KM23(1987)
pag 45 - 62

ASYMMETRISCHE KLEINSTE KWADRATEN:
NIEUWE GEZICHTEN VAN EEN SCATTERPLOT

P.H.C. Eilers*

Samenvatting

In de som van kwadraten van residuen ten opzichte van een regressiemodel, kan men positieve en negatieve residuen verschillend wegen. Daaruit volgen parameterschattingen die informatief zijn over de spreiding in de gegevens. Deze techniek wordt geïllustreerd met bivariate regressielijnen en smoothing van tijdreeksen. Combinatie met een nieuw eenvoudig smoothingalgoritme geeft een flexibel hulpmiddel voor grafische ondersteuning van scatterplots. Voor het oplossen van het asymmetrische kleinste kwadratenprobleem kan men gebruik maken van iteratief gewogen regressie.

*
Dienst Centraal Milieubeheer Rijnmond
's Gravelandseweg 565
3119 XT Schiedam
010-4273699

1. Inleiding

Een doelmatige en wijd verbreide techniek voor het zichtbaar maken van het verband tussen twee variabelen, is het tekenen van een strooiingsdiagram (in het engels: scatterplot of scattergram, in weinig fraai nederlands: strooiplaat). Door voor iedere waarneming een punt of ander teken op papier of beeldscherm te tekenen, met de beide variabelen als coördinaten, ontstaat een puntenwolk. Deze wolk geeft een beeld van ligging en spreiding van beide variabelen en van vorm en sterkte van een mogelijk functioneel verband. Dank zij de grafische mogelijkheden van hedendaagse (micro-)computers, is het tekenen van een scatterplot welhaast een routine-aangelegenheid geworden. Cleveland en McGill (1984) hebben overtuigend aangetoond dat een scatterplot aanmerkelijk aan waarde kan winnen door ondersteuning met grafieken van afgeleide grootheden. Zij spreken van de vele gezichten van een scatterplot. In dit artikel wil ik de portrettengalerij uitbreiden met een nieuw conterfeitsel, dat vooral aandacht schenkt aan de spreiding in de scatterplot.
De techniek die hieronder besproken zal worden, is geïnspireerd door de smoothing-techniek in Cleveland en McGill (1984), voor het tekenen van een glad verlopende "ruggegraat" van de puntenwolk. De basis is een simpele aanpassing van de kleinste kwadraten doelfunctie. Bij normale kleinste kwadraten methoden is de doelfunctie de som van de kwadraten der residuen. Er wordt geen onderscheid gemaakt tussen positieve en negatieve residuen. Doen we dat wel, door de kwadraten van de negatieve residuen meer of minder zwaar te wegen in de doelfunctie, dan volgen interessante en bruikbare resultaten. De toepassingen die volgen zullen dat illustreren.
Voor zover mij bekend, is deze asymmetrische weging van de kwadraten der residuen niet eerder toegepast. Voor L_1-regressie, waar de som van de absolute waarden van de residuen geminimaliseerd wordt, is het werk van Koenker en Bassett (1978) bekend. Hun "regressiequantielen" worden berekend door negatieve en positieve residuen ongelijk te wegen. Veel navolgers hebben zij niet gevonden, vermoedelijk omdat de implementatie van de berekening niet zo eenvoudig is: men moet namelijk op compacte wijze een lineair programma oplossen. Wellington en Narula (1984) geven een implementatie. Het rekenwerk voor asymmetrische kleinste kwadraten kan opgelost worden met herhaalde gewogen regressie. Dat maakt toepassingen gemakkelijk te implementeren.
De smoothing die hier gebruikt wordt, is eenvoudiger in formulering en implementatie, dan de "nearest-neighbour smoothing" van Cleveland en McGill.

2. Asymmetrische kleinste kwadraten

Stel dat waarnemingen y_i en x_{ij} gegeven zijn met $i = 1..m$ en $j = 1..n$. Het klassieke lineaire regressie probleem is nu:

$$\text{minimaliseer} \quad S = \sum_{i=1}^{m} (y_i - \hat{y}_i)^2 \tag{2.1}$$

$$\text{met} \quad \hat{y}_i = \sum_{j=1}^{n} x_{ij} a_j \tag{2.2}$$

Een uitbreiding van (2.1) is gewogen regressie:

$$\text{minimaliseer} \quad S_2 = \sum_{i=1}^{m} w_i (y_i - \hat{y}_i)^2 \tag{2.3}$$

Bij gegeven waarden van w_i is de oplossing van (2.3) even gemakkelijk te berekenen als die van (2.1).
Asymmetrische kleinste kwadraten worden als volgt gedefinieerd: kies

$$\begin{aligned} w_i &= 1 - p \quad \text{als} \quad y_i \leqslant \hat{y}_i \\ w_i &= p \qquad\;\; \text{als} \quad y_i > \hat{y}_i \end{aligned} \tag{2.4}$$

$$\text{met} \quad 0 < p < 1$$

Er is sprake van een lastige koppeling tussen w_i en $\hat{y}_i$: om de gewichten te bepalen moet men $\hat{y}_i$ kennen en om $\hat{y}_i$ te berekenen moeten de gewichten bekend zijn. De standaard uitweg uit dergelijke situaties blijkt ook hier te werken: neem een waarde voor de a_j's aan (bijvoorbeeld het resultaat van ongewogen regressie), bereken daaruit $\hat{y}_i$ en w_i, los (2.3) op om verbeterde waarden voor de a_j's te vinden. Herhaal deze procedure tot de a_j's niet meer veranderen. De appendix gaat in op de convergentie. De praktijk heeft geleerd dat tussen vijf en tien iteraties vrijwel altijd voldoende zijn.

