressioner for generel administration. Én for selskaber, som kun varetager transport af spildevand, én regression for selskaber, som kun varetager rensning og til sidst én regression for selskaber, som varetager både transport og rensning.

Der er foretaget forskellige transformationer af data for at undersøge, hvilken model som bedst beskriver sammenhængen mellem omkostninger til general administration og den debiterede vandmængde. Det er fundet, at den bedste model for alle tre regressioner er en Cobb-Douglas funktion som beskrevet i (32).

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

$$Y = \beta_0 \cdot X_1^{\beta_1} \quad (32)$$

De estimerede værdier for modellerne er præsenteret i tabellerne 12a, 12b og 12c. Tabel 12a viser resultaterne for selskaber, som varetager både rensning og transport, tabel 12b viser resultaterne for selskaber, som kun varetager transport, og i tabel 12c fremgår resultaterne for selskaber, som kun varetager rensning.

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 (32).

Tabel 12a Resultater for regressionsanalysen af Generel administration for selskaber som varetager både rensning og transport af spildevand

Variabel	Koefficienter	Spredning	t-værdi	p-værdi
Log(konstant)	0,4915	2,6842	0,183	0,855
Log(debiteret vandmængde)	0,9851	0,1832	5,378	<0,0001***

Standardafvigelse for fejled = 0,8329

Korrektionsfaktor = 1,4147

Antal observationer: 54

Justeret R²: 0,3451

Tabel 12b Resultater for regressionsanalysen af Generel administration for selskaber som kun varetager transport af spildevand

Variabel	Koefficienter	Spredning	t-værdi	p-værdi
Log(konstant)	-0,6884	5,0759	-0,136	0,89
Log(debiteret vandmængde)	1,0597	0,3488	3,039	0,0078**

Standardafvigelse for fejled = 0,5341

Korrektionsfaktor = 1,1533

Antal observationer: 18

Justeret R²: 0,3263

Tabel 12c Resultater for regressionsanalysen af Generel administration for selskaber som kun varetager rensning af spildevand

Variabel	Koefficienter	Spredning	t-værdi	p-værdi
Log(konstant)	4,5502	3,1897	1,427	0,1875
Log(debiteret vandmængde)	0,6629	0,2132	3,109	0,0125*

Standardafvigelse for fejled = 0,7581

Korrektionsfaktor = 1,3329

Antal observationer: 11

Justeret R²: 0,4642

Omkostningsækvivalenten for generel administration

Omkostningsækvivalenten for generel administration for de selskaber som varetager både rensning og transport af spildevand kan beskrives ved ligning (33):

$$Y = 1,4147 \cdot 1,6347 \cdot X_1^{0,9851} \quad (33)$$

Omkostningsækvivalenten for generel administration for de selskaber som kun varetager transport af spildevand kan beskrives ved ligning (34):

$$Y = 1,1533 \cdot 0,5024 \cdot X_1^{1,0597} \quad (34)$$

Omkostningsækvivalenten for generel administration for de selskaber som kun varetager rensning af spildevand kan beskrives ved ligning (35):

