

Robuustheid regressiemodel voor kapitaalkosten gebaseerd op aansluitdichtheid

Dr.ir. P.W. Heijnen
Faculteit Techniek, Bestuur en Management
Technische Universiteit Delft

6 mei 2010

1 Introductie

De Energiekamer heeft in 2009 onderzoek laten doen door Cosentec naar de relatie tussen kabel- en lijngerelateerde *kapitaalkosten* en de *aansluitdichtheid* in het Nederlandse elektriciteits- en gasnetwerk. Dit onderzoek is beschreven in [Ontwerp-methodebesluit, 2010]. Uit dit onderzoek is voor het elektriciteitsnetwerk een verband geschat tussen deze *kapitaalkosten* en de *aansluitdichtheid* binnen de gebieden van de 9 Nederlandse netbeheerders, m.b.v. gegevens over het jaar 2006. Met het gevonden verband wil de Energiekamer voortaan per netbeheerder een schatting maken van de verwachte *kapitaalkosten* per jaar. Het verschil tussen de geschatte *kapitaalkosten* en de werkelijke *kapitaalkosten* zal daarna worden verrekend met de individuele netbeheerders. De Energiekamer heeft besloten om in 2010 [p272, Ontwerp-methodebesluit, 2010] het regressiemodel opnieuw te bepalen m.b.v. de meest recente data uit 2009 en het resulterende model voor 3 jaar vast te leggen en te gebruiken.

De methode die gebruikt is om de relatie tussen de *kapitaalkosten* en de *aansluitdichtheid* te bepalen, lijkt vrij arbitrair.

Het regressiemodel uit 2006 is slechts gebaseerd op 9 meetpunten en het nieuwe regressiemodel in 2009 zal zelfs maar gebaseerd zijn op 8 meetpunten. Ieder meetpunt komt overeen met één van de Nederlandse netbeheerders. Voor een betrouwbare regressieanalyse is dit een zeer klein aantal. De invloed van individuele meetpunten kan daardoor zeer groot zijn.

Stedin is van oordeel dat de betrouwbaarheid van de regressieanalyse en het daaruit voortvloeiende model terdege moet worden onderzocht voordat kan worden geconcludeerd dat *aansluitdichtheid* een objectief regionaal verschil vormt. In dit rapport is in dit verband nader onderzoek uitgevoerd naar verschillende voorwaarden waaraan moet worden voldaan voordat een regressiemodel gebruikt kan worden om betrouwbare schattingen van de werkelijkheid te geven.

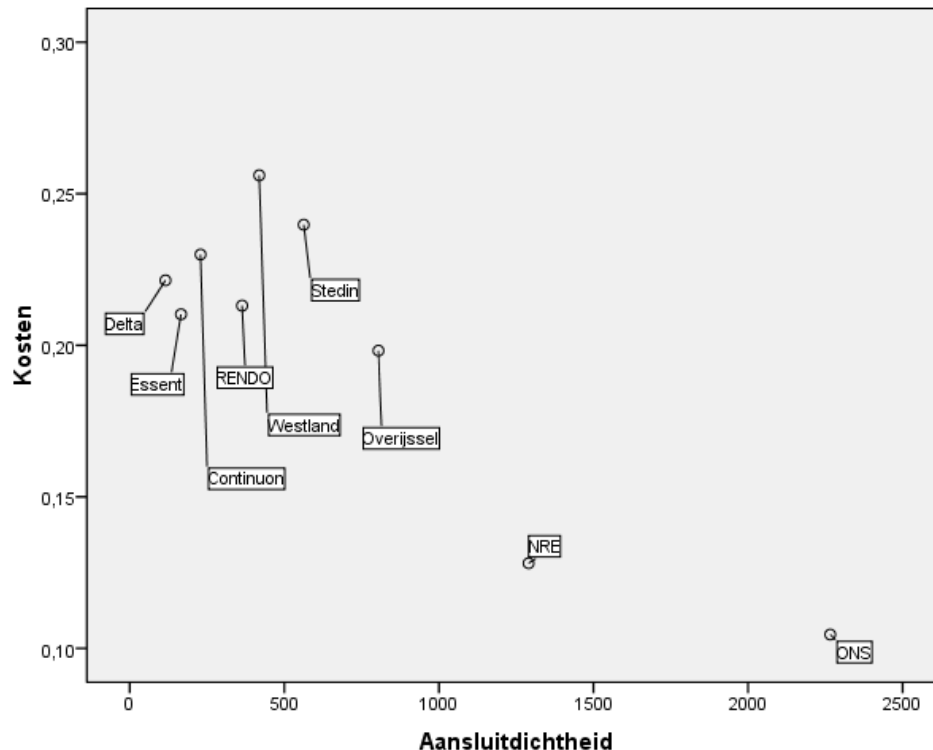
2 Regressiemodel voor kapitaalkosten op basis van aansluitdichtheid

Een simpel regressiemodel, zoals door de Energiekamer voorgesteld, geeft een model van het lineaire verband tussen een afhankelijke en een onafhankelijke variabele. Het regressiemodel kan voor twee doeleinden gebruikt worden. Ten eerste om m.b.v. het regressiemodel een schatting te maken van de verwachte waarde van de afhankelijke variabele en ten tweede om de bekende waarde van een afhankelijke variabele te verklaren d.m.v. de waarde van een onafhankelijke variabele.

Het regressiemodel dat de Energiekamer wil gaan gebruiken geeft het verband weer tussen de afhankelijke variabele *kapitaalkosten* en de onafhankelijke variabele *aansluitdichtheid*. Op basis van de *aansluitdichtheid*, zoals deze is bepaald voor iedere netbeheerder afzonderlijk, wil men m.b.v. het regressiemodel een schatting geven voor de *kapitaalkosten* van de netbeheerder. De werkelijke *kapitaalkosten* van de netbeheerder zullen aan deze schatting worden gerelateerd en afwijkingen hiervan zullen in positieve of negatieve zin worden verrekend. Om ervoor te zorgen dat het totaal aan geschatte waarden gelijk is aan het totaal van de daadwerkelijke *kapitaalkosten*, worden de te verrekenen bedragen geschaald met de quotiënt van de totale werkelijke kosten en de totale geschatte kosten.

Voor de netbeheerders, en in het bijzonder voor die netbeheerders waarvoor de schatting lager uitvalt dan hun werkelijke *kapitaalkosten*, is het dus van groot belang te weten dat er een duidelijk

verband bestaat tussen de *aansluitdichtheid* en de *kapitaalkosten*, dat de schatting gemaakt met het regressiemodel betrouwbaar is en dat het daadwerkelijk een goede indicatie geeft van de *kapitaalkosten* die bij de voor hen specifieke *aansluitdichtheid* haalbaar zouden moeten zijn. Dit rapport zal zich dan ook voornamelijk richten op deze aspecten, allereerst door het regressiemodel over 2006, zoals beschreven in [Ontwerp-methodebesluit, 2010] nader te onderzoeken en daarna door dezelfde analyses toe te passen op de nieuwe meetdata uit het jaar 2009. Ook zal de houdbaarheid van het regressiemodel over de jaren heen nader worden bekeken, omdat de Energiekamer het nieuwe regressiemodel, gebaseerd op de 2009 data, voor de komende 3 jaar vast wil leggen.



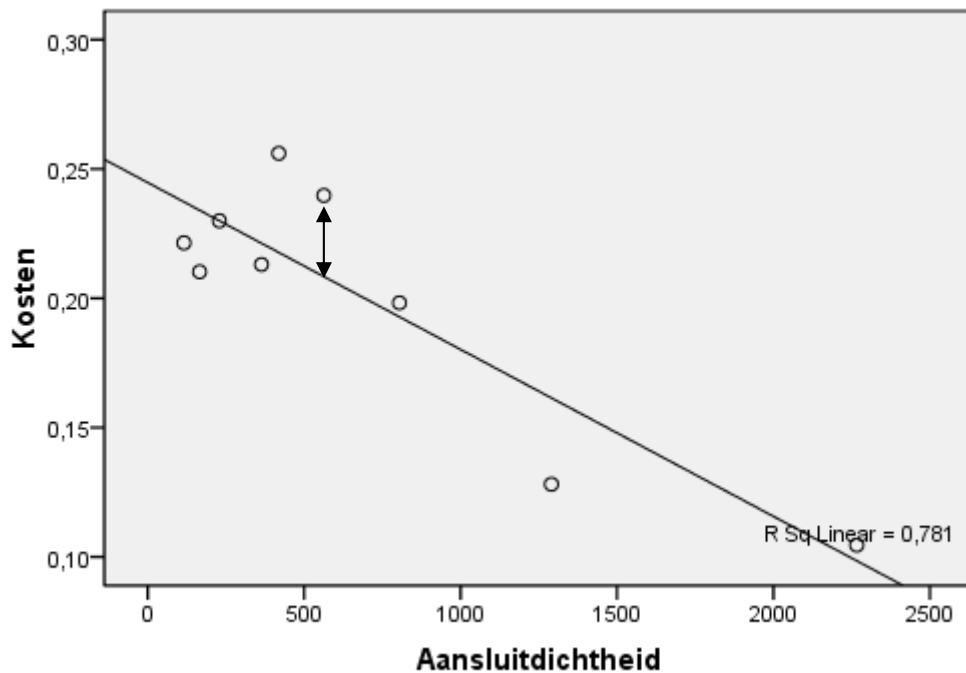
Figuur 1: Spreidingsdiagram van de kapitaalkosten versus de aansluitdichtheid (2006)

Het onderzoek is gedaan op data uit het jaar 2006 [p258, Ontwerp-methodebesluit, 2010]. Figuur 1 toont de gebruikte meetpunten. In 2006 waren er 9 netbeheerders. Van alle 9 netbeheerders zijn de data bekend verondersteld.

