Statistique et analyse des données

J. P. Ley

Régression linéaire robuste

Statistique et analyse des données, tome 4, nº 1 (1979), p. 55-62.
<http://www.numdam.org/item?id=SAD_1979__4_1_55_0>

Statistiques et Analyse des Données
1 - 1979 pp. 55, 62.

REGRESSION LINEAIRE ROBUSTE : UNE NOUVELLE APPROCHE DE LA REGRESSION EN NORME ℓ_1, UN NOUVEL ALGORITHME POUR CERTAINS R-ESTIMATEURS.

J. P. L E Y

I.N.R.A. - C.N.R.Z. , Laboratoire de Biométrie, 78350 JOUY EN JOSAS.

1. INTRODUCTION.

Nous nous intéresserons au modèle linéaire (1.1.) défini par :

(i) $\underset{\sim}{Z}(\alpha, \underset{\sim}{\beta}) = (Z_i(\alpha, \underset{\sim}{\beta}) = Y_i - \alpha - \underset{\sim}{x}'_i \underset{\sim}{\beta}, i = 1 \ldots n)$ est un n-échantillon (collection de variables aléatoires indépendantes équidistribuées) de loi F continue et à information de Fisher finie, où :

(ii) Y_i $(i = 1 \ldots n)$ est aléatoire réelle ("variable à expliquer") ;

(iii) $\underset{\sim}{x}'_i$ $(i = 1 \ldots n)$ est un vecteur ligne contenant la i-ème observation de p variables fixes et connues ("variables explicatives") ; nous noterons X la matrice n x p égale à $(\underset{\sim}{x}_1\ \underset{\sim}{x}_2 \ldots \underset{\sim}{x}_p)'$ (où le ' désigne la transposition de matrice) ;

(iv) α et $\underset{\sim}{\beta}$ sont des paramètres fixes et inconnus (les "coefficients de régression") respectivement scalaire et vectoriel à p dimensions.

Nous examinerons une classe de méthodes robustes de traitement de ce modèle linéaire, concurrentes de la méthode des moindres carrés : les R-estimateurs de JAECKEL (1972).

Nous montrons au paragraphe 2 que la régression en norme ℓ_1 (estimant α et $\underset{\sim}{\beta}$ par les valeurs qui minimisent le critère

$$\ell_1(\underset{\sim}{Z}(\alpha , \underset{\sim}{\beta})) = \sum_{i=1}^{n} |Z_i(\alpha, \underset{\sim}{\beta})|)$$

est un cas particulier de R-estimation, ce qui confirme le rôle central de cette méthode dans le domaine de la régression robuste, et nous verrons quelles bonnes propriétés des R-estimateurs peuvent être ajoutées à celles de la régression en norme ℓ_1.

Nous cherchons ensuite (paragraphes 3 et 4) à exploiter cette identité en sens inverse, c'est-à-dire en voyant quelles bonnes propriétés (tant statistiques, cf. BASSETT et KOENKER, 1978, qu'algorithmiques, cf. GENTLE, 1977) de la régression en norme ℓ_1 peuvent être transposées aux R-estimateurs de JAECKEL (1972) .

Nous essaierons de voir au paragraphe 4 quelles perspectives s'ouvrent aux R-estimateurs.

2. LA REGRESSION EN NORME ℓ_1, CAS PARTICULIER DE R-ESTIMATION.

Nous allons d'abord rappeler ce qui nous paraît l'essentiel de la R-estimation.

Soit $(a(i),\ i = 1 \dots n)$ un ensemble de scores (valeurs numériques) satisfaisant à :

Propriété (2.1.) $\quad \sum_{i=1}^{n} a(i) = 0 \quad$ (moyenne nulle)

Propriété (2.2.) $\quad i > j \implies a(i) \geq a(j) \quad$ (scores croissants)

Par analogie avec la somme des carrés d'un n-échantillon centré $Z = (Z_i,\ i = 1 \dots n)$, qui peut s'écrire

$$\sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} Z_{(i)} \cdot Z_{(i)}$$

où $Z_{(i)}$ est la i-ème statistique d'ordre ("valeur par ordre croissant"), JAECKEL (1972) définit une mesure de dispersion $D_a(\underset{\sim}{Z})$ par :

Définition (2.3.)