$$Y = 1,3329 \cdot 94,6535 \cdot X_1^{0,6629} \quad (35)$$

Kvalitetskontrol af omkostningsækvivalenten

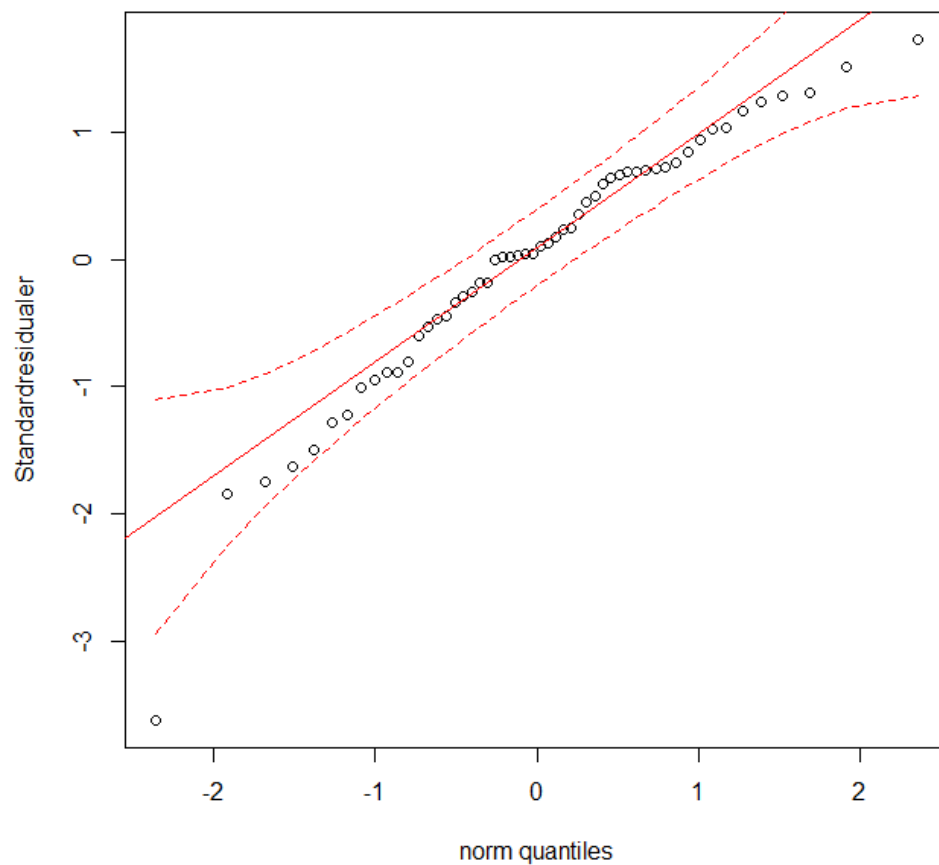
Af tabel 12a fremgår det, at koefficienten for debiteret vandmængde er statistisk signifikant på et 5 pct.-niveau. Før regressionen er fjernet et enkelt selskab, da der ikke er blevet redegjort for uregelmæssigheder i indberetningen.

Af tabel 12b fremgår det, at koefficienten for debiteret vandmængde er statistisk signifikant på et 5 pct.-niveau. Ved outlieranalyse er der fjernet et enkelt selskab, som ved hjælp af Cook's distance er vurderet til at have uhensigtsmæssig stor indflydelse på modellen.

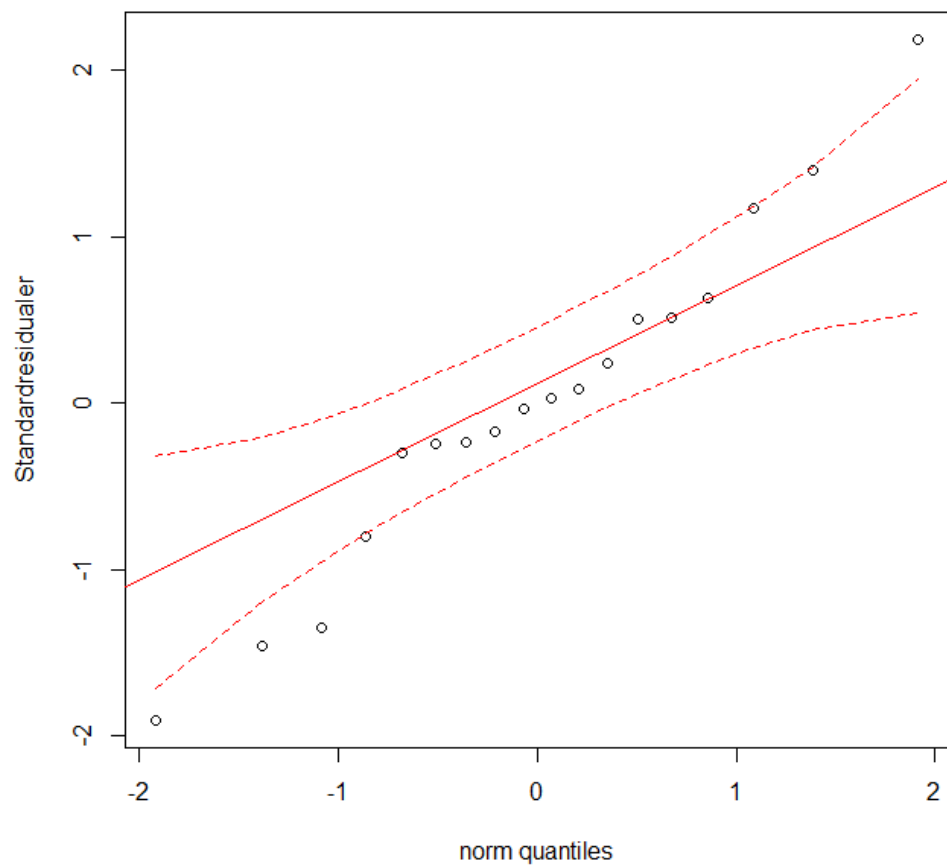
Af tabel 12c fremgår det, at koefficienten for debiteret vandmængde er statistisk signifikant på et 5 pct.-niveau.