De correlatiecoëfficiënt van de *kapitaalkosten* en de *aansluitdichtheid* bedraagt $-0,884$. Dit duidt op een sterke negatieve lineaire relatie tussen de twee variabelen. Hiervan is dan ook gebruik gemaakt voor het schatten van het regressiemodel. Een belangrijke kanttekening hierbij is echter dat de correlatiecoëfficiënt in sterke mate beïnvloed kan zijn door individuele meetpunten die afwijken van de rest van de meetdata. Hier wordt later op teruggekomen.

Het regressiemodel is bepaald met een *ordinary least squares* methode [p258, Ontwerp-methodebesluit, 2010]. Dit is de standaardmethode voor het bepalen van een regressiemodel. Iedere netbeheerder heeft hetzelfde gewicht gekregen [p266, Ontwerp-methodebesluit, 2010] en er wordt

een lineair verband tussen de *kapitaalkosten* en de *aansluitdichtheid* bepaald zodanig dat het verschil tussen de meetpunten en de regressielijn zo klein mogelijk is.



Figuur 2: Spreidingsdiagram en regressielijn (2006)

De Energiekamer gebruikt het Schwartz Bayesian criterium om een keuze te maken tussen een lineair, kwadratisch of loglineair verband [p260, Ontwerp-methodebesluit, 2010]. Op basis van dit criterium is gekozen voor een lineair verband. Figuur 2 toont de meetpunten en de lineaire regressielijn zoals hij bepaald is door Energiekamer voor de data over 2006 [Formule 35, bijlage 1, Ontwerp-methodebesluit, 2010].

Voor de accuraatheid van een regressiemodel wordt over het algemeen de determinatiecoëfficiënt, ook wel de R^2 gebruikt. Dit is een maat die aangeeft welk percentage van de variantie van de afhankelijke variabele wordt verklaard door de onafhankelijke variabele. Voor een simpel regressiemodel met slechts één onafhankelijke variabele is de R^2 gelijk aan het kwadraat van de correlatiecoëfficiënt. In dit geval, verklaart het regressiemodel op basis van de *aansluitdichtheid* 78,1% van de variantie gevonden in de *kapitaalkosten*.

De Energiekamer concludeert hieruit, dat de netbeheerders met een meetpunt boven de regressielijn te hoge *kapitaalkosten* hebben gegeven hun *aansluitdichtheid* [p224, Ontwerp-methodebesluit, 2010].

Voordat dit echter geconcludeerd mag worden, moet er eerst gedegen worden aangetoond dat de *aansluitdichtheid* inderdaad een goede voorspeller is voor de *kapitaalkosten* die verwacht kunnen worden.

In het onderzoek van Cosentec lijkt met onderstaande zaken echter onvoldoende tot geen rekening gehouden te zijn:

- De dataset is compleet. Dat betekent dat er geen sprake is van een steekproeffout of van onzekerheid in de verkregen relaties. De significantiewaarde die volgt uit de statistische analyses kan niet gebruikt worden om aan te tonen dat de resultaten praktische relevantie hebben. Hiervoor is een ander criterium nodig.
- De dataset is erg klein, waardoor de invloed van individuele netbeheerders onaanvaardbaar groot kan worden en het verdwijnen of ontstaan van een nieuwe netbeheerder het regressiemodel structureel kan veranderen.
- De dataset is erg klein, waardoor niet is aan te tonen dat aan een aantal belangrijke regressievoorwaarden is voldaan. De gebruikte regressiemethode is echter niet geldig als niet aan deze voorwaarden is voldaan. In dat geval kan het regressiemodel niet worden gebruikt om betrouwbare schattingen te maken.
- De keuze om alle netbeheerders hetzelfde gewicht te geven in de regressieanalyse kan erg bepalend zijn voor het uiteindelijke resultaat. Het zorgt ervoor dat een zeer kleine netbeheerder dezelfde invloed op het regressiemodel heeft als een hele grote netbeheerder. De keus voor een gewogen regressiemodel kan de uitkomsten structureel veranderen.

Bovengenoemde punten zullen hieronder één voor één besproken worden.

3 Onderzoek op de complete populatie van netbeheerders

In het onderzoek zijn de gegevens van alle 9 netbeheerders over 2006 bekend verondersteld. Er is hier dus geen sprake van een steekproef, maar van een onderzoek op de totale populatie. Dat betekent dat er geen sprake is van een steekproeffout of van een statistische significantie van bijvoorbeeld een correlatiecoëfficiënt of een regressiecoëfficiënt. Bij een steekproefonderzoek geeft de statistische significantie de kans aan om deze waarde (of groter) van de correlatiecoëfficiënt in de steekproef te vinden, terwijl hij in werkelijkheid voor de gehele populatie gelijk is aan 0. Als deze kans heel klein is, kleiner dan 5% in het algemeen, wordt ervan uitgegaan dat de correlatiecoëfficiënt significant is, wat betekent dat dezelfde *relatie* ook voor de gehele populatie zal gelden, maar dit betekent niet dat exact dezelfde *waarde* van de correlatiecoëfficiënt ook voor de populatie geldt.

In het geval van een onderzoek op de complete populatie kent de waarde van een correlatiecoëfficiënt of regressiecoëfficiënt m.b.t. de gebruikte data geen onzekerheid. Wel is er sprake van onzekerheid als het er over gaat of de resultaten kunnen worden toegepast om de *kapitaalkosten* voor komende jaren te schatten op basis van de *aansluitdichtheid*. In [p268, Ontwerp-methodebesluit, 2010] wordt alleen vermeld dat de Energiekamer er vanuit gaat dat het verband houdbaar is over de tijd, hiernaar is echter geen onderzoek gedaan.

In [Ontwerp-methodebesluit, 2010] wordt meerdere malen de statistische significantie van resultaten gebruikt, om de praktische relevantie van de resultaten aan te tonen [p258, p260, Ontwerp-methodebesluit, 2010]. Dit is niet terecht¹. Als het om het verband gaat tussen de *aansluitdichtheid* en de *kapitaalkosten* is alleen de hoogte van de correlatiecoëfficiënt en het daarmee samenhangend percentage verklaarde variantie (de R^2) van belang. De Energiekamer moet zelf een eenduidige ondergrens vastleggen voor de hoogte van de correlatiecoëfficiënt die zij nog relevant acht. Dit kan niet gebeuren op basis van de statistische significantie.

¹ In [p258, Ontwerp-methodebesluit, 2010] wordt zelfs gesproken over “als de t-waarde niet hoger is dan 5%”, hier wordt waarschijnlijk bedoeld, “als de significantie van de t-waarde niet hoger is dan 5%”.

4 Invloed van individuele netbeheerders op het model

Zoals in Figuur 1 en Figuur 2 duidelijk te zien is, wordt de regressielijn in sterke mate bepaald door de netbeheerder ONS en in iets mindere mate door de netbeheerder NRE. Zonder deze twee netbeheerders zou er nauwelijks sprake zijn van een lineair verband tussen de *kapitaalkosten* en de *aansluitdichtheid*. De correlatiecoëfficiënt bedraagt in dat geval nog slechts -0,137 en een regressiemodel zonder deze 2 netbeheerders zal nog slechts 1,9% van de variantie kunnen verklaren van de verschillen in *kapitaalkosten* zoals gevonden voor de verschillende netbeheerders. Hieraan kan geen enkele praktische relevantie meer worden toegekend.

Hieronder zal m.b.v. een aantal statistische parameters worden aangetoond dat de invloed van de netbeheerders ONS en NRE op de regressieresultaten inderdaad erg groot is.

