

Robuuste regressie met behulp van de L1 norm

Citation for published version (APA):

Jansen, F. J. (1988). *Robuuste regressie met behulp van de L1 norm*. (Computing centre note; Vol. 39). Technische Universiteit Eindhoven.

Document status and date:

Gepubliceerd: 01/01/1988

Document Version:

Uitgevers PDF, ook bekend als Version of Record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

L. J. Jansen
Technische Universiteit
Eindhoven

Eindhoven University of Technology
Computing Centre Note 39

**Robuuste regressie
met behulp van de L1 norm**

F. J. Jansen

Stage Statistiek

Docent : Prof. dr. R. Doornbos
Begeleider : Dr. J.B. Dijkstra

februari 1988

Inhoud

	pagina
1. Inleiding	3
2. Methode	4
2.1 Inleiding	4
2.2 De Simplex-methode	8
2.3 De modificatie van Barrodale en Roberts	12
2.4 Implementatie van het Barrodale-Roberts algorithme	19
3. Enige eigenschappen van de L_1 -schatting	21
4. Een simulatie-onderzoek	24
5. Literatuur	26
 Bijlage : de procedure LAVREG	 27

1. Inleiding

Voor het schatten van de onbekende vector van parameters β in het meervoudige regressie model $y = X\beta + \epsilon$ domineert het kleinste kwadraten criterium de statistische literatuur al lange tijd. In de procedure-bibliotheek van het rekencentrum is een procedure "MULTIPLE REGRESSION" (zie [1]) aanwezig, waarmee men de kleinste kwadraten schatter (LS-schatter) kan berekenen.

Deze schattingsmethode is echter niet goed bestand tegen uitschieters in de observaties van de afhankelijke variabele y . Eén enkele uitschieter kan de LS-schatter sterk beïnvloeden. De procedure-bibliotheek bevat ook de procedure "ROBUST MULTIPLE REGRESSION" (zie [2]). Deze regressie methode is redelijk bestand tegen uitschieters in y . Voor deze iteratieve methode wordt door Holland en Welsh [3] als startwaarde de L_1 -schatter aangeraden. Een algoritme om deze schatter te berekenen is nog niet in de procedure-bibliotheek geïmplementeerd.

Deze notitie handelt over de L_1 -schatter, welke door veel auteurs als een robuust alternatief wordt beschouwd van de LS-schatter. De L_1 -schatter is redelijk tegen uitschieters in y bestand.

In tegenstelling tot de LS-schatter bestaat er voor de L_1 -schatter geen expliciete formule. Men kan de L_1 -schatter berekenen door het probleem als een Lineair Programmerings probleem te formuleren en dan technieken uit de Lineaire Programmering te gebruiken. Dit staat beschreven in een artikel van Barrodale en Roberts [4].

In deze notitie zal ik een algoritme beschrijven dat op deze manier de L_1 -schatter berekend. Tevens worden resultaten gegeven van enkele kleine simulaties waarbij de L_1 -schatter met behulp van dit algoritme werd berekend.

2. Methode

2.1 Inleiding

We hebben het volgende model :

$$y_i = \beta_0 + \beta_1 x_{i,1} + \dots + \beta_{p-1} x_{i,p-1} + \epsilon_i.$$

Hierbij is i een index voor het aantal waarnemingen en p is het aantal te schatten parameters. ($i = 1, \dots, n$ en $p < n$)

Definiëer :

$$e_i := y_i - \beta_0 - \beta_1 x_{i,1} - \dots - \beta_{p-1} x_{i,p-1}, \quad i = 1, \dots, n$$

$$y := (y_1, \dots, y_n)^T$$

$$\beta := (\beta_0, \dots, \beta_{p-1})^T$$

$$e := (e_1, \dots, e_n)^T$$

Verder definiëren we :

$$X := \begin{bmatrix} 1 & x_{1,1} & \dots & x_{1,p-1} \\ \vdots & \vdots & \dots & \vdots \\ 1 & x_{n,1} & \dots & x_{n,p-1} \end{bmatrix}$$

De kleinste kwadraten schatter is de oplossing van het volgende probleem :

$$\min_{\beta} \sum_{i=1}^n e_i^2$$

De L_1 -schatter is de oplossing van het probleem :

$$\min_{\beta} \sum_{i=1}^n |e_i|$$

Dit laatste probleem kan als een lineair programmerings probleem (LP probleem) worden geformuleerd.

Een standaard LP probleem heeft de volgende gedaante :

$$\begin{array}{l} \min \quad c^T x \\ \text{onder de voorwaarden} \\ Ax = b, \quad x \geq 0 \end{array}$$

$$\begin{array}{ll} \text{Definiëer} & e_i^+ := \max(0, e_i), & e^+ := (e_1^+, \dots, e_n^+)^T \\ & \text{en} & e_i^- := \max(0, -e_i), & e^- := (e_1^-, \dots, e_n^-)^T. \end{array}$$

$$\text{Dan geldt : } e_i^+ + e_i^- = |e_i| \quad \text{en} \quad e_i^+ - e_i^- = e_i \quad (i=1, \dots, n)$$

Het L_1 -probleem kunnen we nu als volgt herformuleren :

$$\begin{array}{l} \min \quad \sum_{i=1}^n (e_i^+ + e_i^-) \\ \text{o.d.v.} \\ X(\gamma - \delta) + e^+ - e^- = y \quad (\text{a}) \\ e^+ \geq 0, \quad e^- \geq 0 \quad (\text{b}) \\ \gamma \geq 0, \quad \delta \geq 0. \quad (\text{c}) \end{array}$$

(a) en (c) zijn equivalent met : $X\beta + e^+ - e^- = y$,
 β onbeperkt in teken.

Iedere $(\gamma, \delta, e^+, e^-)$ die voldoet aan $X(\gamma - \delta) + e^+ - e^- = y$ heet een oplossing. Als ook nog geldt $e^+ \geq 0$, $e^- \geq 0$, $\gamma \geq 0$, en $\delta \geq 0$, dan is het een toegelaten oplossing.

Definiëer :

$$A := \begin{bmatrix} X & -X & I & -I \end{bmatrix}, \quad \text{een matrix met afmetingen } n \times (2p+2n) \\ \text{en kolommen } a_j.$$

$$w := \begin{bmatrix} \gamma \\ \delta \\ e^+ \\ e^- \end{bmatrix}, \quad c := \begin{bmatrix} 0 \\ 0 \\ u \\ u \end{bmatrix}, \quad \text{u is een vector van n enen.}$$

Het L_1 -probleem wordt nu : $\min c^T w$
o.d.v. $Aw = y, w \geq 0$.

De verzameling van toegelaten oplossingen :

$$F := \left\{ w \mid Aw = y, w \geq 0 \right\}$$

$c^T w$ is de objectfunctie van het probleem.

Iedere $w \in F$ die $c^T w$ minimaliseert heet een optimale oplossing van het probleem.

Iedere toegelaten oplossing w heet een toegelaten basisoplossing als de kolommen van A , die corresponderen met de niet-nul componenten van w , een lineair onafhankelijke verzameling in \mathbb{R}^n vormen.