Som nævnt tidligere er det undersøgt, om transformationer af data kan skabe en bedre model. QQ-plottes for de tre regressioner fremgår af figurerne 3.27a, 3.27b og 3.27c. Figurerne viser ikke tegn på, at transformationer er nødvendige, og observationerne ligger pænt indenfor konfidensintervallet. Den eneste undtagelse er ydreobservationerne for transportselskaberne, som ikke ligger indenfor konfidensintervallet. Ingen af de afprøvede alternative transformationer skaber en bedre model, hvorfor afvigelsen er accepteret.

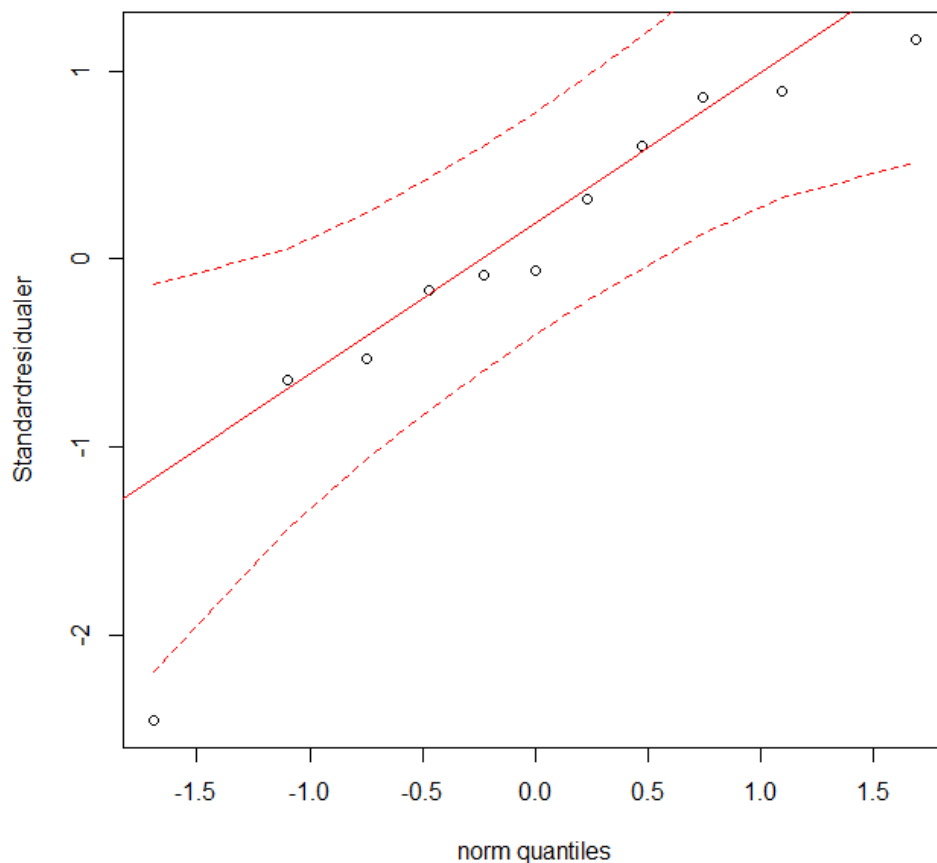
Figur 3.27a QQ-plot for generel administration for selskaber som varetager både rensning og transport af spildevand