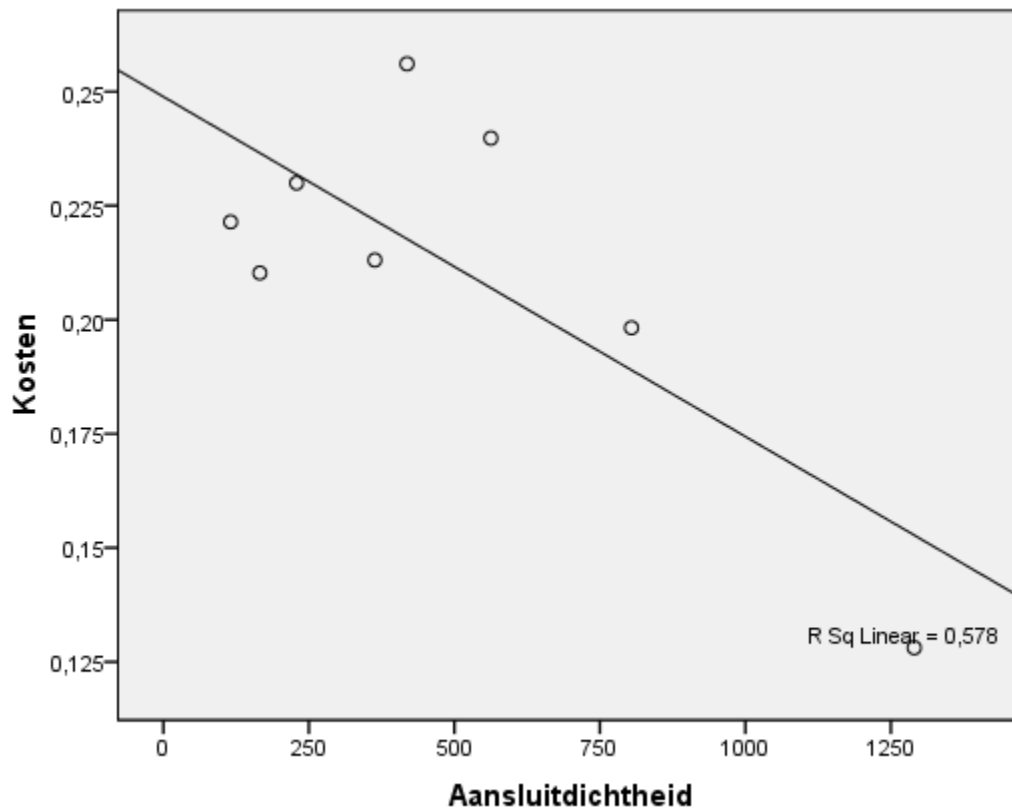
De invloed van de individuele meetpunten op een regressiemodel kan op een aantal manieren worden gemeten:

- **Grootte van het residu:** Dit is het verschil tussen gemeten waarde en geschatte waarde. Meetpunten met een groot residu vertonen afwijkend gedrag van de andere meetpunten en hun waarde wordt door het model niet goed verklaard. De residuen worden over het algemeen gestandaardiseerd, zodat ze een gemiddelde van 0 en een variantie van 1 hebben. Een vuistregel is dat residuen waarvan de absolute waarde groter is dan de 2,5% grens van de bijbehorende Student t-verdeling verwijderd zouden moeten worden uit de analyse [p223, Hair et al., 2006]. Met een dataset met 9 meetpunten ligt deze 2,5% grens op 2,3.
- **Cook's distance:** Dit is een maat voor de verandering van de regressiecoëfficiënt als het meetpunt uit de analyse wordt verwijderd. De Cook's distance moet laag zijn (dichtbij 0), anders wordt het regressiemodel in hoge mate bepaald door dat ene meetpunt, wat voor de betrouwbaarheid van het model niet gewenst is. Als vuistregel geldt dat meetpunten boven $\frac{4}{(n-k-1)}$ [p225, Hair et al., 2006] verwijderd moeten worden. Hierin is n de grootte van de dataset en k het aantal onafhankelijke variabelen in het regressiemodel. In dit geval zouden meetpunten met een Cook's distance groter dan 0,57 ($n = 9$ en $k = 1$) verwijderd moeten worden.
- **Leverage value:** Dit is een maat voor de afstand van het individuele meetpunt tot alle andere meetpunten. Ook de Leverage value moet laag zijn (dichtbij 0), anders vertoont het meetpunt teveel afwijkend gedrag van de andere meetpunten en kan daardoor het regressiemodel teveel beïnvloeden. Als vuistregel geldt dat meetpunten met een Leverage value boven de $\frac{3(k+1)}{n}$ [p224, Hair et al., 2006] teveel invloed hebben en verwijderd moeten worden. In dit geval zouden meetpunten met een Leverage value groter dan 0,67 verwijderd moeten worden.

Netbeheerder	Residul (<2,31)	Cook's distance (<0,57)	Leverage value (<0,67)
Continuon	0,0000	0,0000	0,0550
Overijssel	0,2130	0,0033	0,0033
Delta	-0,6234	0,0591	0,0854
Stedin	1,2345	0,1123	0,0042
Essent	-0,9361	0,1195	0,0711
NRE	-1,3135	0,2777	0,0928
ONS	0,2387	0,3492	0,6411
REDO	-0,3229	0,0098	0,0276
Westland	1,5097	0,1962	0,0191

Figuur 3: Analyse van invloedrijke punten (2006)

Figuur 3 toont aan dat de Netbeheerder ONS een erg grote Leverage value en de grootste Cook's distance heeft van alle netbeheerders in de dataset. Daarmee heeft deze netbeheerder dus veel invloed op de dataset en het zou beter zijn om deze te verwijderen uit de analyse.



Figuur 4: Spreidingsdiagram en regressielijn (zonder ONS) (2006)

Figuur 4 toont opnieuw het spreidingsdiagram en de regressielijn, zonder de netbeheerder ONS. Dit nieuwe regressiemodel verklaart nog 57,8% (R^2) van de variantie in de *kapitaalkosten*.

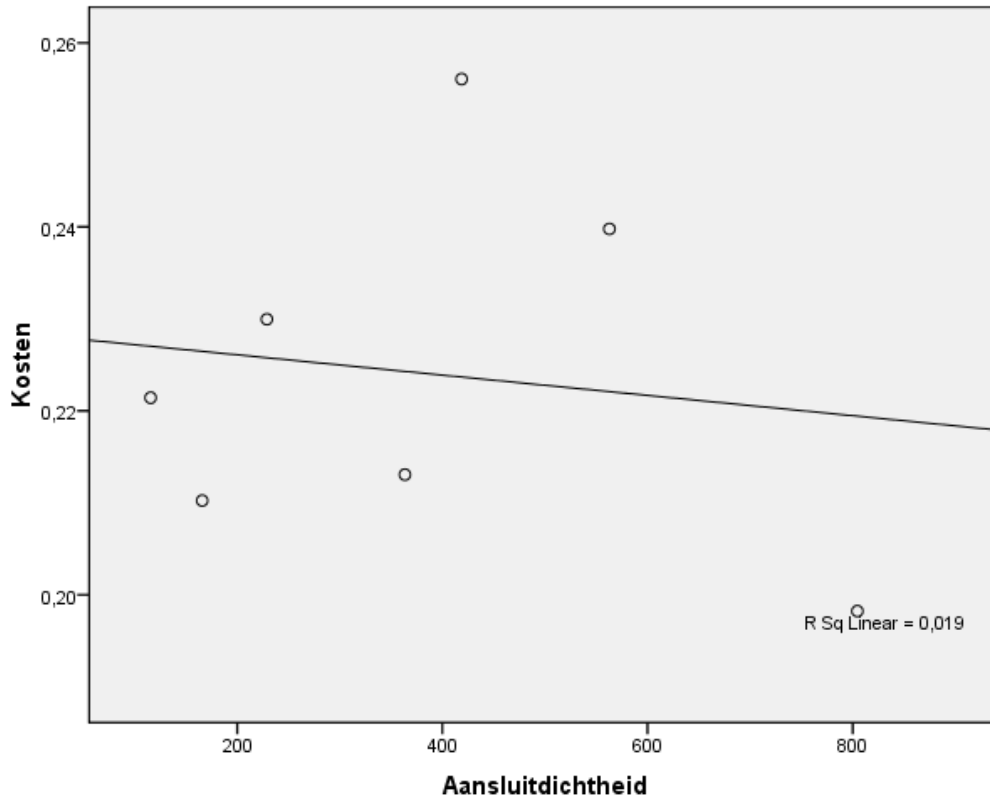
Opnieuw moet echter weer gecontroleerd worden wat voor dit nieuwe model de invloed is van de individuele meetpunten. De grenzen voor Cook's distance en de Leverage value zijn aangepast aan de nieuwe situatie met $n = 8$ meetpunten en $k = 1$.