$$D_a(Z) = \sum_{i=1}^{n} a(i) \cdot Z_{(i)}$$

Naturellement, chaque choix de scores définit une mesure de dispersion, invariante par translation $(D_a(\underset{\sim}{Z} + b.\underset{\sim}{1}) = D_a(\underset{\sim}{Z})$: c'est une conséquence triviale de la propriété 2.1.). On voit que dans la somme, l'un des deux $Z_{(i)}$, susceptible de variations aléatoires "anormales", a été remplacé par un score fixe $a(i)$: la mesure de dispersion fait donc intervenir chaque valeur de l'échantillon par sa grandeur absolue et non par son carré, ce qui laisse espérer pour $D_a(.)$ de bonnes propriétés de robustesse en présence de valeurs aberrantes (c'est d'ailleurs à notre connaissance la seule justification de l'étiquette "robuste" apposée sur les R-estimateurs pour le modèle linéaire).

Remarque : on peut d'ailleurs interpréter les propriétés (2.1.) et (2.2.) comme destinées à faire simuler des "erreurs ordonnées" par les scores ("erreurs" : 2.1., "ordonnées" : 2.2.).

JAECKEL (1972) définit ensuite le R-estimateur associé aux scores $a(.)$ comme un vecteur $\underset{\sim}{\beta}_a$, non nécessairement unique, pour lequel $D_a(\underset{\sim}{Z}(\alpha,\ \underset{\sim}{\beta}))$ atteint son minimum (son existence est assurée par la propriété 2.5. que nous rappellerons plus loin). Il ne propose rien pour l'estimation de α, puisque $D_a(\underset{\sim}{Z}(\alpha,\ \underset{\sim}{\beta}))$ ne dépend pas de α (à cause de l'invariance de D_a par translation). Par contre, McKEAN et HETTMANSPERGER (1976) complètent la procédure en estimant α par inversion du test de rang (cf. LEHMANN, 1975) sur la position des $(Z_i(0,\ \underset{\sim}{\beta}_a),\ i = 1 \dots n)$ défini par les scores caractérisant $D_a(.)$ (par l'association classique, cf. JAECKEL, 1971, entre scores pour le problème "un échantillon" et scores pour "deux échantillons"). Nous entendrons par R-estimation du modèle (1.1.) la procédure de JAECKEL (1972) complétée par celle de McKEAN et HETTMANSPERGER (1976).

Remarque : suivant l'exemple de JAECKEL (1972), nous n'insisterons pas sur la façon de choisir les scores, renvoyant à JURECKOVA (1971). Rappelons toutefois que des critères d'efficacité asymptotique permettent de choisir les scores en fonction de la distribution des erreurs (loi F) (cf. la notion de procédures "asymptotiquement localement les plus puissantes" dans HAJEK et SIDAK, 1967, par exemple). Dans cette association entre F et $a(.)$, la propriété (2.2.) caractérise les distributions F "fortement unimodales" (cf. HAJEK et SIDAK, 1967, p. 20). Une autre propriété qu'imposent HETTMANSPERGER et McKEAN (1977) assure l'invariance de $D_a(.)$ par changement de signe $(D_a(-\underset{\sim}{Z}) = D_a(\underset{\sim}{Z}))$:

Propriété (2.4.) $a(n+1-i) = -a(i)$, $i = 1 \dots n$

Nous ne la retiendrons pas plus que JAECKEL (1972), car elle n'est pas essentielle pour les propriétés fondamentales des R-estimateurs, bien que cette invariance soit agréable pour interprêter la procédure. Elle est en effet caractéristique, dans l'association mentionnée plus haut entre F et $a(.)$, de lois F symétriques, ce qui nous semble une restriction a priori peu souhaitable. Notons enfin que les procédures étudiées ici, ont, en tous cas, l'intérêt d'être asymptotiquement libres de distribution.

Les trois propriétés des R-estimateurs qui nous paraissent particulièrement attrayantes sont les suivantes :

Propriété (2.5.) (JAECKEL, 1972) : $D_a(\underset{\sim}{Z}(\alpha, \underset{\sim}{\beta}))$ est une fonction de $\underset{\sim}{\beta}$ seul, positive, convexe, linéaire par morceaux, atteignant son minimum dans un domaine borné. Nous reviendrons sur l'intérêt de cette propriété au paragraphe 4.

Propriété (2.6. (JAECKEL, 1972) : sous des conditions diverses sur les scores, sur la loi F (à information de Fisher finie, fortement unimodale) et sur le comportement des $\underset{\sim}{x}'_i$ (en particulier, $\lim_{n\to\infty} (X'X/n)^{-1} = \Sigma$ définie positive), $\sqrt{n}(\underset{\sim}{\beta}_a - \underset{\sim}{\beta})$ est asymptotiquement gaussien multivariate, de moyenne nulle et de matrice de covariance $\tau^2.\Sigma$ (où le scalaire τ^2 est déterminé par F et $a(.)$).