Figur 3.27b **QQ-plot for general administration for selskaber som kun varetager transport af spildevand**

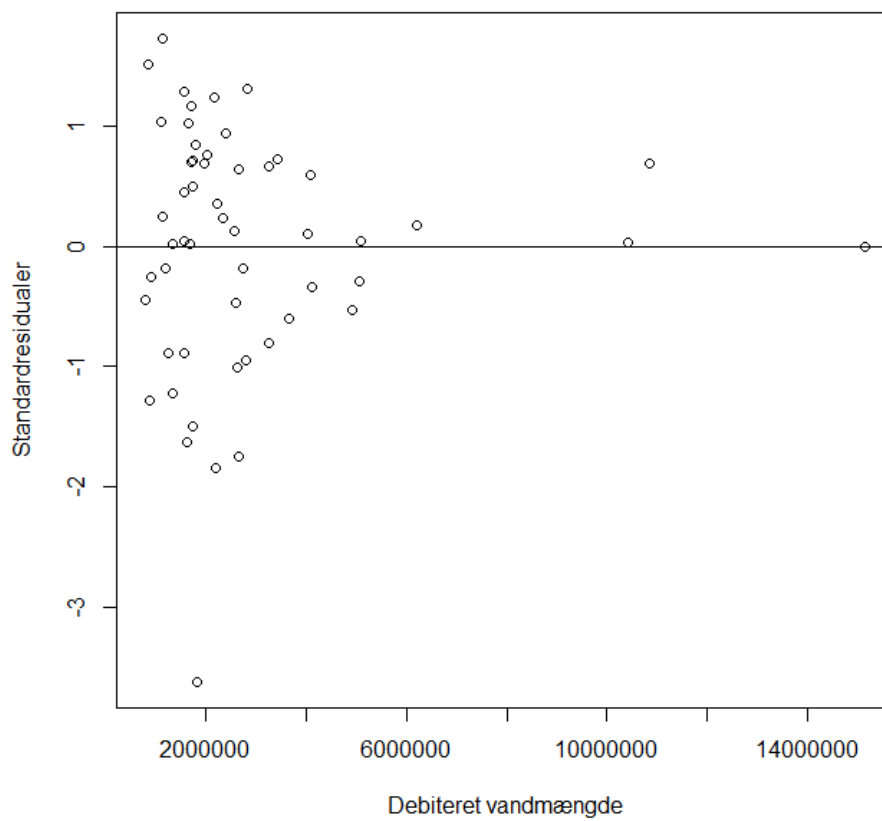


Figur3.27c **QQ-plot for generel administration for selskaber som kun varetager rensning af spildevand**