Een punt $w \in F$ heet een extreem punt van F als w niet te schrijven is als een convexe combinatie van twee andere punten uit F , d.w.z er bestaan geen punten $w_1 \in F$ en $w_2 \in F$, $w_1 \neq w_2$, en geen getal λ , met $0 < \lambda < 1$, zodat $w = \lambda w_1 + (1-\lambda)w_2$.

Men kan bewijzen (zie [5]) :

$w \in F$ is een extreem punt van F dan en slechts dan als de kolommen van A die corresponderen met niet-nul elementen van w een lineair onafhankelijke verzameling vormen in \mathbb{R}^n .

Hier volgt ook uit dat F slechts eindig veel extreme punten heeft, immers er zijn slechts eindig veel manieren om een onafhankelijk stelsel uit de kolommen van A te kiezen.

Ook geldt dat F minimaal één extreem punt heeft.

De volgende twee resultaten komen ook uit de theorie van de lineaire programmering :

- Als F begrensd is, dan neemt $c^T w$ zijn minimum aan in tenminste één van de extreme punten van F .
- Als F onbegrensd is én als $c^T w$ zijn minimum op F bereikt, dan is tenminste één van de extreme punten van F een optimale oplossing.

Hier kunnen we uit concluderen dat bij het zoeken naar een optimale oplossing van het L_1 -probleem volstaan kan worden met het beschouwen van optimale basisoplossingen, dit zijn immers juist de extreme punten van F .

Een methode voor het oplossen van lineaire programmerings problemen is de Simplex-methode. Deze iteratieve methode bepaalt in iedere iteratie een nieuwe toegelaten basisoplossing (d.i. extreem punt van F), zodanig dat in iedere iteratie de waarde van de objectfunctie verminderd wordt.

Na een eindig aantal stappen eindigt deze methode ófwel omdat het probleem strijdig is (er kan niet aan de voorwaarden worden voldaan) , ófwel omdat er een, al dan niet eindige, optimale oplossing is gevonden.

$$\text{Het } L_1\text{-probleem : } \min \sum_{i=1}^n (e_i^+ + e_i^-)$$

$$\begin{aligned} \text{o.d.v.} \quad & X(\gamma - \delta) + e^+ - e^- = y \\ & e^+ \geq 0, \quad e^- \geq 0, \\ & \gamma \geq 0, \quad \delta \geq 0. \end{aligned}$$

bezit een eindige optimale oplossing; we bepalen immers het minimum van een continue functie over een gesloten convexe verzameling en die functie wordt van beneden begrensd door 0.

Rechtstreekse toepassing van de Simplex-methode op het L_1 -probleem is niet verstandig omdat het L_1 -probleem, geformuleerd als LP-probleem, veel (d.i. $2n+2p$) variabelen heeft.

Barrodale en Roberts [4] hebben de Simplex-methode zodanig aangepast dat deze wel bruikbaar is voor oplossing van het L_1 -probleem.

In de volgende paragraaf zal in globale lijnen de Simplex-methode worden besproken en in de daarop volgende paragraaf de zojuist genoemde modificatie.

2.2 De Simplex-methode

Onder een basis van de kolommen a_j , $j=1, \dots, 2n+2p$, van de matrix A verstaan we een deelverzameling $(a_{j_1}, \dots, a_{j_n})$ van deze kolommen met de eigenschappen:¹

1. de basiskolommen a_{j_1}, \dots, a_{j_n} zijn onderling onafhankelijk,
2. alle kolommen van A zijn lineair in de basiskolommen uit te drukken.

Bij het lineaire systeem $Aw = y$, $w \geq 0$ noemt men een basis B van A , $B = (a_{j_1}, \dots, a_{j_n})$, toegelaten als het rechterlid y een niet-negatieve combinatie is van de basiskolommen.

Een basis B van A induceert een splitsing :

$$\begin{aligned} A &= (B \mid N) \\ w^T &= (w_b^T \mid w_n^T) \\ c^T &= (c_b^T \mid c_n^T) \end{aligned}$$

w_b is dus de vector van variabelen w_j die corresponderen met de basiskolommen in B . B is toegelaten betekent : $Bw_b = y$, $w_b \geq 0$. De vector $w^T := (w_b^T, 0^T)$ is een toegelaten basisoplossing, immers $Aw = (B \mid N)w = Bw_b + N0 = y + 0 = y$ en $w \geq 0$

Uit de lineaire programmering komt het volgende optimaliteitscriterium :

Als B een toegelaten basis voor het probleem

$$\min \left\{ c^T w \mid Aw = y, w \geq 0 \right\}, \text{ dan is de bijbehorende}$$

basisoplossing $w(B) := \begin{bmatrix} B^{-1}y \\ 0 \end{bmatrix}$ optimaal als geldt :

$$d^T(B) := c_b^T B^{-1}A - c^T \leq 0$$

Een bewijs hiervan luidt als volgt :

laat w een willekeurige toegelaten oplossing zijn.

Uit $Aw = Bw_b + Nw_n = y$, $w \geq 0$ volgt :

¹ Vanaf nu nemen we aan dat het aantal kolommen van B gelijk is aan n , het aantal rijen van A . Dus bestaat de inverse van B .

$$\begin{aligned}
 Z(w) &:= c^T w = c_b^T w_b + c_n^T w_n = c_b^T B^{-1}(y - Nw_n) + c_n^T w_n = \\
 &= c_b^T B^{-1}y - (c_b^T B^{-1}N - c_n^T)w_n \geq c_b^T B^{-1}y, \\
 \text{immers } 0 &\leq c_b^T B^{-1}A - c^T = c_b^T (I | B^{-1}N) - c^T = (0 | c_b^T B^{-1}N - c_n^T). \quad \square
 \end{aligned}$$

Zij B een toegelaten basis en $\begin{bmatrix} w_b \\ 0 \end{bmatrix}$ de bijbehorende toegelaten basisoplossing van het probleem, dus $Bw_b = y$, $w_b \geq 0$.

Laat a_s een niet-basis kolom zijn.

De kolommen van B vormen een basis voor de kolommen van A , dus geldt $Bv_b + a_s = 0$ voor zekere $v_b \in \mathbb{R}^m$.

Voor iedere $\Delta \geq 0$ geldt :

$$y = Bw_b = Bw_b + \Delta(Bv_b + a_s) = B(w_b + \Delta v_b) + \Delta a_s$$

ofwel : als $w(\Delta) := \begin{bmatrix} w_b \\ 0 \end{bmatrix} + \Delta \begin{bmatrix} v_b \\ e_s \end{bmatrix}$, waarbij e_s de s -de eenheidsvector is in \mathbb{R}^{2n-p} , dan geldt : $(B | N)w(\Delta) = y$.

Wil $w(\Delta)$ toegelaten zijn dan moet ook nog gelden $w(\Delta) \geq 0$.