Voor zover mij bekend, is een asymmetrische weging volgens (2.4) nieuw voor kleinste kwadraten. Het is echter te beschouwen als een veralgemenisering van de "regressiequantielen" van Koenker en Bassett (1978). Zij gebruiken als doelfunctie

$$S_1 = \sum_{i=1}^{m} w_i |y_i - \hat{y}_i| \qquad (2.5)$$

met gewichten volgens (2.4) en $\hat{y}_i$ volgens (2.2). Het oplossen van (2.5) en (2.2) komt neer op het oplossen van een lineair programma. Tenzij men bijzondere voorzorgen neemt, zijn geheugenbeslag en rekentijd zeer omvangrijk voor grote waarden van m. Koenker en Bassett (1978) geven aan dat het effectief is het duale probleem op te lossen. Het construeren van een goed computerprogramma voor dit probleem is geen lichte taak, zie Wellington en Narula (1984).
Het ligt voor de hand (2.3) of (2.5) te veralgemeniseren tot

$$\text{minimaliseer} \quad S_r = \sum_{i=1}^{m} w_i |y_i - \hat{y}_i|^r \qquad (2.6)$$

In principe kan r iedere (positieve?) waarde aannemen. Of andere waarden dan r=1 of r=2 bruikbaar zijn, zal verder onderzoek moeten aantonen. Voor zeer hoge waarden van r is S_r een uitbreiding van de L_∞ norm, zie Philips en Taylor (1973). Bij het benaderen van gegevens of functies met de L_∞ norm, wordt nagestreefd dat de absolute waarden van de grootste residuen $y_i - \hat{y}_i$ zo klein mogelijk zijn. Voor statistische gegevens zal men deze norm niet vaak kiezen maar bij het interpoleren van cijferreeksen van hoge kwaliteit, zoals tabelwaarden, wordt ze vaak toegepast (Chebysher-interpolatie, zie Philips en Taylor, 1973). Door gewichten te kiezen volgens (2.4) kunnen de negatieve residuen zwaarder of minder zwaar gewogen worden dan de positieve. Er zijn toepassingen denkbaar waar overschatten zwaarder weegt dan onderschatten.
De rest van dit artikel zal alleen van S_2 gebruik maken.

3. Toepassing op een bivariate regressielijn

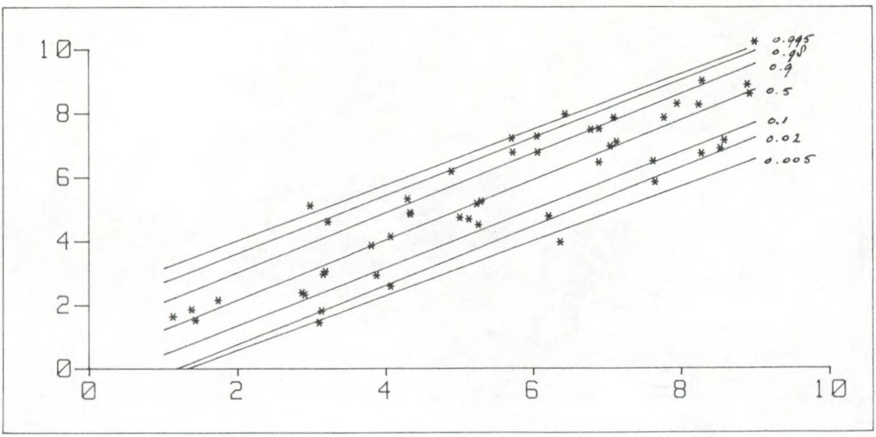
Als een eenvoudige illustratie van de resultaten die het gebruik van asymmetrische kleinste kwadraten op kan leveren, kunnen de figuren 3.1 en 3.2 dienen. In beide figuren zijn scatterplots getekend van gesimuleerde gegevens. Uit een uniforme verdeling ($1 < x_i < 9$) werden 50 waarden getrokken. De bijbehorende y_i werden in figuur 3.1 geconstrueerd als

$$y_i = x_i + \underline{e}_i \tag{3.1}$$

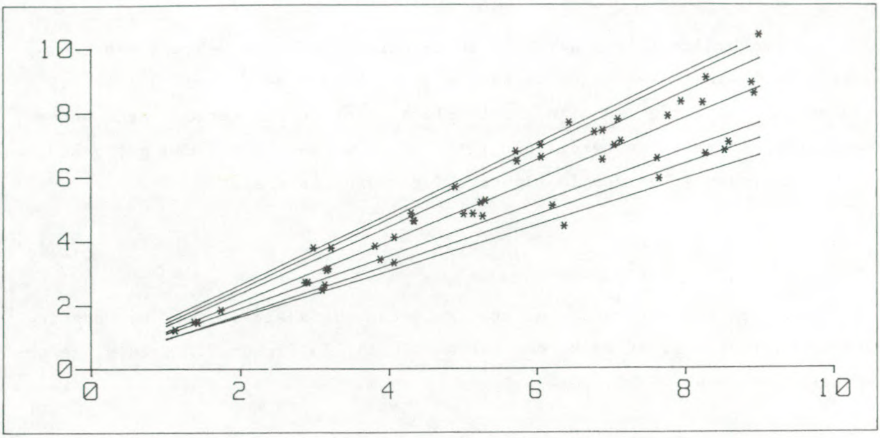
waarin $\underline{e}_i$ een normaal verdeelde stochastische variabele is, met verwachting nul en variantie 1, af te korten tot $\underline{e} \sim N(0,1)$. In figuur 3.2 werd y_i geconstrueerd volgens:

$$y_i = x_i + \underline{e}_i . x_i \tag{3.2}$$

De met x toenemende spreiding is in de laatste figuur goed waarneembaar, door het breder worden van de puntenwolk.
In beide figuren zijn lijnen $y = a(p)x + b(p)$ getekend met S_2 volgens (2.3) en gewichten volgens (2.4); de bijbehorende waarde van p is naast de lijnen aangegeven. De lijn voor $p = 0.5$ komt overeen met de klassieke ongewogen regressielijn van y op x.