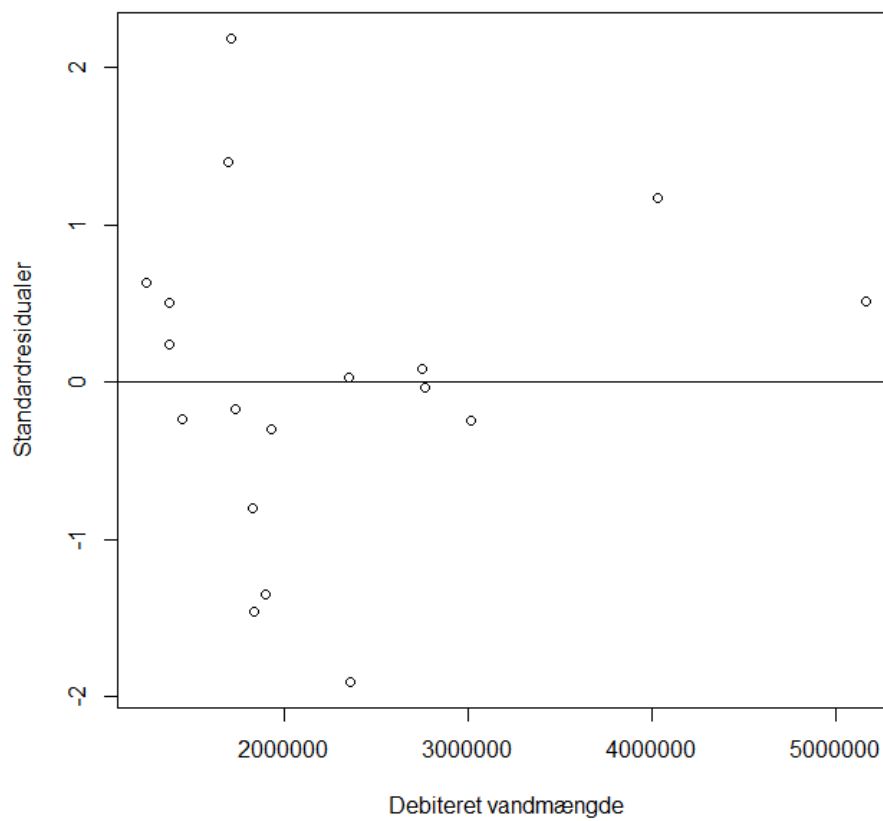


Figurerne 3.28a, 3.28b og 3.28c viser plots af fejleddene fra modellerne mod de debiterede vandmængder. Det ses, at fejleddene ligger pænt fordelt omkring 0. Der ser ud til at være enkelte store selskaber med meget små fejledd samt enkle selskaber med større fejledd, men dog inden for hvad der er acceptabelt. En Breusch-Pagan test viser, at der ikke er problemer med heteroskedasticitet.

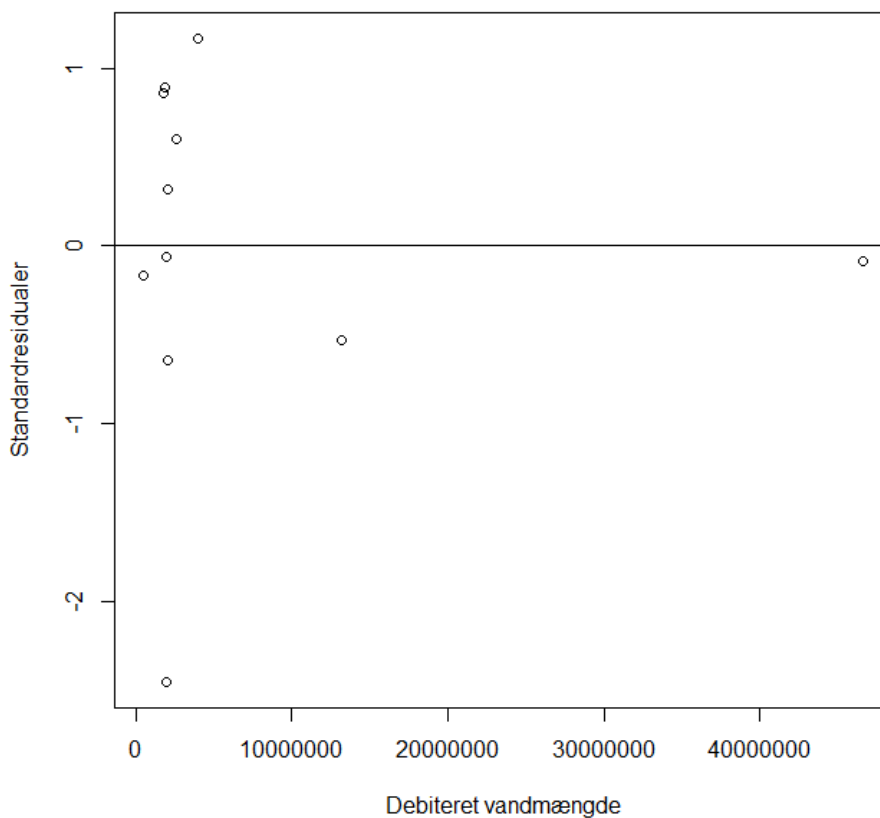
Figur 3.28a **Residualplot for selskaber som varetager både rensning og transport af spildevand**