Netbeheerder	Residul (<2,37)	Cook's distance (<0,67)	Leverage value (<0,75)
Continuon	-0,0697	0,0007	0,0650
Overijssel	0,3434	0,0204	0,0895
Delta	-0,6984	0,1140	0,1326
Stedin	1,2157	0,1262	0,0044
Essent	-0,9743	0,1774	0,0997
NRE	-0,9144	3,6165	0,5879
ONS			
RENDO	-0,3240	0,0100	0,0157
Westland	1,4217	0,1739	0,0052

Figuur 5: Analyse van invloedrijke punten (zonder ONS) (2006)

Uit Figuur 5 blijkt dat zonder netbeheerder ONS, de netbeheerder NRE een te hoge Cook's distance heeft en een in verhouding erg hoge Leverage value en daarmee onaanvaardbaar veel invloed heeft op het gevonden regressiemodel.

Zoals hierboven echter al is besproken, verklaart een regressiemodel zonder de netbeheerders ONS en NRE nog slechts 1,9% van de variantie in de *kapitaalkosten*.



Figuur 6: Spreidingsdiagram en regressielijn (zonder ONS en NRE) (2006)

Uit Figuur 6 blijkt duidelijk dat er zonder de netbeheerders ONS en NRE geen sprake meer is van een eenduidig lineair verband. Verder onderzoek naar de invloed van de individuele netbeheerders heeft dan ook geen zin.

Op basis van de analyses hierboven blijkt dus dat het gevonden regressiemodel niet robuust is omdat het in sterke mate wordt bepaald door 2 van de 9 netbeheerders. Deze 2 netbeheerders hebben samen slechts 2% van alle aansluitingen in hun beheer.

Het regressiemodel op deze manier gebruiken betekent in feite dat alle andere netbeheerders (die samen 98% van alle aansluitingen beheren) worden vergeleken met deze 2 netbeheerders en dat hun verrekening aan de hand van de geschatte kosten afhangt van de prestaties van deze 2 kleine netbeheerders.

5 Regressievoorwaarden waaraan moet zijn voldaan

Bij het gebruik van een ordinary least squares regressiemethode wordt verondersteld dat aan een aantal voorwaarden is voldaan. Zo niet, dan zijn de schattingen met het gevonden regressiemodel niet betrouwbaar.

De voorwaarden waaraan moet worden voldaan luiden [p172 en verder, Hair, 2006]:

1. De meetpunten zijn onafhankelijk.
2. De variabelen zijn van minimaal ordinaal meetniveau.
3. Er is sprake van een lineair verband tussen de afhankelijke en de onafhankelijke variabele.

4. Voor een vaste waarde van de onafhankelijke variabele moet de afhankelijke variabele normaal verdeeld zijn.
5. De afhankelijke variabele moet een constante variantie hebben voor de verschillende waarden van de onafhankelijke variabele.

Hoewel in het vorige hoofdstuk is aangetoond dat het regressiemodel over 2006 in sterke mate bepaald is door twee individuele meetpunten, wordt hier toch uitgegaan van het oorspronkelijke model. De voorwaarden zullen hieronder één voor één besproken worden.

Voorwaarde 1: De meetpunten zijn onafhankelijk

Aangezien het hier alle 9 onafhankelijke netbeheerders betreft, mag er van worden uitgegaan dat aan deze voorwaarde is voldaan.

Voorwaarde 2: De variabelen zijn van minimal ordinaal meetniveau

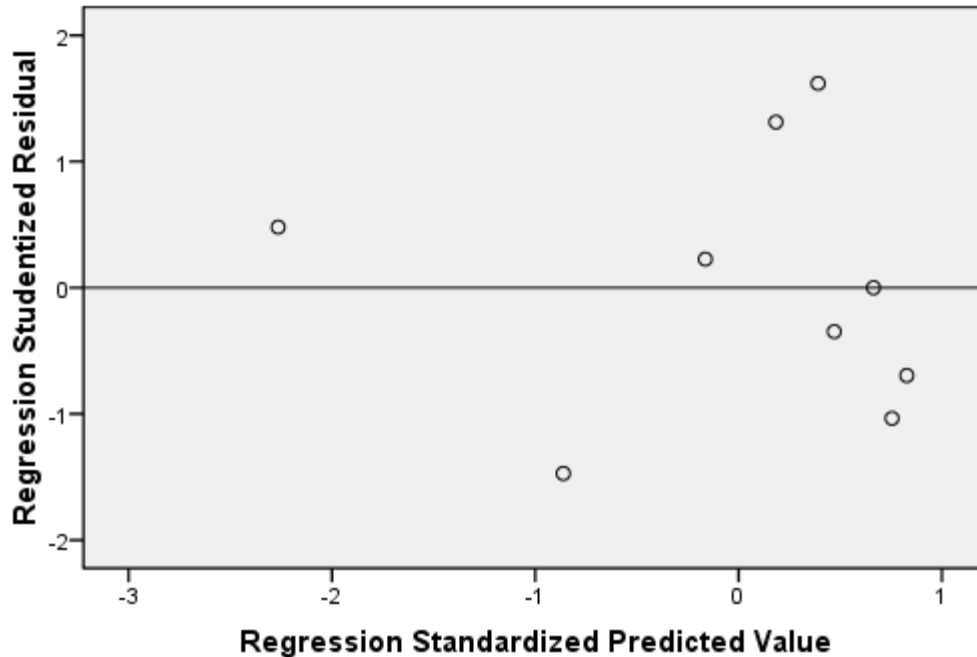
Dit betekent dat er minstens een ordening tussen de waarden van de variabelen aanwezig moet zijn. In dit geval is hier zeker aan voldaan, want beide variabelen *aansluitdichtheid* en *kapitaalkosten* zijn op een numerieke schaal gemeten.

Voorwaarde 3: Er moet een lineair verband aanwezig zijn

Aan deze voorwaarde wordt niet in voldoende mate voldaan. Het spreidingsdiagram en ook de analyses uit het vorige hoofdstuk tonen duidelijk aan dat het statistisch gevonden lineaire verband vooral wordt bepaald door de aanwezigheid van de netbeheerders ONS en NRE.

Voorwaarde 4 en 5: Normaalverdeling en constante variantie van de afhankelijke variabele voor vaste waarden van de onafhankelijke variabele

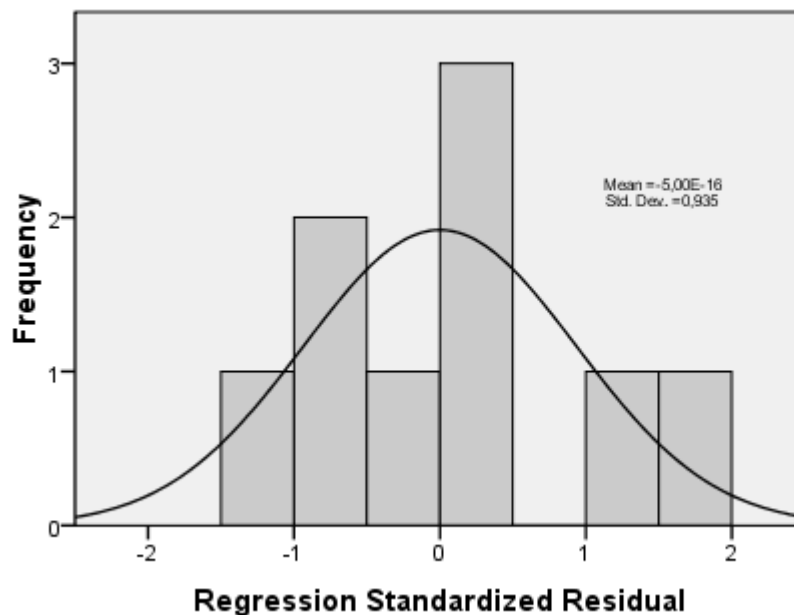
Voor grotere datasets wordt dit aangetoond door naar een aantal plots van de residuen te kijken. In de eerste plot, worden de residuen uitgezet tegen de geschatte waarden van de afhankelijke variabele. Als deze plot een mooie goed verdeeld puntenwolk laat zien, zijn er geen afwijkingen in de verdeling van de afhankelijke variabele te verwachten.



Figuur 7: Spreidingsdiagram van de residuen t.o.v. de geschatte waarden (2006)

Figuur 7 toont deze plot van de residuen. Het is duidelijk dat op basis van zo weinig meetpunten niets zinnigs gezegd kan worden over de verdeling van de residuen.