$w(\Delta)$ is dus toegelaten voor

$$\begin{aligned}
 0 \leq \Delta & \quad , \text{ als } v_b \geq 0 \\
 0 \leq \Delta \leq \Delta_{\max} & := \min \left\{ \frac{(w_b)_i}{(-v_b)_i} \mid (v_b)_i < 0 \right\} \quad , \text{ als } \exists_i (v_b)_i < 0 \\
 & = \min \left\{ \frac{(B^{-1}y)_i}{(B^{-1}a_s)_i} \mid (B^{-1}a_s)_i > 0 \right\}
 \end{aligned}$$

Laat $B = (a_{j_1}, \dots, a_{j_n})$ weer een toegelaten basis zijn. We willen een nieuwe toegelaten basis kiezen, en wel zó dat de waarde van de objectfunctie bij deze nieuwe basis minder is. Stel dat we de niet-basis kolom a_s aan de basis B willen toevoegen en de basis kolom a_r , $r \in \{j_1, \dots, j_n\}$ uit de basis willen verwijderen.

Er geldt $a_s = \sum_{i=1}^n \alpha_i a_{j_i}$, ofwel $B^{-1}a_s = \alpha$
 voor een zekere vector $\alpha = (\alpha_1, \dots, \alpha_n)^T$

Kies nu de niet-basis kolom a_s zo dat :

1. er is een index i met $(B^{-1}a_s)_i > 0$
2. $d_s(B) := c_b^T B^{-1}a_s - c_s > 0$

De vector $d(B) := c_b^T B^{-1}A - c$ noemt men de gereduceerde kosten.
 Vorm nu de basis B_1 welke uit B ontstaat door de basiskolom a_r
 te vervangen door de niet-basis kolom a_s .

Dan geldt :

$w(\Delta_{\max}) := \begin{bmatrix} w_b \\ 0 \end{bmatrix} + \Delta_{\max} \begin{bmatrix} -B^{-1}a_s \\ e_s \end{bmatrix}$ is een basisoplossing met

$$\begin{aligned} Z(w(\Delta_{\max})) &= c^T w(\Delta_{\max}) = c_b^T w_b - \Delta_{\max} (c_b^T B^{-1}a_s - c_s) = \\ &= c_b^T w_b - \Delta_{\max} d_s(B) < c_b^T w_b. \end{aligned}$$

Uitgaande van een startoplossing wordt op deze manier in iedere iteratie een nieuwe basisoplossing bepaald met een lagere functiewaarde. Aangezien er maar eindig veel basisoplossingen zijn en omdat er een eindige oplossing bestaat bij het L_1 -probleem wordt via de Simplex methode na een eindig aantal iteraties een optimale oplossing gevonden.

De Simplex methode werkt volgens het onderstaande algoritme :

stap 1

kies een startoplossing, dit moet een toegelaten basisoplossing zijn.

stap 2

kies een vector a_s die niet in de basis zit, zo dat

$$d_s(B) = c_b^T B^{-1} a_s - c_s = \max \left\{ d_j(B) \mid j \text{ niet in de basis} \right\}.$$

Als $d_s(B) \leq 0$ ga dan naar stap 5.

Indien dit niet het geval is ga dan naar stap 3.

stap 3

$$\text{Kies } r \text{ zo dat } \Delta_r := \frac{y_r}{a_{rs}} = \min \left\{ \frac{y_i}{a_{is}} \mid a_{is} > 0 \right\}.$$

Ga naar stap 4

stap 4

Verander de basis B door de basis kolom a_r te vervangen door de niet-basis kolom a_s .

Vervang de matrix A door de matrix $B^{-1}A$ en het rechterlid y door $B^{-1}y$.

Dus : $A \rightarrow B^{-1}A$ en $y \rightarrow B^{-1}y$.

Ga naar stap 2.

stap 5

Stop. De huidige basisoplossing is optimaal.

Het opnieuw berekenen in iedere iteratie van de vectoren d en y en de matrix A kan op een slimme manier. Het is dus niet nodig iedere keer opnieuw de inverse van matrix B te berekenen.

Men kan op eenvoudige wijze de matrix $B^{-1}A$ afleiden uit de waarde van deze matrix in een vorige iteratie.

De vectoren d en y en de matrix A worden bijgehouden in een zogenaamd Simplex-tableau.² In de volgende paragraaf, bij de modificatie van Barrodale en Roberts, kom ik hier nader op terug.

²Voor een uitvoerige behandeling van de Simplex-methode en het Simplex-tableau zie : P. van Beek en T.H.B. Hendriks, Optimaliseringstechnieken : principes en toepassingen, Bohn, Scheltema & Holland, 1983.

2.3 De modificatie van Barrodale en Roberts

Barrodale en Roberts hebben de Simplex methode zodanig aangepast dat deze te gebruiken is voor oplossing van het L_1 -probleem met behulp van een LP-formulering :

$$\begin{aligned} \min \quad & \sum_{i=1}^n (e_i^+ + e_i^-) \\ \text{o.d.v.} \quad & X(\gamma - \delta) + e^+ - e^- = y \\ & e^+ \geq 0, e^- \geq 0, \\ & \gamma \geq 0, \delta \geq 0. \end{aligned}$$

Bij het L_1 -probleem kan men de volgende toegelaten startoplossing kiezen :

$$e_i^+ := \begin{cases} y_i & , \text{als } y_i > 0 \\ 0 & , \text{anders} \end{cases} \quad , \quad e_i^- := \begin{cases} -y_i & , \text{als } y_i < 0 \\ 0 & , \text{anders} \end{cases} \quad , \quad i=1, \dots, n$$

$$\gamma_i = 0, \quad \delta_i = 0 \quad , \quad i=0, \dots, p-1$$

Men krijgt het volgende Simplex starttableau :

0	0	...	0	0	0	...	0	1	..	1	1	..	1	
γ_0	γ_1	...	γ_{p-1}	δ_0	δ_1	...	δ_{p-1}	e_1^+	..	e_n^+	e_1^-	..	e_n^-	
1	$x_{1,1}$...	$x_{1,p-1}$	-1	$-x_{1,1}$...	$-x_{1,p-1}$	1			-1			y_1
:	:		:	:	:		:			
1	$x_{n,1}$...	$x_{n,p-1}$	-1	$-x_{n,1}$...	$-x_{n,p-1}$			1			-1	y_n
d_0	d_1		d_{p-1}	d_p		...	d_{2p-2}	0	0		-2	-2		Z

tabel 1 ³

In iedere iteratie moet het Simplex tableau worden bijgewerkt.

³Op blz. 13 staat de definitie van de "d-rij".

In een meer compacte notatie ziet het Simplex tableau er als volgt uit :

0^T	0^T	u^T	u^T	← kosten
γ^T	δ^T	$(e^+)^T$	$(e^-)^T$	← variabelen
A				y ← beperkingen
d^T				Z ← gereduceerde kosten

tabel 2

De bovenste rij is de rij van de kosten c_j . De onderste rij is de "d-rij" ofwel de rij van de gereduceerde kosten.

Hiervan luidt de formule : $d_j(B) := c_b^T B^{-1} a_j - c_j$

Hierin is B de huidige basis, c_b de kostenvector van de met deze basis corresponderende basisvariabelen en a_j de j-de kolom van de matrix A. Z is de waarde van de objectfunctie.