Figur 3.28b Residualplot for selskaber som kun varetager transport af spildevand

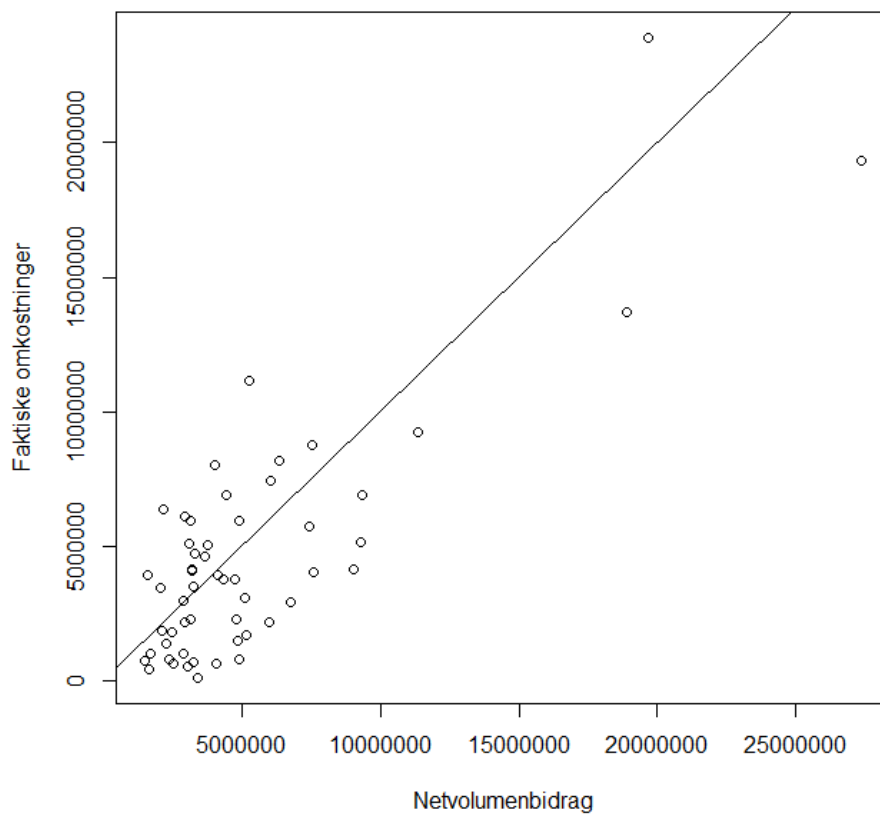


Figur3.28c Residualplot for selskaber som kun varetager rensning af spildevand

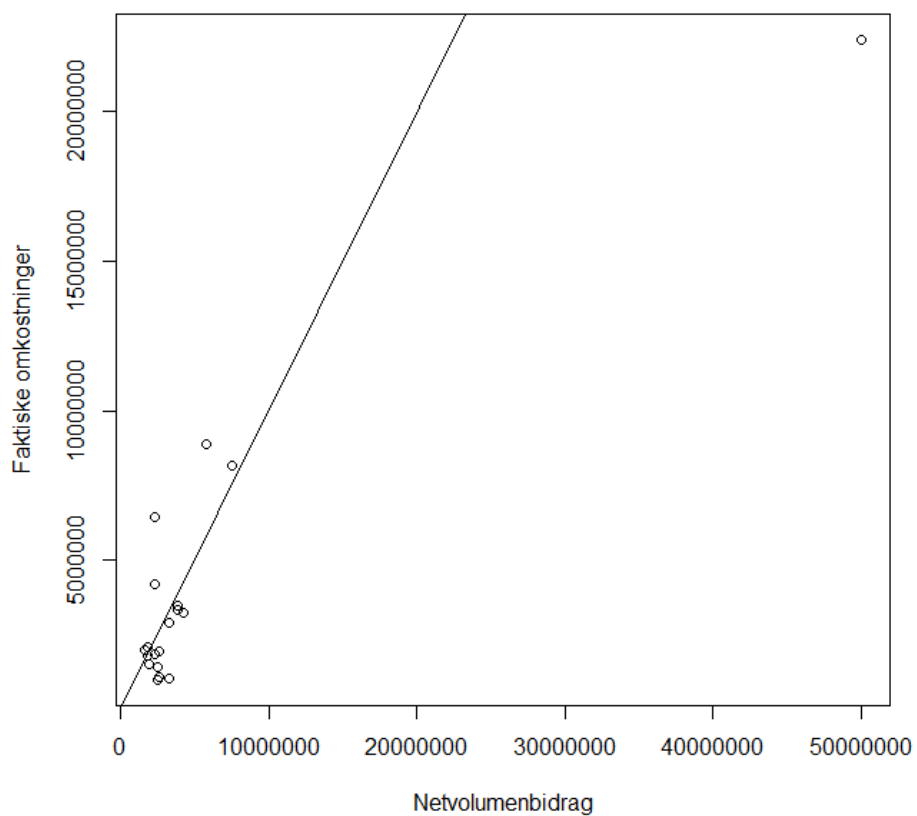


Til sidst viser figurene 3.29a, 3.29b og 3.29c plots af de beregnede værdier mod de faktisk indberettede omkostninger. Den markerede linje er et udtryk for det punkt, hvor de beregnede værdier er lig de