Propriété (2.7.) (McKEAN et HETTMANSPERGER , 1976) proposent un estimateur de τ convergent (à une constante multiplicative près, c'est l'étendue de l'intervalle de confiance provenant du même test de rang que celui servant à estimer α), ce qui permet de réaliser des tests d'hypothèses linéaires sur les $\underset{\sim}{\beta}$, asymptotiquement libres de distribution.

Remarque : τ^2 joue bien le rôle dévolu à σ^2 (variance des $Z_i(\alpha, \underset{\sim}{\beta})$) dans l'estimation par moindres carrés : la définition même de l'estimateur de τ^2 en fait une mesure de la dispersion résiduelle.

Nous pouvons maintenant énoncer notre lemme :

Lemme (2.8.) : la régression en norme ℓ_1 est le cas particulier de R-estimation associé aux scores de la médiane, c'est-à-dire :

$$a_M(i) = \begin{cases} -1 & \text{si } i < (n+1)/2 \\ 0 & \text{si } i = (n+1)/2 \quad \text{(cas } n \text{ impair)} \\ 1 & \text{si } i > (n+1)/2 \end{cases}$$

Démonstration : nous démontrerons cette identité séparément par α et $\underset{\sim}{\beta}$. Pour $\underset{\sim}{\beta}$, nous montrerons que les deux critères $\ell_1(.)$ et $D_{a_M}(.)$ coïncident, ce qui implique l'identité des "domaines de solutions" (rappelons que les solutions ne sont a priori nécessairement uniques ni dans un cas ni dans l'autre).

Soit $m(\underset{\sim}{\beta})$ la médiane des $Z_i(0, \underset{\sim}{\beta})$ que nous noterons simplement Z_i ($Z_{(i)}$ étant la i-ème statistique d'ordre de l'échantillon des Z_i). Alors, $D_{a_M}(\underset{\sim}{Z}(\alpha, \underset{\sim}{\beta}))$ s'écrit :

$$\begin{aligned} D_{a_M}(\underset{\sim}{Z}(\alpha, \underset{\sim}{\beta})) &= D_{a_M}(\underset{\sim}{Z}(0, \underset{\sim}{\beta})) && \text{(invariance par translation)} \\ &= \sum_{i=1}^{n} a_M(i) \,.\, Z_{(i)} && \text{(par définition)} \\ &= \sum_{i=1}^{n} a_M(i) \,.\, (Z_{(i)} - m(\underset{\sim}{\beta})) && \text{(invariance par translation)} \end{aligned}$$

n

$$= \sum_{i=1}^{n} |Z_{(i)} - m(\beta)|$$ (en remplaçant les scores par leurs valeurs, et en analysant simplement les signes des $Z_{(i)}-m(\beta)$)

$$= \sum_{i=1}^{n} |Z_i - m(\beta)|$$ (l'ordre de sommation n'intervient plus)

D'autre part, on peut choisir de minimiser $\ell_1(Z(\alpha, \beta))$ en deux étapes :

a) pour β fixé, chercher $\alpha(\beta)$ minimisant $\ell_1(Z(\alpha, \beta))$. Or

$$\ell_1(Z(\alpha, \beta)) = \sum_{i=1}^{n} |Z_i(\alpha, \beta)| = \sum_{i=1}^{n} |Z_i(0, \beta) - \alpha| = \sum_{i=1}^{n} |Z_i - \alpha|$$

Il suffit donc de prendre $\alpha(\beta)$ = médiane des $Z_i = m(\beta)$.

b) reportant $\alpha(\beta)$ dans $\ell_1(Z(\alpha, \beta))$, il ne reste plus qu'à minimiser la fonction de β seul $\ell_1(Z(\alpha(\beta), \beta)) = \sum_{i=1}^{n} |Z_i - m(\beta)| \equiv D_{a_M}(Z(\alpha, \beta))$, ce qu'il fallait démontrer pour l'identité des estimateurs de β.

Enfin, le R-estimateur de α proposé par McKEAN et HETTMANSPERGER (1976) est obtenu en inversant le test de rang sur la position des Z_i (calculés pour $\beta = \beta_{a_M}$ obtenu en minimisant $D_{a_M}(Z(\alpha, \beta))$), qui, dans le cas des scores de la médiane, est simplement le test du signe : l'estimateur associé est donc bien la médiane (cf. LEHMANN, 1975) des Z_i, coïncidant donc avec l'estimateur de α de la régression en norme ℓ_1. CQFD.