Als een variabele deel uitmaakt van de basis dan geldt dat het corresponderende element uit de d-rij gelijk is aan 0. Tevens volgt dat de bij dit element corresponderende kolom in het simplex tableau gelijk is aan een eenheidsvector in \mathbb{R}^n .

Voor het gemak is aangenomen dat voor alle y_i geldt dat $y_i > 0$. Als geldt $y_r < 0$ dan moet rij r in zijn geheel met -1 worden vermenigvuldigd. In plaats van e_i^+ komt dan e_i^- in de basis.

In het vervolg kan met γ , δ , e^+ , of e^- zowel de variabele als de corresponderende kolom van de matrix A worden bedoeld. Uit de context wordt wel duidelijk wat het geval is.

In het starttableau van de Simplexmethode (tabel 1) is het volgende op te merken :

1. Voor de kolommen γ_j en δ_j ($j=0, \dots, p-1$) geldt $\gamma_j = -\delta_j$.
Voor de kolommen e_j^+ en e_j^- ($j=1, \dots, n$) geldt $e_j^+ = -e_j^-$.

Als we alle kolommen met B^{-1} vermenigvuldigen (d.w.z. een andere basis kiezen) dan blijven deze twee relaties uiteraard gelden.

Voor de variabelen e_j^+ en e_j^- geldt op grond van hun definitie (blz. 5) $e_j^+ e_j^- = 0$. Hetzelfde geldt voor het produkt $\delta_j \gamma_j$.

2. De gereduceerde kosten ($d_j(B)$) van de variabele γ_j zijn gelijk aan :

$$\begin{aligned} c_b^T B^{-1} \gamma_j - c(\gamma_j) &= c_b^T I \gamma_j - 0 \\ &= \sum_{i=1}^n \gamma_{ij} = \sum_{i=1}^n x_{ij} \end{aligned}$$

Evenzo zijn de gereduceerde kosten van de variabele δ_j gelijk aan :

$$\begin{aligned} c_b^T B^{-1} \delta_j - c(\delta_j) &= c_b^T I \delta_j - 0 \\ &= \sum_{i=1}^n \delta_{ij} = \sum_{i=1}^n -x_{ij}. \end{aligned}$$

De som van de gereduceerde kosten van γ_j en δ_j is dus gelijk aan 0.

Deze relatie blijft bij iedere basis B gelden :

$$\begin{aligned} (c_b^T B^{-1} \gamma_j - c(\gamma_j)) + (c_b^T B^{-1} \delta_j - c(\delta_j)) &= \\ c_b^T B^{-1} (\gamma_j + \delta_j) - c(\gamma_j) - c(\delta_j) &= c_b^T B^{-1} (\gamma_j + \delta_j) = 0. \end{aligned}$$

Evenzo geldt voor de som van de gereduceerde kosten van e_j^+ en e_j^- dat deze gelijk is aan -2.

Ook deze relatie blijft bij iedere basis B gelden.

Uit deze twee opmerkingen kunnen we het volgende afleiden :

1. We hoeven niet het volledige Simplex-tableau bij te houden zoals dat in tabel 1 staat af te lezen. Aangezien de kolommen van de basis-variabelen altijd eenheidsvectoren zijn hoeven we deze dus niet expliciet op te slaan. Ook hoeven we niet zowel kolom $B^{-1}e_j^+$ als kolom $B^{-1}e_j^-$ op te slaan; ze zijn immers elkaars tegengestelde. Hetzelfde geldt voor $B^{-1}\gamma_j$ en $B^{-1}\delta_j$.

2. Als we de gereduceerde kosten van variabele e_j^+ kennen, kennen we ook de gereduceerde kosten van variabele e_j^- . Hetzelfde geldt voor de variabelen γ_j en δ_j .

Op deze manier hoeven we in plaats van $2n+2p$ nog maar de kolommen van p variabelen op te slaan.

We slaan kolom γ_j op als zowel γ_j als δ_j niet in de basis zit.

We slaan kolom e_j^+ op als zowel e_j^+ als e_j^- niet in de basis zit.

Het starttableau bevat dus juist de kolommen $\gamma_0, \dots, \gamma_{p-1}$.

Het is nu mogelijk om de Simplex iteraties uit te voeren in een werkarray van afmetingen $(n+2) \times (p+2)$.

Dit werkarray bevat behalve de data uit tabel 1 ook nog labels voor de basis- en de niet-basis vectoren.

Het Barrodale-Roberts algoritme bestaat uit twee fasen.

In fase 1 wordt de keuze, welke variabele tot de basis toetreedt, beperkt tot γ_j en δ_j .

De variabele met de grootste niet-negatieve gereduceerde kosten wordt gekozen om de basis te betreden. De variabele die de basis moet verlaten wordt gekozen uit de variabelen e_j^+ en e_j^- en wel dié variabele die een maximale vermindering van de object-functie geeft.

Aan het einde van fase 1 is de rang k ($\leq p$) van de matrix X bepaald door het totale aantal vectoren γ_j, δ_j in de basis. Angezien k van de vectoren e_j^+ (e_j^-) uit de basis verwijderd zijn (en dus waarde 0 hebben !) representeert het Simplex-tableau op dat moment een parameterschatting die in tenminste k datapunten interpoleert. Het is ook mogelijk dat de schatting in meer dan k punten interpoleert.

In fase 2 mogen variabelen γ_j, δ_j de basis niet verlaten. In deze fase worden alleen niet-basis variabelen e_j^+ of e_j^- met basis variabelen e_i^+ of e_i^- verwisseld.

De variabele met de grootste niet-negatieve gereduceerde kosten wordt gekozen om de basis te betreden. De variabele die de basis moet verlaten wordt gekozen uit de variabelen e_j^+ en e_j^- en

wel dié variabele die een maximale vermindering van de object-functie geeft.

Het algoritme eindigt wanneer alle gereduceerde kosten niet-positief zijn.

Bij de Simplex-methode gaat het meeste rekenwerk zitten in het bijwerken van het Simplex-tableau in iedere iteratie. Barrodale en Roberts hebben een methode ontwikkeld waardoor dit rekenwerk aanzienlijk wordt verminderd.

In iedere iteratie verlaat een basis variabele e_j^+ of e_j^- de basis. In fase 1 wordt deze vervangen door één van de γ_j of δ_j , terwijl in fase 2 de niet-basis variabelen e_j^+ of e_j^- kandidaten zijn om de basis te betreden.

Stel dat een niet-basis variabele, γ_r , is gekozen om een nieuwe basis variabele te worden. De waarde van deze variabele gaat dan vanaf nul stijgen terwijl deze variabele aan de beperkingen van het L_1 -probleem moet blijven voldoen. In de oorspronkelijke Simplex-methode zou deze variabele γ_r net zolang stijgen tot het punt waarop de eerste basisvariabele, zeg e_s^+ , gelijk aan nul wordt. Verdere stijging van γ_r zou tot gevolg hebben dat e_s^+ een niet toegestane waarde (d.i. negatief) zou krijgen. Op dit punt zou bij de normale Simplex-methode het Simplex-tableau worden bijgewerkt en zou γ_r in de plaats van e_s^+ in de basis komen.