indberettede. Generelt ligger selskaberne pænt fordelt omkring linjen. I Figur 3.29b skiller et enkelt selskab sig ud ved at have en beregnet værdi, som er meget højere end den indberettede. Dette selskab er dog fjernet ved outlieranalyse, hvorfor det er forventeligt, at der er en betydelig forskel.

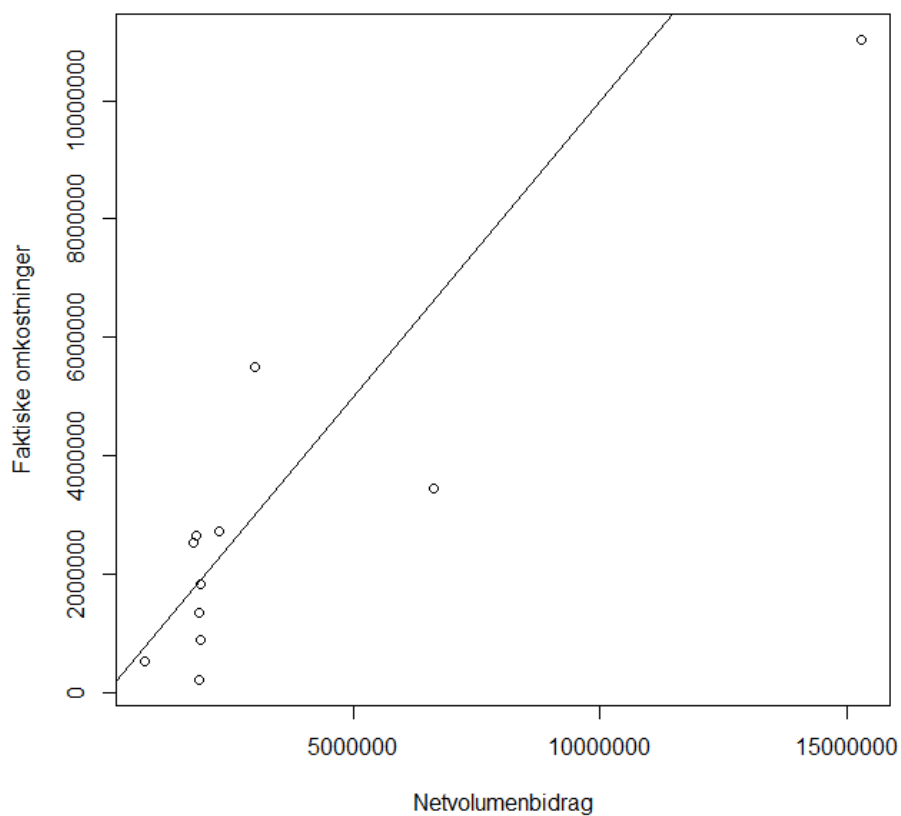
Figur 3.29a Faktiske omkostninger mod beregnede for selskaber som varetager både rensning og transport af spildevand