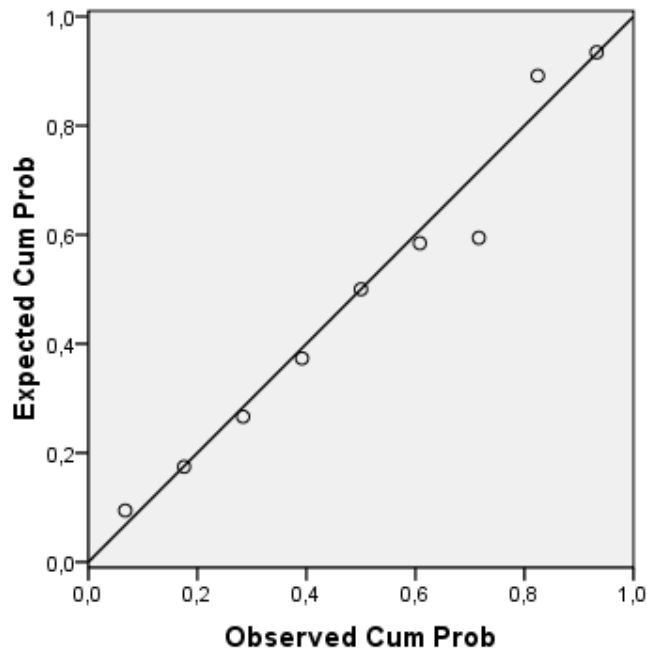
Een histogram van de residuen moet aantonen dat de residuen normaal verdeeld zijn.



Figuur 8: Histogram van de residuen (2006)

Figuur 8 toont dit histogram van de residuen en de bijbehorende (zelfde gemiddelde, zelfde variantie) normaalverdeling. Ook hierbij is duidelijk dat op basis van zo weinig meetpunten niets zinnigs valt af te leiden over de verdeling van de residuen. Om die reden is het onbegrijpelijk dat de

Energiekamer in [Formule 35, Bijlage 1, Ontwerp-methodebesluit, 2010] concludeert “Deze term kent een statistische normale verdeling”, waarbij “Deze term” verwijst naar de residuen. Dit wordt nergens bewezen en kan ook niet worden bewezen.



Figuur 9: Normalwaarschijnlijkheidsplot (2006)

Figuur 9 toont de normalwaarschijnlijkheidsplot van de residuen. Deze plot vergelijkt het gevonden percentage meetpunten kleiner dan een zekere waarde met het verwachte aantal meetpunten kleiner dan die waarde als de normaalverdeling was gevolgd. De normalwaarschijnlijkheidsplot toont een redelijke overeenkomst met de normaalverdeling op een paar punten na. Maar ook hier geldt weer dat de normaalverdeling voor zo’n kleine dataset niet is aan te tonen.

6 Gewogen versus ongewogen regressiemodel

In het regressiemodel, zoals gepresenteerd door de Energiekamer, is uitgegaan van een ongewogen regressiemodel [p266, Ontwerp-methodebesluit, 2010]. Zoals in Hoofdstuk 3 al is aangetoond, is het daardoor mogelijk dat het lineaire verband, waarop het regressiemodel is gebaseerd, wordt veroorzaakt door twee netbeheerders ONS en NRE, die samen slechts 2% van alle aansluitingen beheren.

De keuze om een ongewogen regressiemodel te schatten, is slechts één van de vele keuzes die gemaakt hadden kunnen worden, maar één die wel zeer bepalend is voor het uiteindelijke resultaat.

Een betere keuze zou zijn om de invloed van een netbeheerder uit te drukken in het aantal aansluitingen dat de netbeheerder beheert. In dat geval wordt een gewogen regressieanalyse uitgevoerd. Het gewogen regressiemodel luidt:

$$\text{kosten} = 0,228872 - 0,000016 * \text{aansluitdichtheid}$$

Dit regressiemodel verklaart echter nog maar 4,4% van de variantie in *kapitaalkosten* van de verschillende netbeheerders. De praktische waarde hiervan is nihil.

7 Relatie tussen kapitaalkosten en aansluitdichtheid over 2009

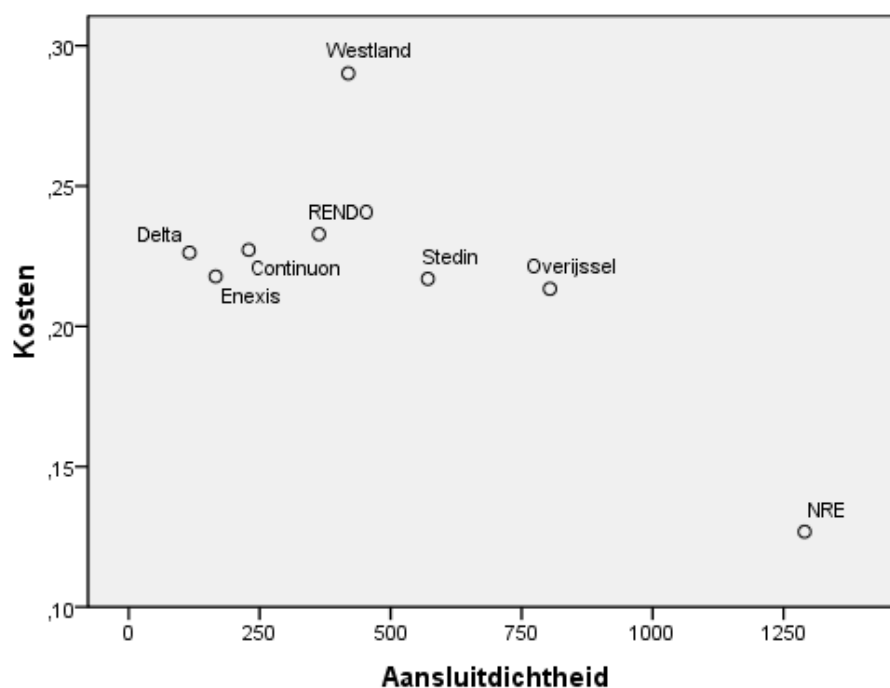
De Energiekamer heeft besloten om in 2010 het regressiemodel opnieuw te bepalen m.b.v. de meest recente data uit 2009 en het resulterende model voor 3 jaar vast te leggen en te gebruiken [p272, Ontwerp-methodebesluit, 2010]. In dit hoofdstuk zal worden onderzocht of de problemen zoals geïdentificeerd voor de data uit 2006 ook voor de data uit 2009 gelden.

Zoals al eerder opgemerkt zal het regressiemodel geschat over de data uit 2009 nog slechts op 8 meetpunten gebaseerd zijn. Dit komt omdat de netbeheerder ONS is overgenomen door de netbeheerder Stedin.

Bij de analyses is uitgegaan van de data zoals weergegeven in Figuur 10.

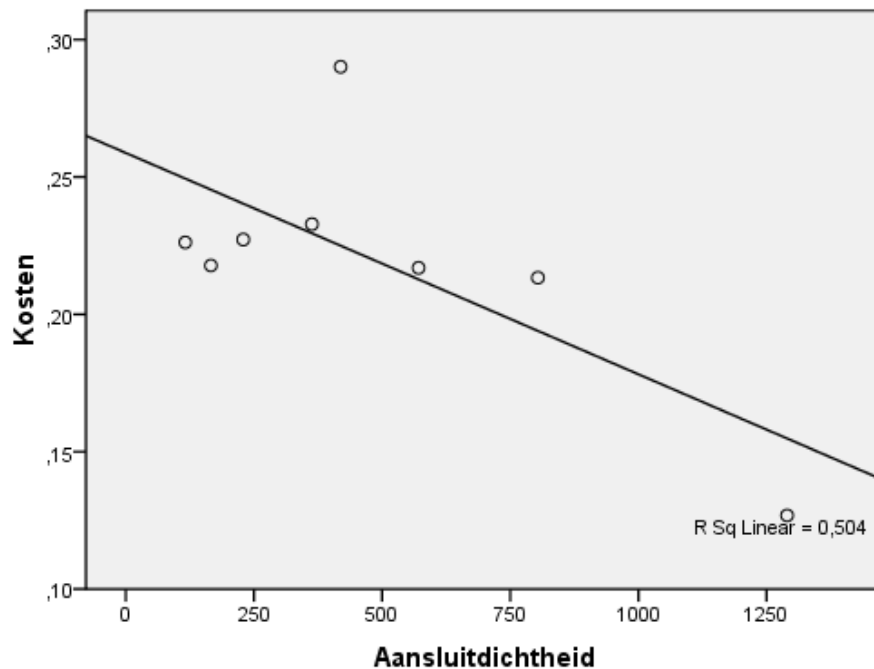
Netbeheerder	Aansluitingen	Kapitaalkosten	Aansluitdichtheid
Continuon	2822488	0,2272	229
Overijssel	52478	0,2133	804
Delta	205379	0,2262	116
Stedin	2068256	0,2169	571
Enexis	2581562	0,2178	166
NRE	105912	0,1268	1290
RENDO	31627	0,2328	363
Westland	53117	0,2901	419

Figuur 10: Meetdata over 2009



Figuur 11: Spreidingsdiagram van de *kapitaalkosten* versus de *aansluitdichtheid* (2009)

Figuur 11 toont de nieuwe meetpunten voor de *kapitaalkosten* en de *aansluitdichtheid* van de 8 netbeheerders over 2009. De correlatiecoëfficiënt van de *kapitaalkosten* en de *aansluitdichtheid* bedraagt nu -0,710.