Barrodale en Roberts hebben nu ingezien dat γ_r mogelijk verder

verhoogd kan worden en de objectfunctie $\sum_{i=1}^n (e_i^+ + e_i^-)$ verder verlaagd zonder dat het Simplex-tableau helemaal bijgewerkt hoeft te worden.

Stel dat $e_{s_1}^+$ (of $e_{s_1}^-$) de eerste basisvariabele is die nul wordt als γ_r vanaf nul wordt verhoogd tot bv. t_1 .

Er moet voldaan worden aan de beperkingen $X(\gamma-\delta) + e^+ - e^- = y$,
 $\delta \geq 0, \gamma \geq 0, e^+ \geq 0, e^- \geq 0$.

Omdat $e_{s_1}^+ e_{s_1}^- = 0$ kan γ_r tot voorbij t_1 worden verhoogd, terwijl $e_{s_1}^+$ en $e_{s_1}^-$ aan de beperkingen blijven voldoen. Dit doen we door de rollen van $e_{s_1}^+$ en $e_{s_1}^-$ te verwisselen.

Dus inplaats van dat $e_{s_1}^+$ (of $e_{s_1}^-$) negatief wordt als γ_r tot voorbij t_1 wordt verhoogd, wordt $e_{s_1}^+$ (of $e_{s_1}^-$) op de waarde nul gehouden, terwijl $e_{s_1}^-$ (of $e_{s_1}^+$) vanaf nul gaat stijgen.

Dit kan herhaald worden. Indien de objectfunctie $\sum_{i=1}^n (e_i^+ + e_i^-)$ nog steeds daalt, kan γ_r verder verhoogd worden tot zeg t_2 , $t_2 > t_1$, waar een volgende basisvariabele $e_{s_2}^+$ (of $e_{s_2}^-$) voor het eerst nul wordt. De rollen van deze twee variabelen wordt dan weer verwisseld; $e_{s_2}^+$ (of $e_{s_2}^-$) wordt op nul gehouden en $e_{s_2}^-$ (of $e_{s_2}^+$) stijgt als γ_r verder stijgt.

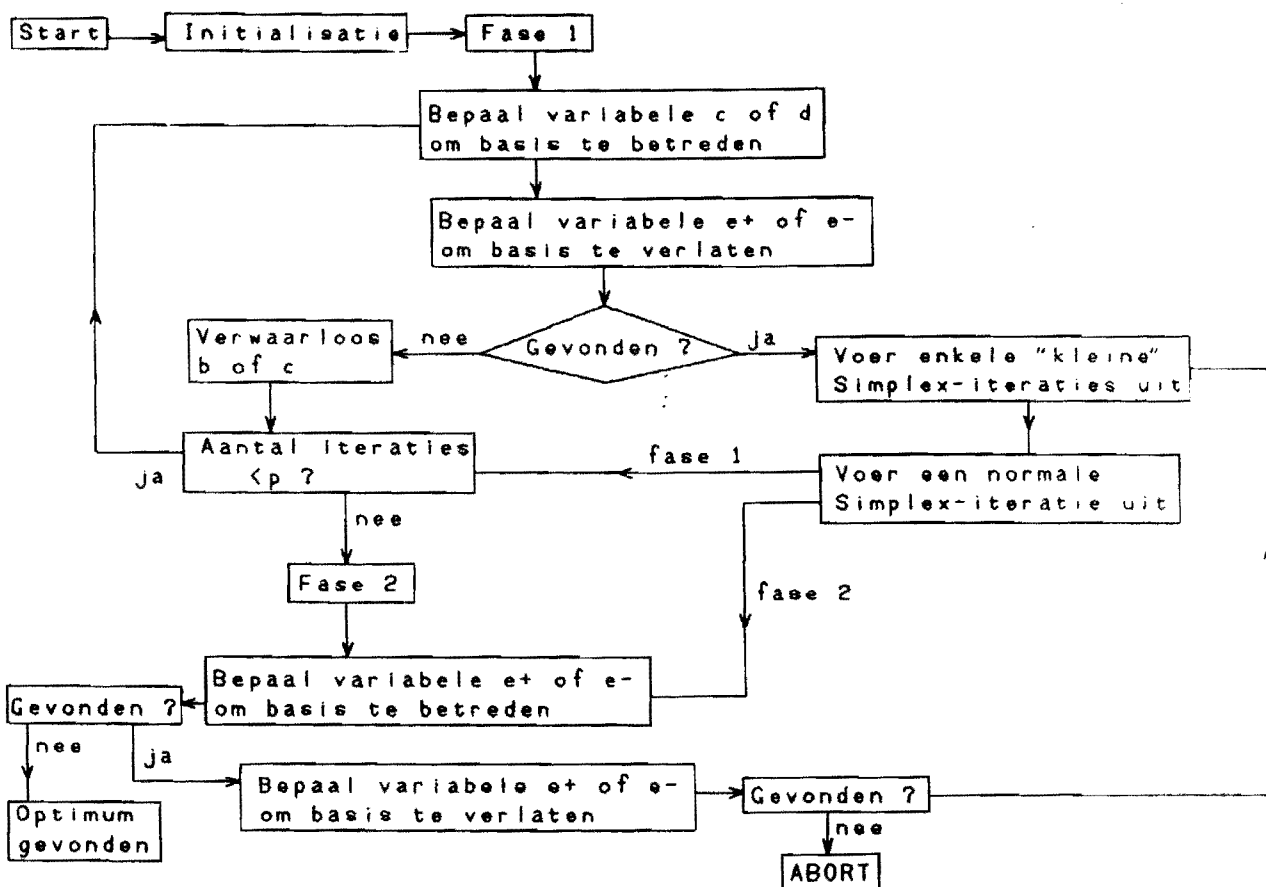
Zolang $\sum_{i=1}^n (e_i^+ + e_i^-)$ in waarde afneemt als γ_j stijgt, gaat de verwisseling van $e_{s_1}^+$ en $e_{s_1}^-$ door en hoeft het Simplex tableau niet helemaal bijgewerkt te worden. Het enige wat in het Simplex-tableau veranderd hoeft te worden is dat er een rij met -1 vermenigvuldigd hoeft te worden en dat de d-rij opnieuw berekend moet worden. Dit laatste kan met enkele eenvoudige berekeningen.

Uiteindelijk bereikt γ_r een waarde, zeg t_m , waar een basisvariabele $e_{s_m}^+$ (of $e_{s_m}^-$) voor het eerst nul wordt, terwijl

verdere stijging van γ_r er voor zou zorgen dat $\sum_{i=1}^n (e_i^+ + e_i^-)$ weer gaat stijgen. Op dit moment vervangt de variabele γ_r de basisvariabele $e_{s_m}^+$ (of $e_{s_m}^-$) in de basis en wordt het Simplex-tableau op de normale manier bijgewerkt.

