REVUE DE STATISTIQUE APPLIQUÉE

MOHAMED EL ARBI CHAFFAI

Les estimateurs robustes sont-ils vraiment robustes en pratique ?

Revue de statistique appliquée, tome 37, n° 3 (1989), p. 57-73

<http://www.numdam.org/item?id=RSA_1989__37_3_57_0>

Rev. Statistique Appliquée, 1989, **XXXVII** (3), 57-73

LES ESTIMATEURS ROBUSTES SONT-ILS VRAIMENT ROBUSTES EN PRATIQUE ?

Mohamed El Arbi CHAFFAI

Assistant à la Faculté des Sciences Economiques et de Gestion de Sfax - Route de l'aérodrome Km4 - Tunisie

RESUMÉ

On considère la classe générale des M et W-estimateurs dans le cadre d'un modèle de régression multiple. Nous montrons que l'estimateur de Huber est souvent sensible en pratique à certains paramètres, et nous proposons dans un premier temps une procédure adaptative qui tient compte du degré de perturbation dans l'échantillon. Nous étudions aussi un autre estimateur dans la classe des W-estimateurs plus stable que l'estimateur de Huber et d'Andrews, ce qui nous a amené à proposer une généralisation des W-estimateurs.

Mots clés : *M-estimateurs, MG-estimateurs, régression robuste, moindres carrés repondérés et itérés.*

ABSTRACT

On considering the general class of M and W-estimators in the linear regression model, our first findind is that Huber estimator is often sensitive in practice to some parameters, so we propose an adaptative procedure which takes into account the degree of contamination of errors in the sample. The second finding is that, one other estimator in the above mentioned class of W-estimators is more stable, than Huber and Andrews estimator. So we propose a generalisation of W-estimators.

INTRODUCTION

L'estimation des modèles économétriques par la méthode des moindres carrés (MC) est souvent sensible à des observations mesurées dans les queues de distributions, à des erreurs grossières sur les variables, ou bien encore à une mauvaise spécification de la relation entre la variable endogène et les variables exogènes du modèle.

Certains auteurs Belsley et al (1980), Cook et Weisberg (1980) et Cook (1977), se sont intéressés aux points aberrants dans les modèles de régression, comme étant un problème d'identification, c'est à dire que la méthode des MC reste optimale, si l'on arrive à éliminer "correctement" les observations qui influencent l'ajustement. Plusieurs ouvrages récents d'économétrie appliquée consacrent de longs chapitres à ces méthodes, Weisberg (1980), Fox (1984) et Montgomery et Peck (1982).

Parrallèlement d'autres auteurs Huber (1973), (1981), Andrews (1974), Krasker (1980) et Krasker et Welsch (1982), ont développé des méthodes automatiques d'ajustement robuste, qui ont l'avantage d'être presque aussi efficaces que la méthode des MC lorsqu'il n'y a pas de points aberrants, mais plus efficaces en présence d'observations atypiques ou bien encore lorsque la distribution de l'erreur dans le modèle suit une distribution à queues longues.

On distingue dans la littérature, les M-estimateurs, les W-estimateurs et les MG-estimateurs (les M-estimateurs généralisés). En pratique, nous avons constaté que l'estimateur de Huber est souvent sensible à certain paramètres, ce qui nous a amené à proposer une procédure adaptative qui tient compte du degré de perturbation dans l'échantillon.

De plus, nous avons trouvé que certains estimateurs de la classe des W-estimateurs sont en pratique plus stables que l'estimateur de Huber; nous avons alors proposé une généralisation des W-estimateurs : les WG-estimateurs.

Le papier sera organisé de la façon suivante : dans une première partie, nous ferons des rappels sur les M-estimateurs et les W-estimateurs; nous aborderons le problème de l'efficacité de l'estimateur de Huber et nous développerons l'estimateur adaptatif de Huber. Une deuxième partie sera consacrée aux MG-estimateurs et aux WG-estimateurs, et enfin dans une dernière partie nous comparerons ces estimateurs sur le modèle de Henderson et Velleman (1981).

I. M-estimateurs et W-estimateurs

On considère le modèle de régression multiple :

$$y_i = \underline{x}_i'\underline{\beta} + u_i \quad i = 1....,T \tag{1.1}$$

y_i représente une observation de la variable à expliquer $\underline{y}, \underline{x}_i'$ un vecteur ligne à K composantes représentant les valeurs de l'observation i sur les variables explicatives $X_1,\ X_2...X_K,\ \underline{\beta}$ un vecteur inconnu dont on cherchera une estimation et u_i, la $i^{ème}$ composante du vecteur aléatoire $\underline{u}$, de moyenne zéro et de matrice des variances covariances $\sigma^2 I(I$ étant la matrice unité). On suppose aussi que les variables explicatives sont non stochastiques et non collinéaires.