Figuur 12: Spreidingsdiagram en regressielijn

De 2009 data geven geen aanleiding om op zoek te gaan naar een ander dan lineair verband, zoals ook voor de 2006 data is besloten door de Energiekamer [p260, Ontwerp-methodebesluit, 2010]. Met een lineair regressiemodel voor de 2009 data wordt door de *aansluitdichtheid* 50,4% van de variantie in de *kapitaalkosten* van de verschillende netbeheerders verklaard. Dit is een stuk minder dan de verklaarde variantie (78,1%) van het oorspronkelijke regressiemodel gebaseerd op de data van alle netbeheerders over 2006.

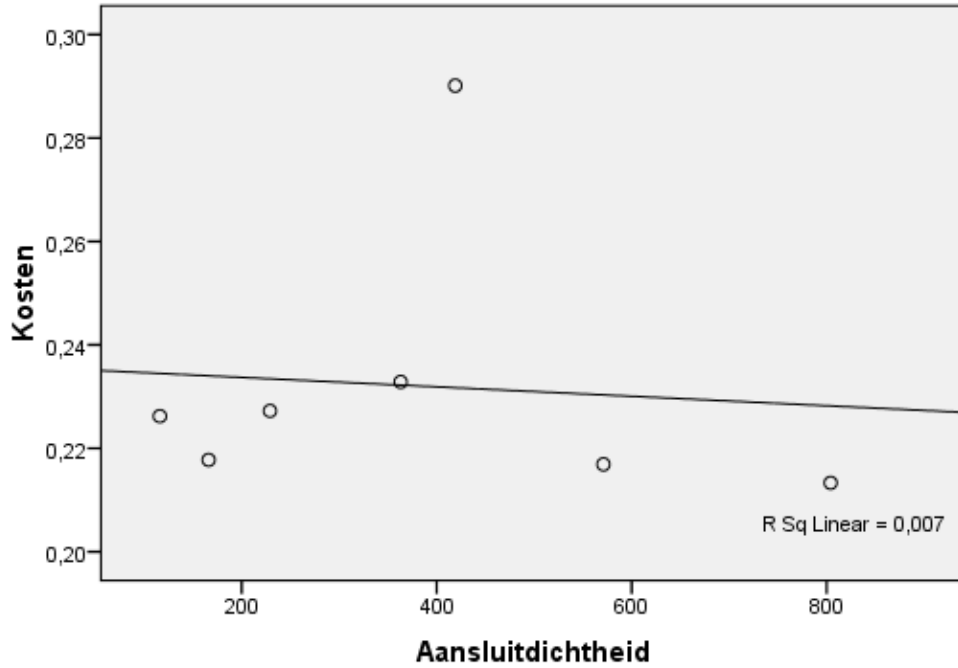
Invloedrijke punten

Om te onderzoeken of dit nieuwe regressiemodel robuust is, zal de invloed van de individuele meetpunten worden onderzocht op dezelfde manier als dit is uitgevoerd voor de 2006 data.

Netbeheerder	Residul (<2,37)	Cook's distance (<0,67)	Leverage value (<0,75)
Continuon	-0,3860	0,0216	0,0654
Overijssel	0,5706	0,0562	0,0886
Delta	-0,6841	0,1096	0,1329
Stedin	0,1225	0,0013	0,0053
Enexis	-0,8147	0,1244	0,1002
NRE	-0,8261	2,9063	0,5861
RENDO	0,0978	0,0009	0,0161
Westland	1,9201	0,3176	0,0053

Figuur 13: Analyse van invloedrijke punten (2009)

Uit Figuur 13 blijkt dat de netbeheerder NRE teveel invloed heeft op het regressiemodel, gezien de veel te hoge Cook's distance en de vrij hoge Leverage value. Deze netbeheerder zou uit de analyse verwijderd moeten worden.



Figuur 14: Spreidingsdiagram en regressielijn (zonder NRE) (2009)

Uit Figuur 14 blijkt duidelijk dat er zonder de netbeheerder NRE nauwelijks sprake is van een lineair verband. De correlatie tussen *kapitaalkosten* en *aansluitdichtheid* bedraagt nog slechts 0,084 en aansluitdichtheid verklaart nog slechts 0,7% van de variantie in kapitaalkosten. De praktische waarde hiervan is dus nihil.

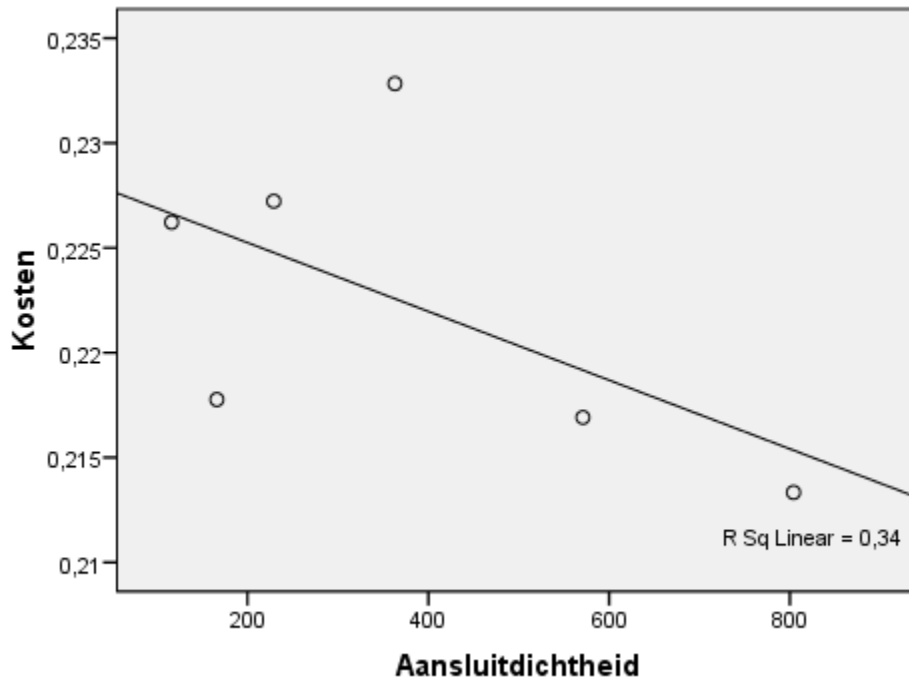
In Figuur 14 is te zien dat de netbeheerder Westland sterk afwijkt van de overige netbeheerders.

Opnieuw zal de invloed van de individuele meetpunten worden onderzocht.

Netbeheerder	Residul (<2,45)	Cook's distance (<0,80)	Leverage value (<0,86)
Continuon	-0,2150	0,0077	0,0650
Overijssel	-0,5139	0,6741	0,5018
Delta	-0,2855	0,0318	0,1973
Stedin	-0,4637	0,0459	0,1012
Enexis	-0,5620	0,0814	0,1299
NRE			
RENDO	0,0211	0,0000	0,0009
Westland	2,0191	0,4114	0,0040

Figuur 15: Analyse van invloedrijke punten (2009)

Uit Figuur 15 blijkt dat de Netbeheerder Westland een erg groot residu heeft t.o.v. de andere netbeheerders. Hij wijkt daarmee erg af van de overige netbeheerders.



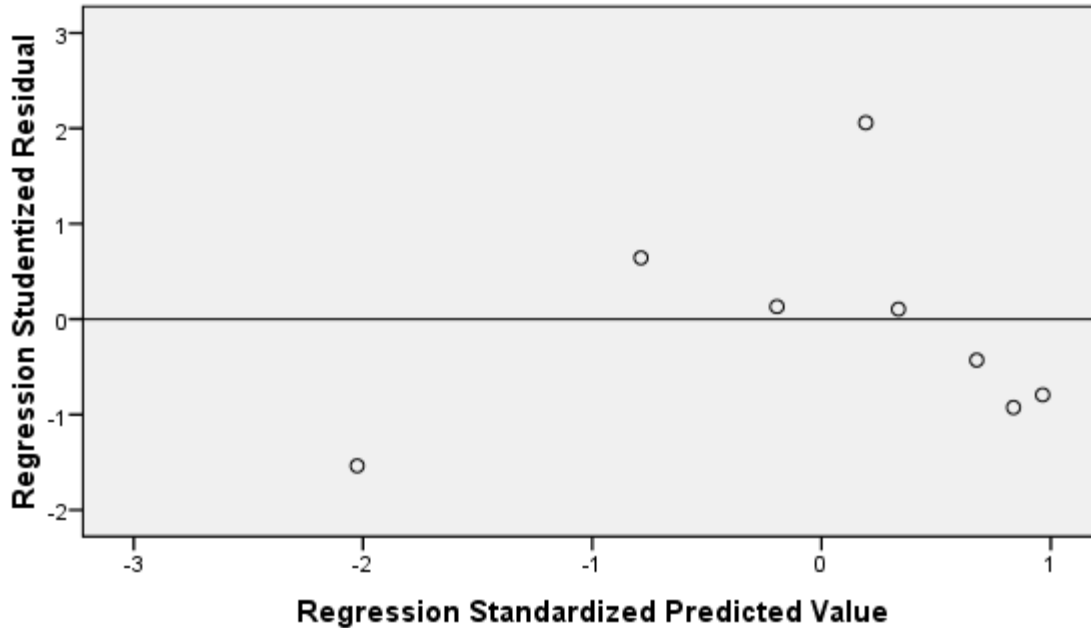
Figuur 16: Spreidingsdiagram en regressielijn (zonder NRE en Westland) (2009)

Figuur 16 laat zien dat er ook zonder de netbeheerder Westland geen sprake is van een duidelijk lineair model. Het blijven verwijderen van meetpunten leidt zeker niet tot een meer robuust model en dient verder geen nut.