L'intérêt de cette identité est double, correspondant à l'application des propriétés (2.6.) et (2.7.) des R-estimateurs au cas particulier de la régression en norme ℓ_1 :

a) la propriété (2.6.) établit les propriétés asymptotiques de la régression en norme ℓ_1, les conditions sur les scores étant vérifiées (convergence et donc "unicité asymptotique", normalité asymptotique, efficacité asymptotique relative par rapport à l'estimation par moindres carrés, bien définie puisque les matrices de covariance asymptotiques des deux estimateurs sont proportionnelles) par une voie différente de celle offerte par BASSETT et KOENKER (1978) ;

b) et la propriété (2.7.) définit une procédure de tests d'hypothèses linéaires sur les β en régression en norme ℓ_1, dont l'absence étant jusqu'à présent un défaut majeur de cette méthode.

Remarque : il est amusant de noter que HETTMANSPERGER et McKEAN (1977) ont pratiqué la régression en norme ℓ_1 "sans le savoir" (et qu'ils avaient déjà forgé la procédure de tests d'hypothèses associée), plus précisément en pratiquant la R-estimation avec les scores de la médiane.

3. UN NOUVEL ALGORITHME POUR CERTAINS R-ESTIMATEURS.

HETTMANSPERGER et McKEAN ne se sont en fait que peu intéressés aux scores de la médiane, essentiellement parce que l'efficacité asymptotique relative des R-estimateurs par rapport aux moindres carrés est exactement parallèle à celle des procédures de rang associées aux scores en statistique non paramétrique classique (cf. par exemple HAJEK et SIDAK, 1967, ou LEHMANN, 1975). Les scores de WILCOXON assurant une efficacité supérieure dans le cas gaussien ($3/\pi \simeq 95$ % contre $2/\pi \simeq 64$ % pour les scores de la médiane), et bornée inférieurement pour les autres distributions symétriques (par la valeur $108/125 \simeq 86,4$ % cf. LEHMANN 1975 p.80),

ils se sont intéressés essentiellement aux R-estimateurs définis par ces scores :
$a_W(i) = i - (n+1) / 2$.
Or ils ont remarqué (1978) , après WAINER et THISSEN (1976), que la mesure de dispersion associée, $D_{a_W}(.)$ est algébriquement équivalente au coefficient "de la moyenne des différences" de GINI (cf. par exemple DAVID, 1968) qui lui-même est proportionnel à $D(.)$ défini par :

$$D(\underset{\sim}{Z}) = \sum_{1 \leq i < j \leq n} |Z_i - Z_j|$$

Appliquons cette mesure aux $Z_i(\underset{\sim}{\beta})$:

$$D(\underset{\sim}{Z}(\underset{\sim}{\beta})) = \sum_{1 \leq i < j \leq n} |Z_i(\underset{\sim}{\beta}) - Z_j(\underset{\sim}{\beta})|$$
$$= \sum_{1 \leq i < j \leq n} |(Y_i - Y_j) - (\underset{\sim}{X}'_i - \underset{\sim}{X}'_j)\underset{\sim}{\beta}|$$

Sous cette forme, on voit immédiatement que la minimisation de $D(\underset{\sim}{Z}(\underset{\sim}{\beta}))$, et donc de $D_{a_W}(\underset{\sim}{Z}(\alpha, \underset{\sim}{\beta}))$ se ramène à la régression en norme ℓ_1 "expliquant" les $n(n-1)/2$ différences $(Y_i - Y_j)$ par les $(\underset{\sim}{X}'_i - \underset{\sim}{X}'_j)$, ce qui constitue le nouvel algorithme pour ce type de R-estimateurs : un gros avantage de la régression en norme ℓ_1 est en effet que son calcul peut s'effectuer par des algorithmes de programmation linéaire performants (cf. par exemple BARRODALE et ROBERTS, 1974) et sûrs (on aboutit à une solution " numériquement exacte" en un nombre fini d'étapes).
Nous avons écrit un programme FORTRAN fondé sur cette remarque et sur l'algorithme de BARRODALE et ROBERTS (1974). Le programme commence par effectuer la régression en norme ℓ_1 ordinaire, puis effectue la régression par R-estimateurs à scores de WILCOXON "expliquant" les résidus de la première (l'idée est analogue à celle de SPOSITO et HAND, 1977, qui proposent de pratiquer la régression en norme ℓ_1 "expliquant" les résidus de la régression par moindres carrés, afin de diminuer le nombre d'itérations de l'algorithme de programmation linéaire).
Exemple : nous avons repris les données analysées (par les scores de la médiane et les scores de WILCOXON) par HETTMANSPERGER et McKEAN (1977) (ces données sont présentées en détail dans DANIEL et WOOD, 1971, chap. 6). Les résultats (tableau 1) confirment ceux obtenus par HETTMANSPERGER et McKEAN, bien qu'illustrant de légères imprécisions dans les résultats obtenus par l'algorithme de minimisation générale du type "plus grande pente" ou NEWTON-RAPHSON modifié utilisé par HETTMANSPERGER et McKEAN.