Figur 3.29b Faktiske omkostninger mod beregnede for selskaber som kun varetager transport af spildevand



Figur 3.29c Faktiske omkostninger mod beregnede for selskaber som kun varetager rensning af spildevand



Kapitel 4

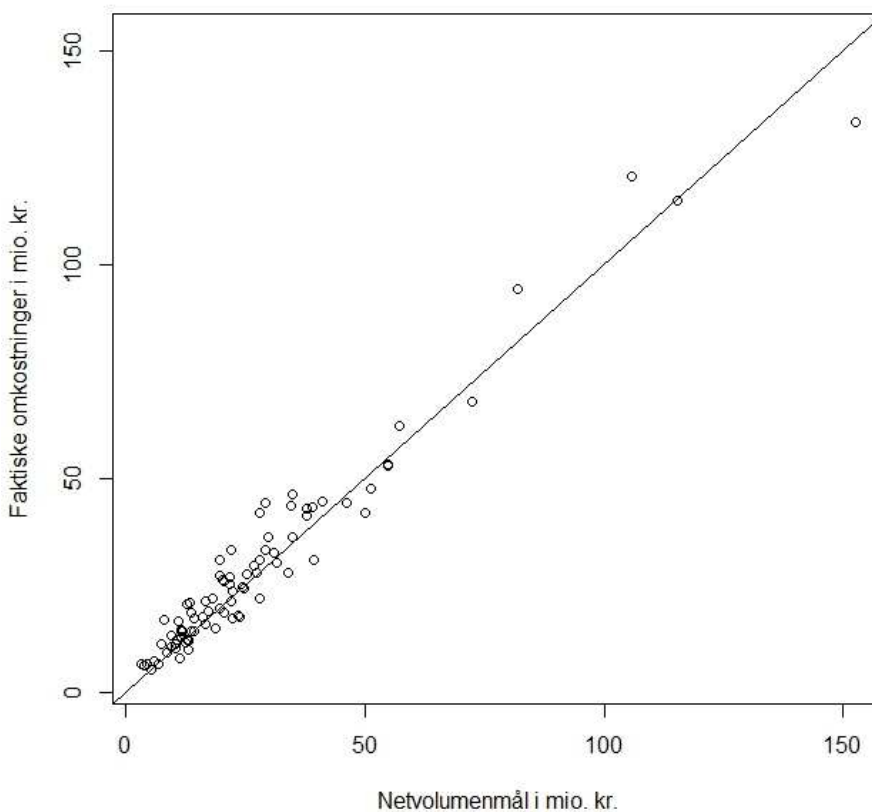
Samlet netvolumenmål

4.1 Indledning

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

Vi har i de tidligere kapitler gennemgået de enkelte costdrivere, hvor vi blandt andet har vist, hvor godt modellerne forudsiger selskabernes faktiske omkostninger. Dette kapitel viser, hvor god den samlede model er til at forudsige selskabernes samlede faktiske driftsomkostninger. Resultaterne fra den reviderede omkostningsækvivalent for renseanlæg er ikke inkluderet i de følgende to figurer. I Figur 4.1 ses et plot over selskabernes samlede omkostninger holdt op i mod deres samlede netvolumenmål.

Figur 4.1 Faktiske driftsomkostninger mod netvolumenmål



Figur 4.1 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. Yderligere ses det i Figur 4.2, at selskabernes faktiske driftsomkostninger og samlede netvolumenmål ligger tæt på hinanden.

Figur 4.2 Sammenligning af netvolumenmål og faktisk driftsomkostninger