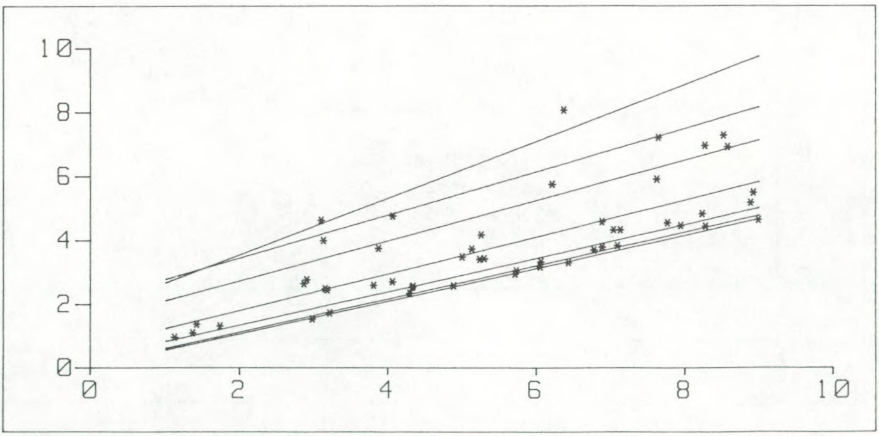


Figuur 3.1 Asymmetrische kleinste kwadratenlijnen bij normaal verdeelde verstoringen met constante variantie.

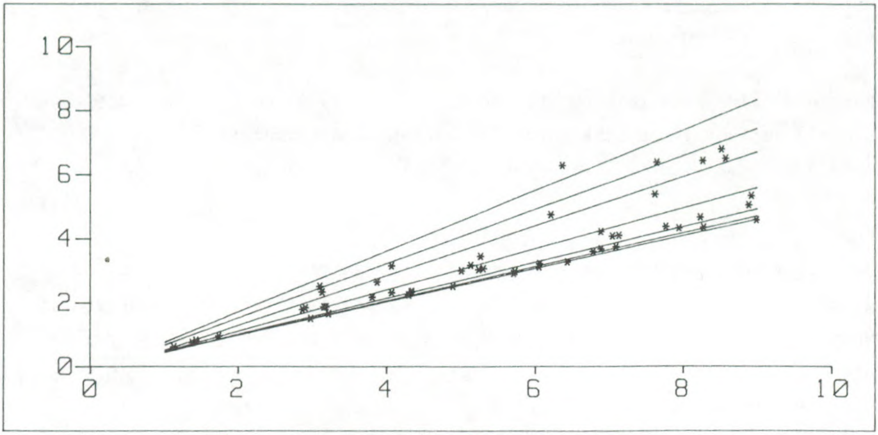


Figuur 3.2 Asymmetrische kleinste kwadratenlijnen bij normaal verdeelde verstoringen met variabele variantie.

In figuur 3.1 blijken de lijnen min of meer parallel te lopen; in figuur 3.2 lopen ze met toenemende x uiteen. Het verloop van de lijnen weerspiegelt het verloop van de spreiding van $\underline{e}$.
Merk op dat de bundels min of meer symmetrisch uitwaaieren ten opzichte van de lijn voor p = 0.5. Dit geeft aan dat de verdeling van $\underline{e}$ symmetrisch is.



Figuur 3.3. Asymmetrische kleinste kwadratenlijnen voor verstoringen met exponentiële verdeling en constante variantie.



Figuur 3.4 Asymmetrische kleinste kwadratenlijnen voor verstoringen met exponentiële verdeling en variabele variantie.

Voorbeelden voor niet-symmetrisch verdeelde verstoringen geven de figuren 3.3 en 3.4. De x-waarden zijn op dezelfde wijze gesimuleerd als in de figuren 3.1 en 3.2. In figuur 3.2 is y gesimuleerd volgens

$$y_i = x_i/2 + \underline{e}_i \qquad (3.3)$$

en in figuur 3.4 volgens

$$y_i = x_i/2 + \underline{e}_i . x_i \qquad (3.4)$$

Hier is $\underline{e}_i$ getrokken uit een exponentiële verdeling met verwachtingswaarde 1.

Nu blijken voor lage waarden van p de lijnen dichter bij elkaar te liggen dan voor hoge waarden van p. De afstanden tussen de lijnen weerspiegelen het verloop van de kansdichtheid van $\underline{e}$.
In de figuren 3.1 - 3.4 zijn de waarden van p: 0.005, 0.02, 0.1, 0.5, 0.9, 0.98, 0.995. De onderste lijn correspondeert met de laagste waarde van p.

4. Smoothing

4.1. Het smoothing-algoritme

Ruim een halve eeuw geleden stelden Whittaker (1923) en Henderson (1924) een simpele smoothing-techniek voor. Zij kiezen de doelfunctie

$$S = \sum_{i=1}^{m} (y_i - z_i)^2 + \beta \sum_{i=k+1}^{m-k+1} (\nabla^k z_i)^2 \qquad (4.1)$$

Hier is y_i een gegeven reeks op equidistante punten ($i = 1..m$) en z_i de gesmoothede versie. Met ∇^k wordt de k-de centrale differentie aangegeven, voor de waarden van i waarvoor deze gedefinieerd is. Hoe groter β, des te sterker de smoothing.
Uitstekende resultaten geeft $k = 2$:

$$\nabla^2 z_i = z_{i-1} - 2\, z_i + z_{i+1} \qquad i = 2..m-1 \qquad (4.2)$$

Een uitbreiding van (4.1) met k=2 is:

$$S = \sum_{i=1}^{m} w_i (y_i - z_i)^2 + \beta \sum_{i=2}^{m-1} (z_{i-1} - 2z_i + z_{i+1})^2 \qquad (4.3)$$