Geconcludeerd moet worden dat een regressiemodel gebaseerd op alle 8 netbeheerders niet gebruikt mag worden omdat de invloed van één netbeheerders erg groot is. Zonder deze invloedrijke netbeheerder verklaart het model geen variantie in *kapitaalkosten* meer.

Regressievoorwaarden

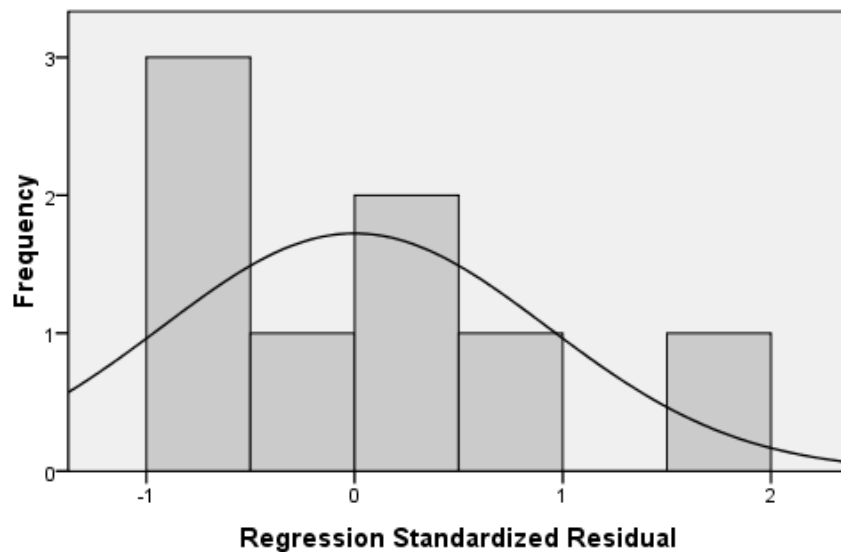
Hoewel al aangetoond is dat het regressiemodel gebaseerd op de gegevens van alle 8 netbeheerders over 2009 in hoge mate bepaald wordt door één netbeheerder, zal opnieuw ook onderzocht worden of aan de vereiste regressievoorwaarden is voldaan.



Figuur 17: Spreidingsdiagram van de residuen t.o.v. de geschatte waarden (2009)

Figuur 17 toont de plot van de residuen uitgezet tegen de gestandaardiseerde verwachte *kapitaalkosten*. Het is duidelijk dat opnieuw op basis van zo weinig meetpunten niets zinnigs gezegd kan worden over de verdeling van de residuen.

Een histogram van de residuen moet aantonen dat de residuen normaal verdeeld zijn.



Figuur 18: Histogram van de residuen (2009)

Figuur 18 toont dit histogram van de residuen en de bijbehorende (zelfde gemiddelde, zelfde variantie) normaalverdeling. Ook hierbij is duidelijk dat op basis van zo weinig meetpunten niets zinnigs valt af te leiden over de verdeling van de residuen. Het histogram geeft in ieder geval geen reden om aan te nemen dat de residuen normaal verdeeld zouden zijn.

Geconcludeerd moet worden dat niet valt aan te tonen dat met zo weinig meetpunten aan de vereiste regressievoorwaarden wordt voldaan, terwijl bij het bepalen van het regressiemodel op basis van ordinary least squares method hier wel van wordt uitgegaan.

Gewogen versus ongewogen regressiemodel

In het voorafgaande is reeds aangetoond dat de netbeheerder NRE een te sterke invloed op het regressiemodel heeft en dat de netbeheerder Westland erg afwijkend gedrag vertoont van de overige netbeheerders. Beide netbeheerders beheren slechts een klein aantal aansluitingen. NRE beheert 1,3% van alle aansluitingen en Westland slechts 0,7% van alle aansluitingen.

Opnieuw zou een betere keuze zijn om de invloed van een netbeheerder uit te drukken in het aantal aansluitingen dat de netbeheerder beheert. In dat geval wordt een gewogen regressieanalyse uitgevoerd. Het gewogen regressiemodel luidt voor de 2009 data:

$$\text{kosten} = 0,231839 - 0,000036 * \text{aansluitdichtheid}$$

Dit regressiemodel verklaart 32% van de variantie in *kapitaalkosten* van de verschillende netbeheerders. Om een goede schatting van de verwachte *kapitaalkosten* te kunnen maken is dit nog steeds te weinig.

Al met al moet geconcludeerd worden dat voor de meetdata uit 2009 dezelfde problemen te verwachten zijn als eerder aangetoond voor de meetdata uit 2006. De invloed van (kleine) netbeheerders is groot, er kan niet worden aangetoond dat aan de vereiste regressievoorwaarden is voldaan en de verklaarde variantie van de gevonden regressiemodellen is laag.

8 Robuustheid van het regressiemodel door de tijd

Uit de eerdere analyses over 2006 en 2009 is gebleken dat, afhankelijk van de verschillende keuzen die gemaakt worden, de resultaten sterk kunnen verschillen. Deze resultaten hangen o.a. af van

- Of netbeheerders met een afwijkend gedrag wel of niet worden meegenomen in de analyse,
- Of het aantal aansluitingen wat een netbeheerder beheert wel of niet als gewicht wordt meegenomen in de analyse,
- Voor welk jaar het regressiemodel voor *kapitaalkosten* gebaseerd op de *aansluitdichtheid* wordt geschat.

Voor een model dat gebruikt gaat worden om voor drie jaren de verwachte *kapitaalkosten* behorend bij een zekere *aansluitdichtheid* te schatten en hierop de verrekening met de individuele netbeheerders te baseren is het niet gewenst dat de resultaten afhangen van niet goed onderbouwde keuzes.

De Energiekamer wil op basis van de 2009 data een regressiemodel schatten op basis van *aansluitdichtheid* waarmee voor de komende 3 jaar de *kapitaalkosten* per netbeheerder geschat zal worden en zal worden vergeleken met de werkelijke *kapitaalkosten* van de netbeheerder [p272, Ontwerp-methodebesluit, 2010]. Aangezien de data voor toekomstige jaren nog niet bekend is, kan het effect hiervan niet direct beoordeeld worden.

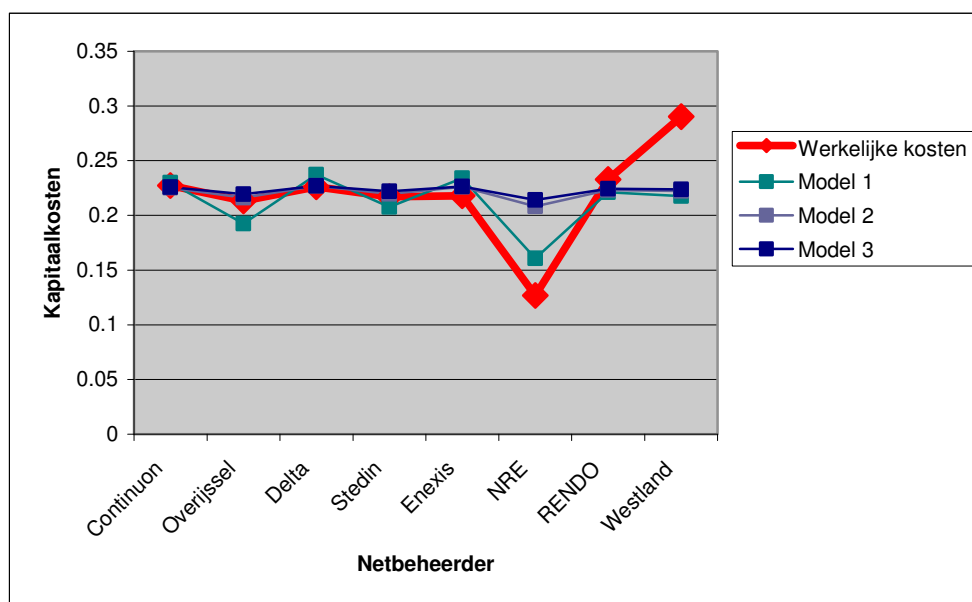
Wel kan onderzocht worden wat het effect zou zijn als de data uit 2006 gebruikt zouden worden om een inschatting te maken van de *kapitaalkosten* over 2009. Dit geeft een indicatie voor de betrouwbaarheid van de modellen over de jaren heen.