		norme ℓ_1 ou scores de la médiane		scores de WILCOXON	
		minimisation générale	programmation linéaire	minimisation générale	programmation linéaire
Estimations de	α	-39.70	-39.69	-39.95	-39.85
	β_1	0.83	0.83	0.80	0.79
	β_2	0.58	0.57	0.90	0.90
	β_3	-0.06	-0.06	-0.11	-0.11

Tableau 1. Comparaison des résultats des deux algorithmes sur les données de DANIEL et WOOD (1971) chap. 6

Nous n'avons écrit ce programme que pour la comparaison des résultats des deux algorithmes. En effet, le "nombre d'observations à expliquer par une régression en norme ℓ_1", dans le cas des scores de WILCOXON, est $n(n-1)/2$ et croit donc très vite avec le "vrai" nombre d'observations (n). Un simple coup d'oeil au tableau 2 suffira à s'en convaincre (précisons que le programme occupe 196 Koctets).

Nous tenons néanmoins le programme à la disposition des lecteurs intéressés .

P	1	2	3	4	5	10	20	30
n	102	100	94	89	85	70	55	46

Tableau 2. Nombre maximum d'observations (n) pour un nombre donné (p) de variables explicatives.

4. PERSPECTIVES.

Nous travaillons à améliorer la R-estimation dans deux directions : algorithmique et statistique.

Algorithmique : il est clair que l'algorithme précédent, pour les scores de WILCOXON, a grand besoin d'être amélioré. Il est vraisemblable que ceci se fera en modifiant le "tableau de contraintes" de l'algorithme de programmation linéaire : en effectuant des combinaisons linéaires simples (des différences) des lignes de ce tableau, on peut le rendre très "creux" tout en le laissant "bien structuré", ce qui laisse entrevoir de grosses économies d'encombrement, au prix d'un effort de programmation qui reste à faire.

D'autre part, ni JAECKEL (1972) ni HETTMANSPERGER et McKEAN (1976 a et b, 1977, 1978 a et b) ne semblent avoir remarqué l'importance pratique de la propriété 2.5. et plus particulièrement de la"linéarité par morceaux" de la fonction (de $\underset{\sim}{\beta}$) $D_a(\underset{\sim}{Z}(\alpha, \underset{\sim}{\beta}))$: elle signifie pourtant que, quel que soit le choix des scores, la minimisation est un problème de programmation linéaire. Nous travaillons à réexprimer dans ce sens le problème de minimisation, afin de profiter plus pleinement des avantages liés aux algorithmes de programmation linéaire :

- solutions "numériquement exactes" en un nombre fini d'étapes ;
- possibilités d'indications sur l'unicité des solutions obtenues ;
- possibilités d'inclure des contraintes linéaires sur les $\underset{\sim}{\beta}$ (cf. BARRODALE et ROBERTS, 1977) ;
- possibilités de construire des algorithmes recherchant le(s) "meilleur(s)" sous ensemble(s) de variables explicatives (cf. ROODMAN, 1974);
- efficacité des algorithmes, diminuant l'intérêt, dans le cas d'échantillons "raisonnables" (quelques centaines d'observations), des R-estimateurs "en une étape" proposés par McKEAN et HETTMANSPERGER (1978), ceux-ci conservant bien entendu tout leur intérêt dans le cas de très grands échantillons.