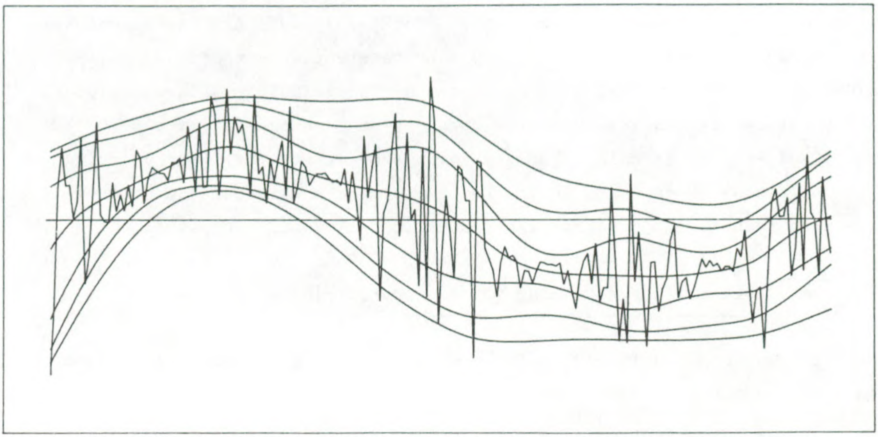
De gewichten in (4.3) worden gekozen volgens (2.4), om positieve en negatieve residuen verschillend te wegen. Een iteratieve oplossing van de uit (4.3.) volgende vergelijkingen is dan nodig. Die vergelijkingen zijn op zich gemakkelijk op te lossen. Door differentiëren van (4.3) volgt:

$$c'_i\, z_{i-2} + b'_i\, z_{i-1} + (\alpha w_i + a_i)\, z_i + b_i\, z_{i+1} + c_i\, z_{i+2} = \alpha w_i\, y_i \qquad (4.4)$$

Hierin geldt:

$$\alpha = 1/\beta \qquad (4.5)$$

$c'_i = 0,\ b'_i = 0,\ a_i = 1,\ b_i = -2,\ c_i = 1$ als $i = 1$
$c'_i = 0,\ b'_i = -2,\ a_i = 5,\ b_i = -4,\ c_i = 1$ als $i = 2$
$c'_i = 1,\ b'_i = -4,\ a_i = 6,\ b_i = -4,\ c_i = 1$ als $2 < i < m-1$
$c'_i = 1,\ b'_i = -4,\ a_i = 5,\ b_i = -2,\ c_i = 1$ als $i = m-1$
$c'_i = 1,\ b'_i = -2,\ a_i = 1,\ b_i = 0,\ c_i = 0$ als $i = m$

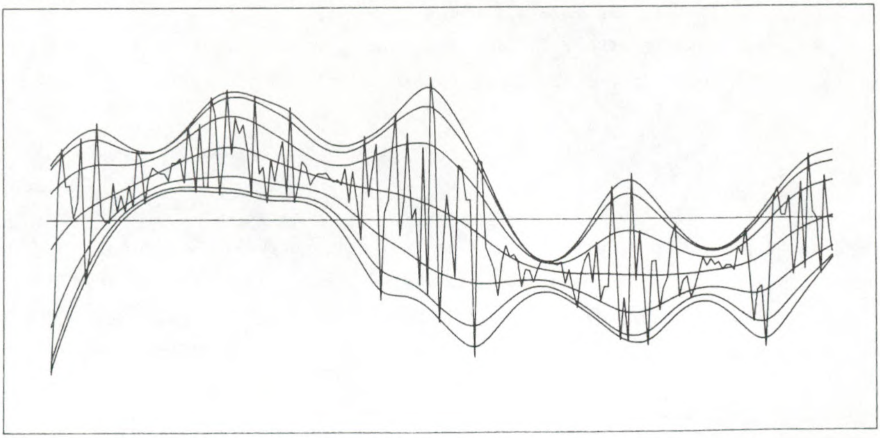


Figuur 4.1 Asymmetrisch gewogen smoothing

Het stelsel vergelijkingen heeft dus een bandvorm met twee nevendiagonalen. Met Gausse eliminatie of Cholesky-decompositie laat het zich snel oplossen. Een illustratie van smoothing met asymmetrisch gewogen kleinste kwadraten ($\beta = 10^4$) geeft figuur 4.1. De data zijn gesimuleerd volgens

$$y_i = \sin(fi) + e_i(\cos(2fi) + 0.3) \qquad (4.6)$$

met $e_i \sim N(0,1)$. Het valt in figuur 4.1 op dat de dalen min of meer afgerond worden voor waarden van p dicht bij 0 of 1. De curven zijn niet "soepel" genoeg.



Figuur 4.2 Asymmetrisch gewogen smoothing; korrektie met p(1-p)

Dit is te wijten aan het feit dat de eerste term van S relatief kleiner wordt dan de tweede naarmate p dichter bij 0 of 1 ligt. Als ad-hoc oplossing is β in (4.3) vervangen door $\beta'p(1-p)$. De referent suggereerde $\beta'(p(1-p))^r$ als algemener alternatief. Met r kan men de "soepelheid" variëren. Figuur 4.2 geeft resultaten voor dezelfde data, na deze aanpassing, met $\beta = 10^4 p(1-p)$. Er is duidelijk sprake van een verbetering. Een fundering van deze oplossing is gewenst. Gezien het exploratieve karakter van de voorgestelde techniek lijkt deze ad-hoc constructie echter acceptabel.

4.2 Symmetrische smoothing van een scatterplot

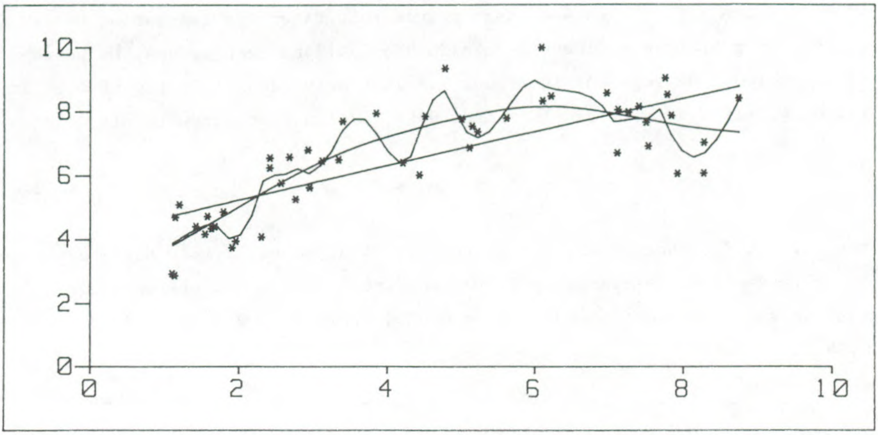
Stel dat puntenparen (x_i, y_i), i= 1..m, gegeven zijn. Verdeel het x-domein in n even brede stroken, genummerd 1..n, tussen de grenzen x_{min} en x_{max}. Hier is x_{min} de laagst voorkomende waarde van x_i en x_{max} de hoogste. Iedere x_i kan aan een strook toegewezen worden en krijgt daardoor een index j(i):