De methode van Barrodale en Roberts zorgt ervoor dat verschillende Simplex-iteraties uitgevoerd kunnen worden zonder de grote hoeveelheid rekenwerk die daar normaal bijhoort. Dit alles is te danken aan de speciale structuur van het L_1 -probleem.

In figuur 1 (blz.18) staan in een stroom-diagram nog eens de stappen van het Barrodale-Roberts algorithme.



Het Barrodale-Roberts algoritme.

figuur 1

2.4 Implementatie van het Barrodale-Roberts algoritme.

Het in de vorige paragraaf beschreven algoritme van Barrodale en Roberts is door mij geïmplementeerd in de programmeertaal Pascal op het Burroughs 7900 Large System. Het algoritme is geschreven in de vorm van de procedure LAVREG. LAVREG is de afkorting van Least Absolute Values REGression. In de bijlage van deze notitie staat een voorlopige beschrijving van deze procedure.

Bij deze implementatie behoren een aantal opmerkingen :

1. In fase 1 van het algoritme kan het gebeuren dat er een variabele γ_r (of δ_r) wordt bepaald om de basis te betreden terwijl er geen geschikte basisvariabele e_s^+ of e_s^- gevonden kan worden om de basis te verlaten. Dit kan gebeuren als de rang van de matrix X kleiner is dan p, het aantal te schatten parameters. In dit geval kan γ_r verder verwaarloosd worden en hoeft deze verder niet in de berekeningen te worden meegenomen.
 In fase 2 kan dit in theorie niet gebeuren omdat er een optimale oplossing van het L_1 -probleem bestaat. In de praktijk echter kan dit geval toch optreden en wel door afrondfouten.
 Barrodale en Roberts schrijven dat dit kan gebeuren doordat de matrix X elementen bevat die in orde van grootte erg van elkaar verschillen. Mocht dit gebeuren, dan raden ze aan om de matrix X te schalen.
 Gelukkig melden ze tegelijkertijd dat deze situatie hoogst zeldzaam is en dat, zelfs als het gebeurt, het waarschijnlijk is dat de oplossing op dat moment in de buurt ligt van de optimale oplossing.
 Als deze situatie in de procedure LAVREG optreedt, dan wordt het programma met een passende mededeling afgesloten.

2. In de procedure LAVREG wordt een klein getal gebruikt waar beneden de absolute waarde van een grootte gelijk wordt beschouwd aan nul. Bij de implementatie van het algoritme voor de Burroughs heb ik voor dit getal de waarde 10^{-9} gebruikt.

3. Enige eigenschappen van de L_1 -schatter.

Stel dat $(\gamma, \delta, e^+, e^-)$ een oplossing is van het L_1 -probleem

$$\begin{aligned} \min \quad & \sum_{i=1}^n (e_i^+ + e_i^-) \\ \text{o.d.v.} \quad & X(\gamma - \delta) + e^+ - e^- = y \\ & e^+ \geq 0, e^- \geq 0, \gamma \geq 0, \delta \geq 0. \end{aligned}$$

Men kan bewijzen dat als X van rang k ($\leq p$) is, dan is er een optimale oplossing die in tenminste k datapunten interpoleert. Dit wil zeggen dat tenminste k punten residu nul hebben. Het is mogelijk dat het L_1 -probleem meerdere optimale oplossingen heeft.

Een andere eigenschap die men eenvoudig kan bewijzen is deze :

Stel dat N^+ het aantal punten in de optimale oplossing is waarvoor geldt $e_i^+ > 0$, en N^- het aantal punten waarvoor geldt $e_i^- > 0$. Dan geldt : $|N^+ - N^-| \leq p$

Deze eigenschap geldt onder de voorwaarde dat geen verzameling van $p+1$ datapunten op een hypervlak van p dimensies ligt. Voor $p=2$ betekent dit dat er geen verzameling van 3 datapunten is dat op een rechte lijn in \mathbb{R}^2 ligt.

Zoals al eerder in deze notitie werd opgemerkt, is de L_1 -schatter redelijk bestand tegen uitschieters in de afhankelijke variabele y .

Met behulp van enige theorie uit de Lineaire Programmering (post-optimale analyse) kan men redelijk eenvoudig het volgende aantonen :

Stel dat $(\gamma, \delta, e^+, e^-)$ een oplossing is van het L_1 -probleem. Laat i een datapunt zijn met $e_i^+ > 0$. Dit betekent dat het punt i een positief residu heeft (voor $p=2$ betekent dit dat het punt i "boven" de optimale aangepaste lijn ligt). Als we de waarde van de observatie y_i nu zó veranderen dat ten opzichte van de optimale L_1 -schatter blijft gelden $e_i^+ > 0$, dan heeft dit geen invloed op de L_1 -schatter.

Hetzelfde geldt voor punten met een negatief residu, dat wil zeggen punten waarvoor $e_i^- > 0$. We mogen y_i zoveel veranderen als we willen, zolang we er maar voor blijven zorgen dat $e_i^- > 0$ verandert de L_1 -schatting niet.

De genoemde eigenschappen worden geïllustreerd op blz. 23.

Het model dat hierbij hoort is het volgende : ($p=2$)

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \dots, 10$$

De data staan in de onderstaande tabel :

		y			
		grafiek A	grafiek B	grafiek C	grafiek D
x	1	1	1	-19	9
	2	7	7	7	10
	3	8	65	8	19
	4	8	8	8	21
	5	4	4	4	28
	6	13	13	13	-
	7	12.5	12.5	12.5	-
	8	9	9	9	-
	9	16	16	16	-
	10	18.5	18.5	18.5	3

tabel 3

De volgende parameter-schattingen werden gevonden :

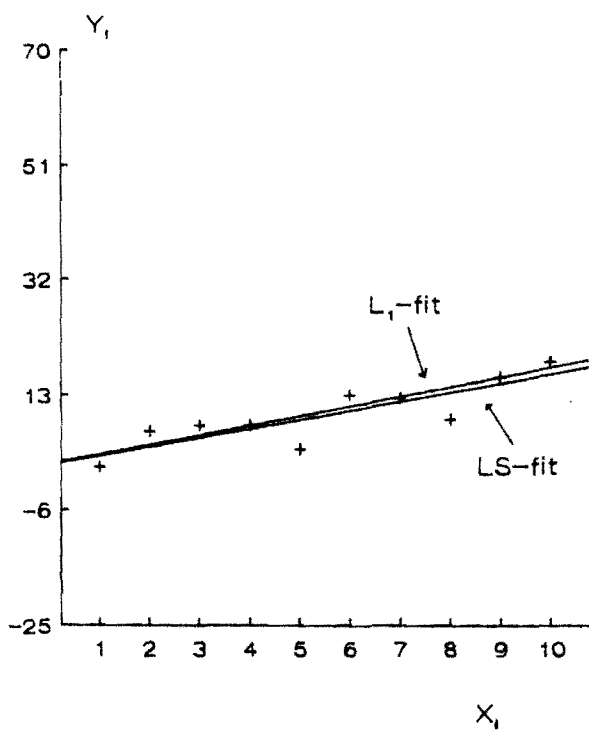
		A	B	C	D
β_0	LS	1.43	16.63	-6.57	17.87
	L_1	1.60	1.60	1.60	25.86
β_1	LS	1.50	-0.22	2.59	-0.69
	L_1	1.60	1.60	1.60	-2.29