Hieronder zal voor de verschillende keuzes die gemaakt kunnen worden, de resultaten worden vergeleken met de werkelijke *kapitaalkosten* (zie Figuur 10) in 2009.

Model	Jaar	Netbeheerders	(On)gewogen	Model constante	Regressie coëfficiënt	Verklaarde variantie
1	2006	Alle 9	Ongewogen	0,244728	-0,000065	78,1%
2	2006	Alle 9	Gewogen	0,228872	-0,000016	4,4%
3	2006	Zonder ONS en NRE	Ongewogen	0,228301	-0,000011	1,9%

Figuur 19: Verschillende modelkeuzen gebaseerd op data uit 2006

Figuur 19 geeft de diverse modellen die bepaald zouden kunnen worden met de data uit 2006. Deze modellen zijn in de voorgaande hoofdstukken reeds besproken. Figuur 20 laat zien in hoeverre de schattingen voor de *kapitaalkosten* gemaakt met de diverse modellen afwijken van de werkelijke *kapitaalkosten* over 2009.



Figuur 20: Modellen uit 2006 toegepast op de data uit 2009

Figuur 21 geeft de percentuele fout, $\frac{y_i - \hat{y}_i}{y_i}$, t.o.v. van de werkelijke kosten per netbeheerder.

Hierin is y_i de werkelijke *kapitaalkosten* in 2009 voor netbeheerder i en \hat{y}_i de geschatte *kapitaalkosten* m.b.v. één van de regressiemodellen uit 2006.

Een negatieve (positieve) waarde betekent dat de netbeheerder lagere (hogere) kapitaalkosten had dan verwacht kon worden aan de hand van het model en dat de netbeheerder dus een positieve (negatieve) verrekening zou kunnen verwachten.

Netbeheerder	Model 1	Model 2	Model 3
Continuon	-1,15%	0,89%	0,64%
Overijssel	9,78%	-1,25%	-2,87%
Delta	-4,85%	-0,35%	-0,36%
Stedin	4,29%	-1,30%	-2,36%
Enexis	-7,43%	-3,88%	-4,00%
NRE	-26,87%	-64,22%	-68,86%
RENDO	5,03%	4,20%	3,66%
Westland	25,03%	23,42%	22,90%

Figuur 21: Percentuele fout per netbeheerder

Uit Figuur 21 blijkt duidelijk dat afhankelijk van de gemaakte modelkeuzes, de verrekening voor de individuele netbeheerders totaal anders kan uitpakken. Dit soort willekeur ondermijnt het vertrouwen in de gebruikte methode.

Er is geen reden om aan te nemen dat bij het gebruik van een model gebaseerd op de data uit 2009 voor de verrekening van kapitaalkosten in de komende 3 jaar dit probleem niet op zal treden.

9 Conclusies

Op basis van het in dit rapport besproken onderzoek, kunnen de volgende conclusies worden getrokken.

Conclusie 1:

Het regressiemodel zoals het is gebaseerd op de dataset met gegevens over de aansluitdichtheid en kapitaalkosten van 9 netbeheerders uit 2006 toont niet aan dat er een sterke relatie is tussen de aansluitdichtheid en de kapitaalkosten.

Deze conclusie is gebaseerd op de volgende observaties:

- Het onderzoek is uitgevoerd op een complete dataset. Er is dus geen sprake van onzekerheid in de gevonden relatie. De significantiewaarde die volgt uit de statistische analyses kan zeker niet gebruikt worden om aan te tonen dat de resultaten enige praktische relevantie hebben. Alleen de hoogte van de correlatiecoëfficiënt en de daarmee samenhangende verklaarde variantie door het geschatte regressiemodel is hiervoor bepalend. Er bestaat echter geen eenduidige ondergrens waarboven sprake is van een sterke relatie tussen de twee variabelen *aansluitdichtheid* en *kapitaalkosten*.
- De dataset is compleet maar met 9 netbeheerders erg klein. De invloed van twee individuele netbeheerders is hierdoor onaanvaardbaar groot en zij bepalen voor het grootste gedeelte de relatie tussen *aansluitdichtheid* en *kapitaalkosten*. Na verwijdering van deze 2 netbeheerders, die samen slechts verantwoordelijk zijn voor 2% van alle aansluitingen, wordt nog nauwelijks enige variantie (1,9%) in *kapitaalkosten* van de overige netbeheerders verklaard.
- De dataset is erg klein, hierdoor is het niet aan te tonen dat aan een aantal belangrijke regressievoorwaarden is voldaan. De gebruikte regressiemethode is echter niet geldig als niet aan deze voorwaarden is voldaan. In dat geval kan het regressiemodel niet worden gebruikt om betrouwbare schattingen te maken.

- De keuze om alle netbeheerders hetzelfde gewicht te geven in de regressieanalyse is zeer bepalend voor het uiteindelijke resultaat. Als een gewogen regressieanalyse wordt uitgevoerd, waarbij het gewicht van een netbeheerder wordt bepaald door het aantal aansluitingen dat zij beheren, dan blijft er van het gevonden verband tussen *aansluitdichtheid* en *kapitaalkosten* nauwelijks meer wat over en de verklaarde variantie (4,4%) is ook in dat geval te laag om nog enige praktische relevantie te bezitten.

Conclusie 2:

Het regressiemodel zoals het zou worden vastgesteld op de dataset met gegevens over de aansluitdichtheid en kapitaalkosten van 8 netbeheerders uit 2009 toont niet aan dat er een sterke relatie is tussen de aansluitdichtheid en de kapitaalkosten.

Deze conclusie is gebaseerd op de volgende observaties:

- De dataset is compleet maar met 8 netbeheerders erg klein. De *aansluitdichtheid* van deze netbeheerders verklaart 50% van de *kapitaalkosten*, maar de invloed op dit resultaat van één individuele netbeheerder is onaanvaardbaar groot. Deze netbeheerder beheert echter slechts 1,3% van alle aansluitingen. Na verwijdering van deze netbeheerder wordt nog slechts 0,7% van de variantie in *kapitaalkosten* van de overige netbeheerders verklaard door de *aansluitdichtheid*.
- De dataset is zelfs kleiner dan voor de data uit 2006. Hierdoor is opnieuw niet aan te tonen dat aan een aantal belangrijke regressievoorwaarden is voldaan. De gebruikte regressiemethode is echter niet geldig als niet aan deze voorwaarden is voldaan. In dat geval kan het regressiemodel niet worden gebruikt om betrouwbare schattingen te maken.
- Een gewogen regressieanalyse geeft een model waarmee nog slechts 32% van de variantie in *kapitaalkosten* wordt verklaard door de *aansluitdichtheid*.

Conclusie 3:

Een robuust regressiemodel op basis van de aansluitdichtheid dat houdbaar is voor meerdere jaren en dat als basis kan dienen voor een vergelijking van de verwachte kapitaalkosten met de werkelijke kapitaalkosten en de daarbij behorende verrekening, kan niet bepaald worden voor deze kleine groep netbeheerders.

Deze conclusie is gebaseerd op de volgende observatie:

- Met de meetdata uit 2006 kunnen verschillende regressiemodellen worden geschat. De verschillen tussen deze modellen zijn gebaseerd op het wel of niet meenemen van individuele netbeheerders die de resultaten sterk beïnvloeden óf het wel of niet mee laten wegen van het aantal aansluitingen dat een netbeheerder beheert. Voor een aantal individuele netbeheerders leveren deze modellen totaal verschillende resultaten op, waarbij het zelfs met het ene model tot een negatieve verrekening en met het andere model tot een positieve verrekening leidt.

Al met al moet geconcludeerd worden dat *aansluitdichtheid* niet bepalend is voor de *kapitaalkosten* van een netbeheerder en dus als zodanig niet kan dienen als een objectiveerbaar regionaal verschil.

Referenties

[Hair et al., 2006]

Hair, J.F., R.E. Anderson, R.L. Tatham, W.C. Black,
Multivariate data analysis, Prentice Hall, USA, 6th edition
2006, ISBN 0-13-032929-0

[Ontwerp-methodebesluit, 2010]

Nederlandse Mededingsautoriteit, Ontwerp-methodebesluit
vijfde reguleringsperiode regionale netbeheerders elektriciteit,
2010