$$j(i) = \mathrm{int}((x_i - x_{min})/(x_{max} - x_{min})) + 1 \qquad (4.7)$$

Van alle waarnemingen binnen strook j, kan een gemiddelde u_j gevormd worden alsmede een gewicht v_j dat gelijk is aan het aantal waarnemingen in strook j. Als we nu de reeks u_j smoothen, met deze gewichten v_j, ontstaat een reeks z_j die de trend van de puntenwolk weergeeft. Het aantal stroken (n) is niet kritisch, waarden van 20 à 50 voldoen goed. Door de interpolerende eigenschappen van (4.3) voor gewichten die nul zijn, kunnen stroken zonder waarnemingen geen kwaad. De waarde van β kan men variëren om een geschikte balans tussen details en grote lijnen te bereiken. Zeer hoge waarden van β leveren de lineaire regressielijn van y op x op.
Een illustratie geeft figuur 4.3. De data zijn gesimuleerd. De x_i's zijn getrokken uit een uniforme verdeling tussen 1 en 9. De y_i's voldoen aan:

$$y_i = 8 - (x_i-6)^2/5 - e_i \qquad i = 1..50 \qquad (4.8)$$

met $\underline{e} \sim N(0.1)$. Het aantal stroken is 50.
Drie waarden van β, namelijk 1, 10^3 en 10^6, zijn gebruikt. De laagste waarde geeft een te grillige lijn; de hoogste waarde geeft vrijwel de lineaire regressielijn. De waarde 10^3 geeft een op het oog acceptabele aanduiding van de trend in de wolk. Het kost weinig rekenwerk met β te experimenteren, omdat de data eerst samengevat worden tot u_j en v_j voor j = 1..n, en omdat het oplossen van de smoothing-vergelijkingen snel gaat.



Figuur 4.3. Smoothing van een scatterplot ($\beta = 1,\ 10^3,\ 10^6$)

4.3. Asymmetrisch smoothing van een scatterplot

Alvorens over te gaan tot de toepassing van asymmetrische kleinste kwadraten, is het nuttig de hierboven gegeven wijze van smoothen scherper te formuleren. Daartoe introduceren we een indicator array G met elementen g_{ij}, i = 1..m, j = 1..n. Dit array wordt als volgt geconstrueerd : $g_{ij} = 1$ als waarneming i in strook j valt, anders $g_{ij} = 0$. Op iedere rij van G komt dus maar één 1 voor. Verder geldt:

$$v_j = \sum_{i=1}^{m} g_{ij} \tag{4.9}$$

Het array G dient alleen voor een gemakkelijke notatie; voor de uitvoering van de berekeningen hoeft het niet expliciet geconstrueerd te worden.
Het smoothing algoritme uit § 4.2 laat zich nu formuleren als: minimaliseer

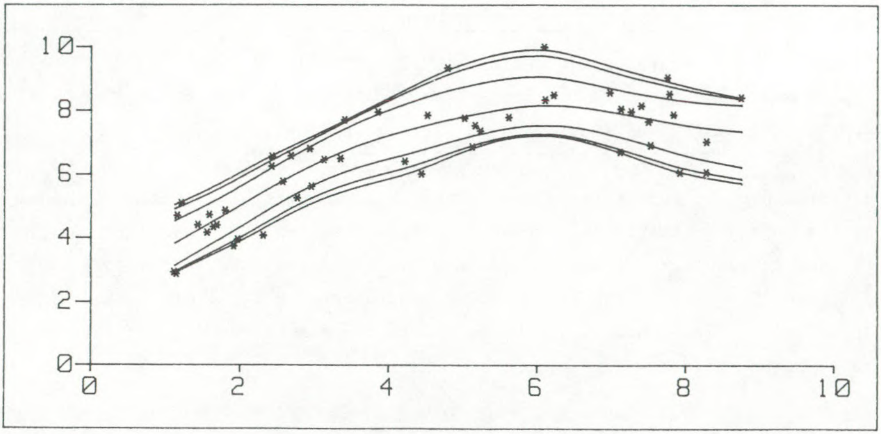
$$S = \sum_{i=1}^{m} w_i (u_i - \sum_{j=1}^{m} g_{ij}\ z_j)^2 + \beta \sum_{j=2}^{m-1} (z_{j-1} - 2z_j + z_{j+1})^2 \tag{4.10}$$

met $w_i = 1$. Uitbreiding naar asymmetrische kleinste kwadraten volgt onmiddellijke door w_i te kiezen volgens (2.4) en β te vervangen door $\beta p(1-p)$. De berekeningswijze is nu iteratief. Experimenteren met waarden van β is beduidend "duurder", omdat voorafgaande samenvatting van de gegevens in de u_j's en v_j's niet meer mogelijk is.