tabel 4

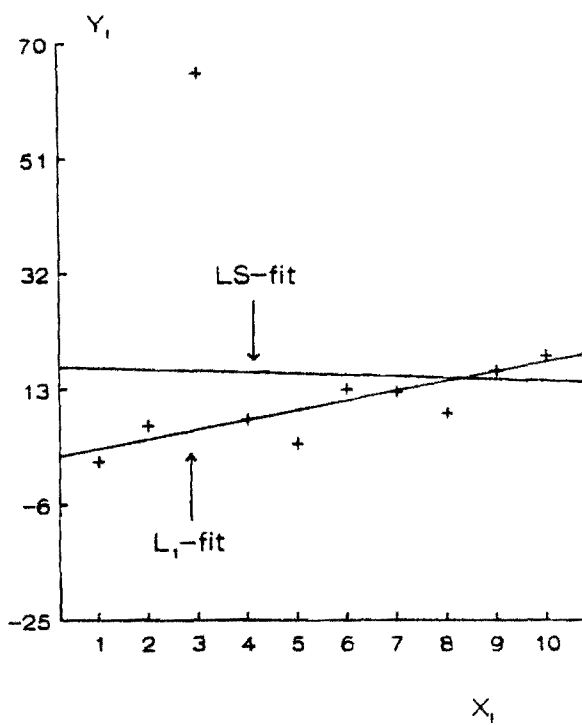
In de grafieken A, B en C is de LS-schatting verschillend, terwijl de L_1 -schatting niet verandert.

Grafiek D laat zien dat de L_1 -schatting, evenals de LS-schatting, niet goed bestand is tegen uitschieters in de onafhankelijke variabelen x_j .

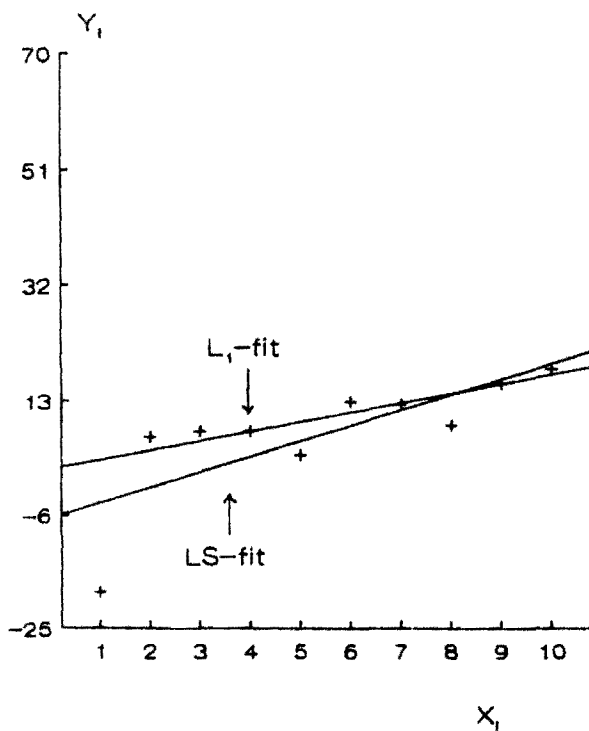
Grafiek A



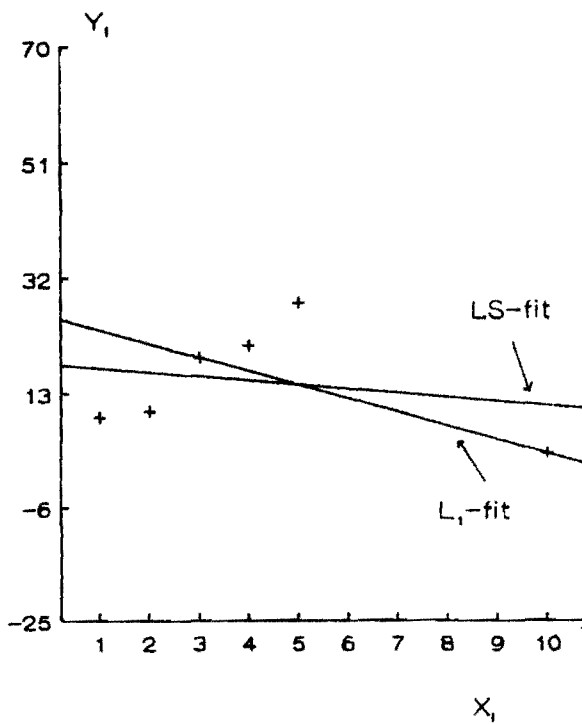
Grafiek B



Grafiek C



Grafiek D



4. Een simulatie-onderzoek.

In een simulatie-onderzoek werd het volgende model gebruikt :

$$y_i = 1 - 1.5 x_{i1} + 23.7 x_{i2} + \underline{\epsilon}_i, \quad i = 1, \dots, 1000$$

Hierbij werden x_{i1} en x_{i2} getrokken uit een uniforme verdeling over het interval (0,10).

Voor de verdelingsfunctie van $\underline{\epsilon}_i$ werden een viertal functies gebruikt :

1. Een Laplace-verdeling, met dichtheid

$$F(x) = \frac{1}{2} \exp(-|x-1|)$$

2. Een Cauchy-verdeling, met dichtheid

$$F(x) = \frac{1}{\pi} \frac{1}{1+(x-1)^2}$$

3. De standaard normale verdeling, $N(0,1)$

4. Een "vervuilde" normale verdeling :

$\underline{\epsilon}_i$ werd met kans 0.15 uit een $N(0, \sigma^2)$ -verdeling getrokken en met kans 0.85 uit een $N(0,1)$ -verdeling. Hierbij werd σ gelijk genomen aan 10.

Het aantal observaties bedroeg 1000.

Voor iedere verdelingsfunctie werd het experiment 10 keer herhaald.

Dit leverde per verdelingsfunctie voor ieder van de drie te schatten parameters 10 schattingen op.

De parameterschattingen werden zowel met het L_1 - alswel met het LS- criterium berekend.

In tabel 5 staan per verdeling voor zowel het L_1 - als het LS-criterium de gemiddelden over de 10 parameterschattingen voor ieder van de drie te schatten parameters.

Tussen haakjes staat steeds de variantie binnen de groep van 10 parameterschattingen.

verdeling van ϵ_i		echte waarde	L1-norm		LS-norm	
Laplace	β_0	1	1.0239	(0.0043)	1.0017	(0.0081)
	β_1	-1.5	-1.5011	(0.0002)	-1.5025	(0.0003)
	β_2	23.7	23.7004	(0.0001)	23.7049	(0.0001)
Cauchy	β_0	1	1.0369	(0.0098)	-0.6439	(40.5691)
	β_1	-1.5	-1.5018	(0.0005)	-1.2860	(0.5438)
	β_2	23.7	23.7005	(0.0002)	23.9268	(1.6500)
Standaard Normaal	β_0	1	0.9472	(0.0071)	0.9873	(0.0086)
	β_1	-1.5	-1.4941	(0.0001)	-1.4987	(0.0002)
	β_2	23.7	23.7044	(0.0001)	23.7022	(0.0001)
Vervuild Normaal	β_0	1	0.9206	(0.0142)	0.9240	(0.0908)
	β_1	-1.5	-1.4911	(0.0002)	-1.4885	(0.0013)
	β_2	23.7	23.7062	(0.0002)	23.7094	(0.0018)

tabel 5

In de tabel valt af te lezen dat bij Cauchy verdeelde fouten de L_1 -schatting superieur is aan de LS-schatting.