Statistique : nous cherchons d'une part à mieux préciser les propriétés de robustesse des R-estimateurs (courbe d'influence par exemple, cf. HAMPEL, 1974), et d'autre part à explorer plus complètement les possibilités qu'offre le choix des scores :

- les scores normaux, aboutissant à des procédures d'efficacité toujours supérieure ou égale à celle des moindres carrés (cf. LEHMAN, p. 97 par exemple) devraient prendre une grosse importance, bien qu'ils n'aient jusqu'à présent pas été utilisés (sans doute parce qu'ils sont très peu utilisés en statistique non paramétrique classique, considérée à tort comme ensemble de techniques utilisables "à la main" : les scores normaux ne sont guère maniables qu'en ordinateur, mais c'est un "défaut" qu'ils partagent avec toute régression linéaire multiple !) ;
- les procédures "adaptatives", où l'on choisit les scores en fonction des données, constituent aussi une voie de recherche attrayante (cf. HOGG et RANDLES, 1975)
- enfin, les scores ne satisfaisant pas la propriété (2.4.), adaptés au cas de distributions F non symétriques, n'ont jamais été même seulement essayés : nous pensons qu'ils pourraient être très intéressants, particulièrement si on les inclut dans le choix possible pour une procédure adaptative.

Remerciements .

Nous tenons à remercier Monsieur le Professeur CAUSSINUS, de l'Université Paul Sabatier de Toulouse, dont les remarques n'ont pas peu contribué à rendre notre texte lisible. Que nos collègues du laboratoire de Biométrie du C.N.R.Z. soient aussi remerciés des discussions fructueuses que nous avons eues sur le sujet.

BIBLIOGRAPHIE.

BARRODALE, I. et ROBERTS, F. D. K. (1974). Algorithm 478 : solution of an overdetermined system of equations in the ℓ_1-norm. *Comm. A.C.M.* 17, 319-320.

BARRODALE, I. et ROBERTS, F. D. K. (1977). Algorithms for restricted least absolute value estimation. *Comm. Statist. Computa. Simula.* B6(4), 353-363.

BASSETT, G. Jr et KOENKER, R. (1978). Asymptotic theory of least absolute error regression. *J. Amer. Statist. Assoc.* 73, 618-621.

DANIEL, C. et WOOD, F.S. (1971) *Fitting equations to data*. John Wiley, New-York

DAVID, H.A. (1968) Gini's mean difference rediscovered. *Biometrika* 55, 573-575.

GENTLE, J.E. (1977) Least absolute values estimation : an introduction. *Commun. Statist. Simula. Computa.* B6(4), 313-328.

HAJEK, J. et SIDAK, Z. (1967). *Theory of rank tests*. Academic Press, New-York.

HAMPEL, F.E. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* 69, 383-393.

HETTMANSPERGER, T.P. et McKEAN, J.W. (1976). Computational problems involved in analysis of linear models based on ranks. *Proc. Statist. Computa. Sect. Amer. Statist. Assoc.*, 88-94.

HETTMANSPERGER, T.P. et McKEAN, J.W. (1977). A robust alternative based on ranks to least squares in analyzing linear models. *Technometrics* 19, 275-284.

HETTMANSPERGER, T.P. et McKEAN, J.W. (1978). Statistical inference based on ranks. *Psychometrika* 43, 69-79.

HOGG , R.V. et RANDLES, R.H. (1975). Adaptive distribution free regression methods and their applications. *Technometrics* 17, 399-407.

JAECKEL, L.A. (1971). Robust estimates of location : symmetry and asymmetric contamination. *Ann. Math. Statist.* 42, 1020-1034.

JAECKEL, L.A. (1972) Estimating regression coefficients by minimizing the dispersion of the residuals. *Ann. Math. Statist.* 43, 1449-1458.

JURECKOVA, J. (1971). Nonparametric estimate of regression coefficients. *Ann. Math. Statist.* 42, 1328-1338.

McKEAN, J.W. et HETTMANSPERGER, T.P. (1976). Tests of hypothesis based on ranks in the general linear model. *Commun. Statist. Theor. Meth.* A5(8), 693-709.

McKEAN, J.W. et HETTMANSPERGER, T.P. (1978). A robust analysis of the general linear model based on one step R-estimates. *Biometrika* 65, 571-579.

LEHMANN, E.L. (1975). *Nonparametrics : Statistical methods based on ranks*. Holden Day, San Francisco.

ROODMAN, G.M. (1974). A procedure for optimal stepwise MSAE regression analysis. *Operations Research* 23, 393-399.

SPOSITO, V.A. et HAND, M.L. (1977) Using an approximate ℓ_1-estimator. *Comm. Statist. Simula Computa.* B6(3), 263-268.

WAINER, H. et THISSEN, D. (1976). Three steps towards robust regression. *Psychometrika*, 41, 9-34.