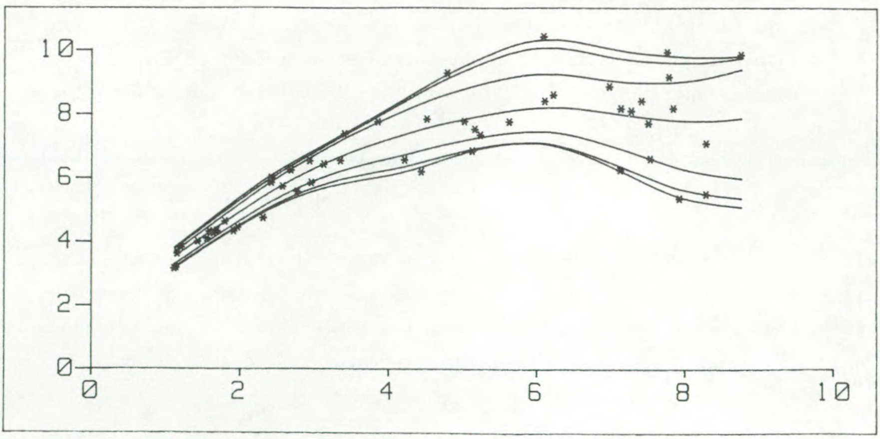
In de figuren 4.4, 4.5 en 4.6 staan enkele resultaten van toepassing van (4.10). De gebruikte waarden van p zijn bij de lijnen aangegeven. De waarde van β is 1000. De gegevens in figuur 4.4 zijn dezelfde als in figuur 4.3. In figuur 4.5 zijn de x_i's dezelfde maar de y_i's zijn geconstrueerd als:

$$y_i = 8 - (x_i - 6)^2/5 - e_i.x_i/5 \tag{4.12}$$

met $\underline{e} \sim N(0,1)$. Voor figuur 4.6 is formule (4.8) gebruikt maar met $\underline{e}$ exponenticel verdeeld met verwachting 2. De onderste lijnen horen bij de laagste waarden van p. De waarden van p zijn 0.005, 0.02, 0.1, 0.5, 0.9, 0.98 en 0.995.

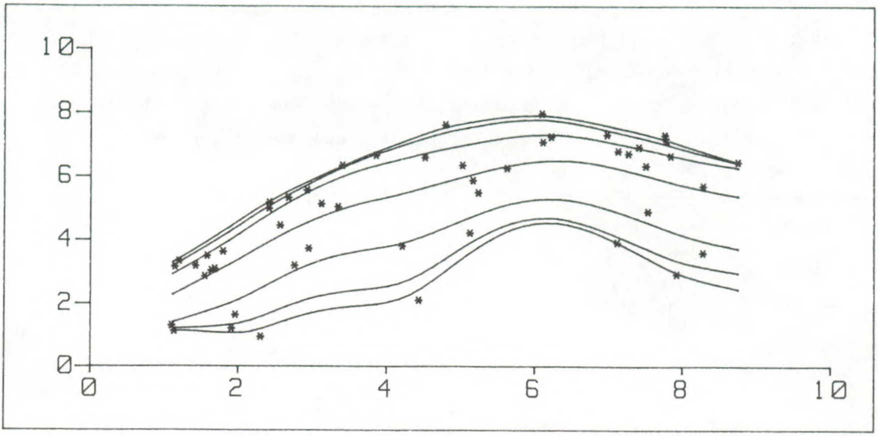


Figuur 4.4 Asymmetrische smoothing van een gesimuleerde puntenwolk



Figuur 4.5 Asymmetrische smoothing van een gesimuleerde puntenwolk

In beide figuren sluiten de lijnen goed aan bij de visuele impressie die men van de puntenwolk krijgt. De verschillen in kansdichtheid van de waarnemingen worden door het meer of minder dicht naast elkaar lopen van de lijnen goed weergegeven.



Figuur 4.6 Asymmetrische smoothing van een gesimuleerde puntenwolk

Figuur 4.7 toont een toepassing op niet-gesimuleerde gegevens. Ze zijn afkomstig uit een onderzoek naar de betrouwbaarheid van ruimtelijke interpolatie van metingen die met 31 meetpunten voor zwaveldioxide uitgevoerd werden. In de figuur is de correlatiecoëfficiënt van de logaritmen van de gemeten concentraties voor elk paar stations uitgezet tegen de onderlinge afstand van de stations. Er zijn 31x30/2 = 465 punten.

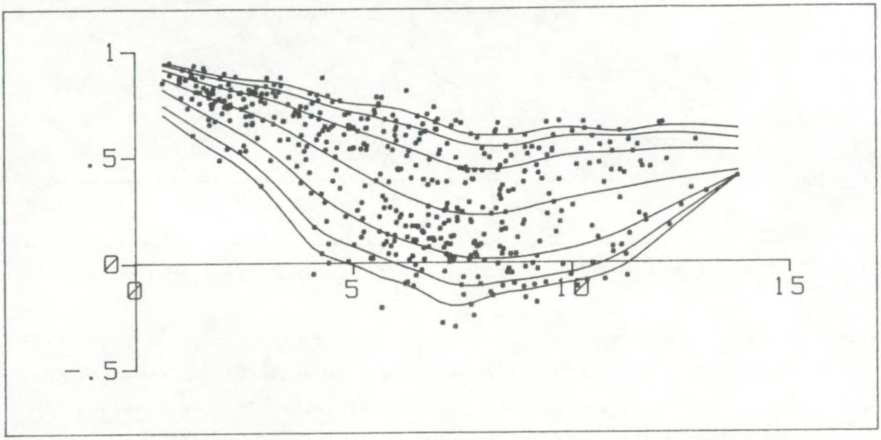
De vijf curven in figuur 4.7 zijn het resultaat van asymmetrisch gewogen smoothing; β is 10000 en het aantal stroken 50. Opmerkelijk is de tendens naar hogere correlaties bij afstanden boven ca. 8 km, die in de puntenwolk zelf niet opvalt.