Bij de standaard normale verdeling geeft de LS-schatting betere resultaten dan de L_1 -schatting. Dit was te verwachten, immers bij normaal verdeelde fouten is de LS-schatting BLUE, dat wil zeggen dat deze schatting de kleinste variantie onder de zuivere lineaire schatters.

Het is bekend dat bij normaal verdeelde fouten de LS-schatting de maximum-likelihood schatting is.

Bij Laplace verdeelde fouten is de L_1 -schatting de maximum likelihood schatting.

Bij de Laplace verdeling en de vervuilde (eng. contaminated) normale verdeling is geen duidelijke uitspraak te doen over de superioriteit van een der beide schatters.

Wel is het zo dat in het geval van vervuilde normaal verdeelde fouten de variantie duidelijk kleiner is bij de L_1 -schatting.

Het is niet mogelijk om op grond van deze resultaten definitieve conclusies te geven.

Daarvoor zijn verdergaande simulaties nodig.

5. Literatuur

- [1] Meervoudige regressie en correlatie (1987)
Rekencentrum-informatie PP-4.3.
TUE-RC 69889
- [2] Robuuste statistische methoden (1986)
Rekencentrum-informatie PP-4.20.
THE-RC 66330
- [3] Holland, P.W. and R.E. Welsch (1977)
Robust regression using iteratively reweighted
least-squares
Communications in Statistics A6(9), 813-827
- [4] Barrodale, I and F.D.K. Roberts (1973)
An improved algorithm for discrete l_1 linear
approximation
SIAM Journal of Numerical Analysis (10), 839-848
- [5] Arthanari, T.S. and Y. Dodge (1981)
Mathematical Programming in Statistics
John Wiley, Interscience division, New York

Bijlage

LAVREG, een Procedure voor het robuust schatten van parameters bij meervoudige lineaire regressie met behulp van de L_1 -norm.
(voorlopige uitgave)

LEAST ABSOLUTE VALUE REGRESSION

Korte functiebeschrijving.

Er wordt een aantal punten $(y_i, x_{i1}, \dots, x_{in})$ beschouwd. Bij de vergelijking $y = a_m x_m + \dots + a_n x_n + b$ worden de a_j 's en de b zodanig bepaald dat de som van de absolute afstanden van de punten tot het bij de a_j 's en b behorende hypervlak minimaal is. Deze regressie is redelijk bestand tegen uitschieters in de waarnemingen van de afhankelijke variabele y .

Procedure heading.

```
PROCEDURE LAVREG(VAR X           : ARRAY2DR;
                 MI,NI,MJ,NJ     : INTEGER;
                 VAR Y,A         : ARRAY1DR;
                 VAR B           : REAL;
                 VAR RES         : ARRAY1DR;
                 VAR ITER        : INTEGER;
                 VAR UNIQUE      : BOOLEAN );
```

Typen door de gebruiker te declareren :

```
TYPE VLSTRING = STRING(80)
   ARRAY1DR = ARRAY[0..BOV] OF REAL;
   ARRAY2DR = ARRAY[0..BOV,0..NJ+4];
   ARRAY1DI = ARRAY[0..BOV] OF INTEGER;
```

Hierin is $BOV = \max(NI+3, NJ+1)$.

Bij aanroep moeten MI en MJ groter dan of gelijk aan nul zijn. Tevens moet gelden dat NI resp. NJ groter dan of gelijk is aan MI resp. MJ .

Formele parameters.

X bevat bij aanroep de observaties voor de onafhankelijke variabelen.

MI,NI kleinste resp. grootste index voor de observaties.

MJ,NJ kleinste resp. grootste index voor de variabelen in het array $X[MI..NI,MJ..NJ]$.

Y bevat bij aanroep de onservaties voor de af-

A	hankelijke variabele (loopt van MI tot en met NI). bevat na afloop de coëfficiënten (geïndiceerd van MJ tot NJ).
B	bevat na afloop de constante.
RES	bevat na afloop de residuen.
ITER	bevat na afloop het aantal uitgevoerde simplex-iteraties.
UNIQUE	bevat na afloop de waarde <u>true</u> als de coëfficiënten en de constante uniek bepaald zijn. Als de constante en/of de coëfficiënten mogelijk niet uniek bepaald zijn, krijgt UNIQUE de waarde <u>false</u> .

Methode.

Zij X de matrix van de geobserveerde x-en, uitgebreid met een kolom enen.

Deze enen worden door de procedure LAVREG zelf toegevoegd.

$$X = \begin{bmatrix} 1 & x_{mi,mj} & \cdots & x_{mi,nj} \\ \vdots & \vdots & & \vdots \\ 1 & x_{ni,mj} & \cdots & x_{ni,nj} \end{bmatrix}$$

y is de vector van de geobserveerde waarden.

De gezochte coëfficiënten a_{mj} , ..., a_{nj} en de constante b worden bepaald uit het volgende minimaliseringsprobleem :

$$\min_i \sum_{i=mi}^{ni} | y_i - \sum_{j=mj}^{nj} a_j x_{ij} + b |$$

We kunnen dit minimaliseringsprobleem als een Lineair Programmerings probleem formuleren :

$$\min_i \sum_{i=mi}^{ni} (u_i + v_i)$$

onder de voorwaarden :

$$y_i = \sum_{j=mj}^{nj} (b_j - c_j) x_{ij} + u_i - v_i, \quad i = mi, \dots, ni,$$

$$b_j \geq 0, c_j \geq 0, u_i \geq 0, v_i \geq 0.$$

Dit probleem wordt opgelost door gebruik te maken van een gemodificeerde Simplex methode. Het gebruikte algoritme is beschreven door Barrodale en Roberts [1].

Externe relaties.

Uit de procedurebibliotheek wordt aangeroepen :

ABORTED

Opmerking.

Theoretisch wordt het genoemde Lineair Programmerings probleem in een eindig aantal iteraties opgelost.

Echter, door cummulatie van afrondfouten kan het voorkomen dat er geen oplossing gevonden kan worden. Het programma wordt dan afgebroken en dit wordt met een passende medeling weergegeven.

Literatuur.

- [1] Barrodale, I. and F.D.K. Roberts (1973)
An improved algorithm for discrete l_1 linear approximation
SIAM Journal of Numerical Analysis (10), 839-848.
- [2] Jansen, F.J. (1988)
Robuuste regressie met behulp van de L_1 -norm
Computing Centre Note 39, Eindhoven University of Technology