Het smaller worden van de puntenwolk in figuur 4.7 bij hogere waarden van de correlatiecoëfficiënt is deels te beschouwen als een artefact, omdat de waarden nu eenmaal niet hoger dan 1 kunnen worden. Een bekende transformatie is Fisher's z, die berekend wordt als (Kendall en Stuart, 1977):

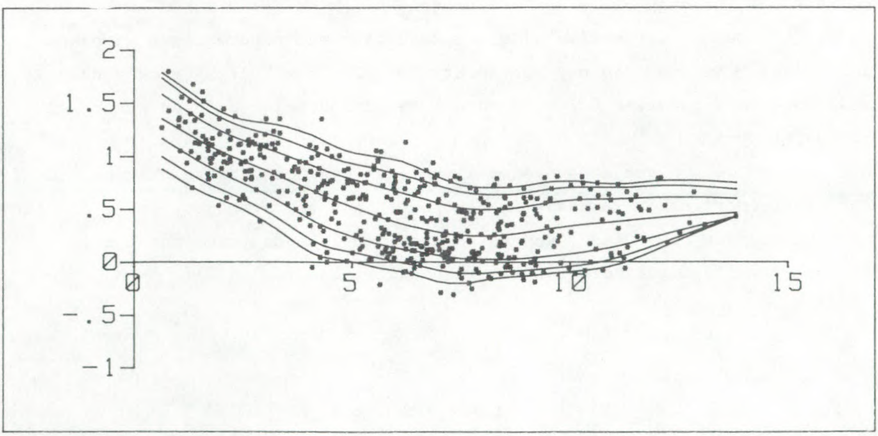
$$z = \frac{1}{2} \ln \frac{(1 + r)}{1 - r} \tag{4.12}$$

In figuur 4.8 is het verloop van z tegen de afstand tussen stations weergegeven, met daarbij de asymmetrische smoothing. De waarden van p zijn dezelfde als in figuur 4.11. De curven blijven op vrijwel gelijke afstanden van elkaar lopen. De transformatie blijkt dus met succes de spreiding te stabiliseren.

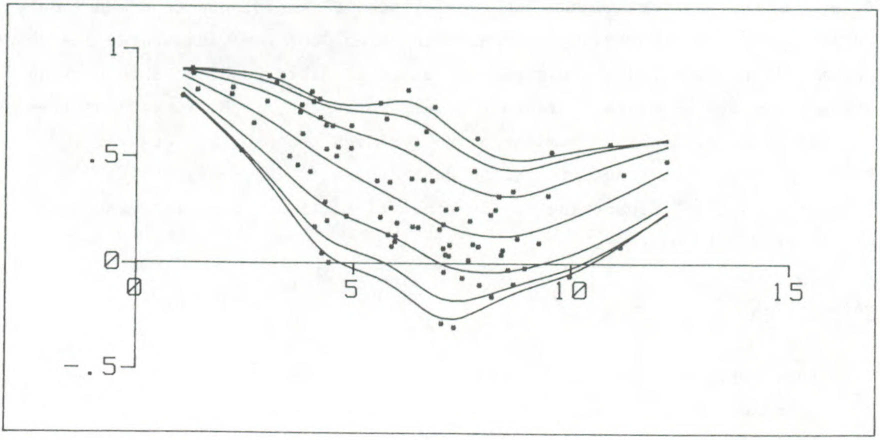
Om een indicatie te krijgen van de invloed van de afmetingen van de steekproef zijn de berekeningen herhaald met een deel van de gegevens. In figuur 4.9 zijn de resultaten getekend die bereikt worden met 20% van de gegevens uit figuur 4.7, aselect gekozen. De overeenkomst tussen beide figuren blijkt groot te zijn.



Figuur 4.7 Asymmetrische smoothing van correlatiecoëfficiënten tegen afstand tussen meetpunten



Figuur 4.8 Asymmetrische smoothing van de gegevens uit figuur 4.7, na toepassing van Fisher's z-transformatie



Figuur 4.9 Gegevens als in figuur 4.7; 20% aselect gekozen

5. Discussie

Ongelijke weging van positieve en negatieve residuen in de som van kwadraten is een nieuwe techniek die goede diensten kan bewijzen bij het onderzoeken van de spreiding van gegevensmateriaal. In het bovenstaande zijn enkele toepassingen op bivariate problemen beschreven. De nadruk lag daar op het exploratieve gebruik, met name voor het onderzoeken van strooiingsdiagrammen.
Een groot voordeel van asymmetrische weging is, dat men voor de implementatie gebruik kan maken van bestaande regressieprogramma's, door middel van iteratieve weging. Hoewel pathologische situaties geconstrueerd kunnen worden, bleken in de praktijk meer dan tien iteraties niet nodig te zijn voor volledige convergentie.
Voor het onderzoeken van strooiingsdiagrammen werd gebruik gemaakt van een eenvoudige maar effectieve discrete smoothing. In de gepresenteerde vorm is het algoritme niet robuust, zoals de door Cleveland en McGill (1984) gebruikte "robust nearest neighbour smoothing". Dat zou ook niet zinvol zijn, omdat we juist de verticale spreiding in de puntenwolk willen onderzoeken. Wenst men echter alleen de "as" van de puntenwolk te benadrukken, met gereduceerde invloed van extreme waarnemingen, dan kan men in plaats van asymmetrische kleinste kwadraten gebruik maken van adaptieve gewichten (Eilers, 1986), of van robuste invloedsfunkties (Huber, 1979). Het enige dat verandert is de berekening van gewichten uit de residuen.

De toepassing van asymmetrisch gewogen kleinste kwadraten hoeft niet beperkt te worden tot exploratie. Men kan theoretische en empirische ééndimensionale verdelingen voor een variabele y karakteriseren door $\bar{y}(p)$, de waarde die de asymmetrisch gewogen som van kwadraten minimaal maakt. Voor de asymmetrisch gewogen som van absolute residuen (Koenker en Basset, 1978) is $\bar{y}(p)$ de (empirische) quantiel. Voor de kwadraten van residuen zou men van "quadrielen" kunnen spreken. Naar analogie van de door Parzen (1979) voorgestelde "density-quantile" benadering van de statistiek, lijken veel interessante mogelijkheden te bestaan.

Referenties

- Adby P.R., Dempster M.A.H. (1974), Introduction to Optimization Methods, Chapman and Hall
- Cleveland W.S., McGill R. (1984), The Many Faces of a Scatterplot, Journal of the American Statistical Society 79, 807-822
- Eilers P.H.C. (1986), Adaptieve Gewichten, een Exploratieve Techniek voor Uitbijters en Mengsels van Regressiemodellen, Kwantitatieve Methoden, te verschijnen
- Henderson R. (1924), A New Method of Graduation, Transactions of the Actuarial Society of America 25, 29-40
- Huber P.J. (1979), Robust Smoothing, in: Robustness in Statistics, R.L. Launer, G.N. Wilxinson (eds.), Academic Press
- Kendall M.G., Stuart A. (1977) The advanced Theory of Statistics Vol 1, Hafner
- Koenker R, Basset G. (1978), Regression Quantiles, Econometrica 46, 33-50
- Parzen E. (1979), A Density Quantile Function Perspective on Robust Estimation, in: Robustness in Statistics, R.L. Launer, G.N. Wilkinson (eds.), Academic Press
- Phillips G.M., Taylor P.J. (1973), Theory and Application of Numeric Analysis, Academic Press
- Wellington J.F., Narula S.C. (1984), An Algorithm for Regression Quantiles, Communications in Statistics, Simulation and Computation 13, 683-704
- Whittaker E.T. (1923), On a New Method of Graduation, Proceedings of the Edinburgh Mathematical Society 41, 63-75

We beschouwen het algemene geval van n-dimensionale asymmetrisch gewogen lineaire regressie. De doelfunctie is:

$$S = \sum_{i=1}^{m} w_i \left(y_i - \sum_{j=1}^{n} x_{ij}a_j\right)^2 \qquad \text{(A.1)}$$

met
$$w_i = 1 - p \text{ als } \hat{y}_i = \sum_{j=1}^{n} x_{ij}a_j < y_i \qquad \text{(A.2)}$$

$$w_i = p \quad \text{ als } \hat{y}_i = \sum_{j=1}^{n} x_{ij}a_j > y_i$$

en
$$0 < p < 1 \qquad \text{(A.3)}$$

De gradient $\partial S/\partial a_k$ wordt gegeven door:

$$g_k = \frac{\partial S}{\partial a_k} = 2 \sum_{i=1}^{m} w_i \left(y_i - \sum_{j=1}^{n} x_{ij}a_j\right) (- x_{ik}) \qquad \text{(A.4)}$$

De matrix H van tweede partiële afgeleiden bevat de elementen

$$h_{kl} = \frac{\partial^2 S}{\partial a_k \partial a_l} = 2 \sum_{i=1}^{m} \frac{\partial w_i}{\partial a_l} \left(y_i - \sum_{j=1}^{n} x_{ij}a_j\right) (x_{ik}) + 2 \sum_{i=1}^{m} w_i x_{ik} x_{il} \qquad \text{(A.5)}$$

De eerste term in de rechterzijde van (A.5) is altijd nul. Er zijn namelijk twee mogelijkheden: y_i is al dan niet gelijk aan $\hat{y}_i$. In het eerste geval volgt het resultaat direct. In het tweede geval zal een infinitesimale verandering van a_l geen invloed hebben op het teken van $y_i - \hat{y}_i$ en dus ook niet op w_i. Bijgevolg geldt dan $\partial w_i/\partial a_l = 0$. Er volgt dus:

$$h_{kl} = \frac{\partial^2 S}{\partial a_k \partial a_l} = 2 \sum_{i=1}^{m} w_i x_{ik} x_{il} \qquad \text{(A.6)}$$

Als de matrix X geen afhankelijke rijen heeft en alle gewichten w_i in (A.6) positief zijn - en dat zijn ze - zal H positief definiet zijn (een bewijs via de singuliere decompositie van X is gemakkelijk te geven).
Door zijn definitie is S een convexe functie. De matrix van tweede afgeleiden is altijd positief definiet. De Newton-Raphson methode (zie bijvoorbeeld hoofdstuk 3 van Adby en Dempster, 1974) is dan toepasbaar.

In matrix-notitie zijn (A.4) en (A.6) te schrijven als:

$$g = 2\ (X^TWXa - X^TWy) \quad ; \quad H = 2X^TWX \qquad (A.7)$$

Hier is W een diagonaal-matrix met de gewichten w_i op de hoofddiagonaal. Om de gradiënt dichter bij nul te brengen berekent de Newton-Raphson methode een correctie Δa zodanig dat $g + H\Delta a$ nul wordt. Dat betekent:

$$g + H\Delta a = 2\ (X^TWXa - X^TWy + X^TWXa) = 0 \qquad (A.8)$$

Hieruit volgt:

$$X^TWX(a+\Delta a) = X^TWy \qquad (A.9)$$

Dit zijn exact de normaalvergelijkingen voor gewogen regressie, met gewichten w_i, van y op X. Het blijkt dus dat het toepassen van iteratief gewogen regressie neerkomt op het toepassen van de Newton-Raphson methode.
Dit hoeft niet automatisch te betekenen dat convergentie altijd snel plaatsvindt. Voor een ééndimensionaal probleem (regressie door de oorsprong) kan men gemakkelijk voorbeelden construeren die het aantal iteraties vrijwel gelijk maken aan het aantal datapunten, voor bepaalde waarden van p. Bij het toepassen van asymmetrisch gewogen kleinste kwadraten op allerlei gegevens is tot nu toe echter niet één geval waargenomen waarin meer dan 10 iteraties nodig waren.
Tot slot een opmerking over de conditie van de matrix H. Weliswaar zal deze matrix wiskundig gezien altijd positief definiet zijn. Bij waarden van p dicht bij 0 of 1 kan echter de numerieke conditie slecht worden. Afrondingen bij de inversie van H kunnen mogelijk problemen veroorzaken, die leiden tot oscilleren van de oplossing. In de praktijk is dit tot nu toe geen probleem gebleken.

Ontvangen: 28-08-86
Geaccepteerd: 21-10-86