I.1 Les M-estimateurs

Huber (1972) qualifie de "dogme" que de supposer l'hypothèse de normalité sur les perturbations u_i dans (1.1) et propose une famille d'estimateurs qui généralisent les estimateurs du maximum de vraisemblance : il s'agit des M-estimateurs. L'avantage de ces estimateurs, c'est qu'ils sont peu sensibles aux observations aberrantes, (en particulier pour des distributions des u_i à queues longues) et assez proches de l'estimateur des MC, lorsque la distribution de l'erreur est gaussienne dans le modèle.

L'estimateur de Huber (1972) de $\underline{\beta}$ est défini par :

$$\operatorname*{Min}_{\underline{\beta}} \sum_{i=1}^{T} \rho\left(y_i - \underline{x}_i'\underline{\beta}\right) \tag{1.2}$$

où $\rho(.)$ désigne une fonction convexe, continue et dérivable. On retrouve l'estimateur des MC pour $\rho_1(t) = 0.5t^2$ et l'estimateur minimisant la somme des valeurs absolues des écarts MSAE pour $\rho_2(t) = |t|$.

Huber (1964), a proposé une fonction $\rho_H(.)$ qui a les mêmes propriétés que $\rho_1(t)$ pour les écarts faibles et celles de $\rho_2(t)$ pour les écarts importants. Si $\psi(.)$ désigné la dérivée de $\rho(.)$, la fonction de Huber est définie par :

$$\psi_H(t) = \mathrm{Min}(H, \mathrm{Max}(-H, t)) \tag{1.3}$$

H désigne un paramètre positif à fixer par l'utilisateur.

Le problème que pose ce type de fonctions (1.3), est que les estimateurs ne vérifient plus les propriétés d'invariance par changement d'échelle; si par exemple on multiplie les écarts par une constante b, l'estimateur $\underline{\beta}$ vérifiant (1.2) ne devrait pas changer.

Pour vérifier cette propriété Huber propose une légère modification de (1.2); si d désigne une mesure de dispersion des écarts, on aura :

$$\mathrm{Min}\ Q(\underline{\beta}, d) = \sum_{i=1}^{T} \rho_H\left[(y_i - \underline{x}_i'\underline{\beta})/d\right] \tag{1.4}$$

Comme d est inconnu, (estimation appropriée de la dispersion des écarts σ), on cherchera alors une estimation simultanée de $\underline{\beta}$ et d solution du système suivant (où $\underline{x}_i$ désigne le vecteur colonne transposé de $\underline{x}_i'$) :

$$\left.\begin{aligned}
\partial Q(\underline{\beta}, d)/\partial \underline{\beta} &= \sum_{i=1}^{T} \psi_H\left[\left(y_i - \underline{x}_i'\underline{\beta}\right)/d\right]\underline{x}_i = \underline{0} \\
\partial Q(\underline{\beta}, d)/\partial d &= -(1/\mathrm{d}^2) \sum_{i=1}^{T} (y_i - \underline{x}_i'\underline{\beta})\psi_H\left[\left(y_i - \underline{x}_i'\underline{\beta}\right)/d\right] = 0
\end{aligned}\right\} \tag{1.5}$$

Huber (1981), note que la résolution du système (1.5) peut présenter des problèmes de convergence; il propose alors de minimiser une autre expression pour estimer simultanément $\underline{\beta}$ et d solution de :

$$\mathrm{Min}Q(\underline{\beta}, d) = ad + \sum_{i=1}^{T} \rho_H\left[\left(y_i - \underline{x}_i'\underline{\beta}\right)/d\right] d \tag{1.6}$$

La constante a est incluse ici pour que, asymptotiquement d soit presque égal à σ (écart type des perturbations).

La résolution du système (1.6) s'effectue en général par procédure itérative, où à chaque étape on devra estimer simultanément $\underline{\beta}$ et d. On trouvera dans Dutter (1977) plusieurs algorithmes de calcul, dans Hogg (1979) et Peters et al (1982), une présentation détaillée d'un algorithme valable seulement pour la fonction de Huber (1.3).

Sous certaines conditions de régularité, et lorsque σ est connu, l'estimateur de $\underline{\beta}$ obtenu en minimisant (1.6) suit approximativement une distribution normale multivariée de moyenne $\underline{\beta}$ et de matrice des variances covariances définie par Hubert (1973), (1981) :

$$\mathrm{Var}(\underline{\beta}) = \sigma^2 \frac{E\left[\psi_H^2(\underline{u}/\sigma)\right]}{\left(E\left[\psi'_H(\underline{u}/\sigma)\right]\right)^2} \left(\sum_{i=1}^{T} \underline{x}_i \underline{x}'_i\right)^{-1} \tag{1.7}$$

Notons que plusieurs autres fonctions robustes ont été proposées dans la littérature; les plus utilisées dans les travaux empiriques sont celles de Huber, d'Andrews (1974), de Hampel (1974) et de Fair. On trouvera ces fonctions robustes dans Holland et Welsch (1977) et Montgomery et Peck (1982).

I.2. Les W-estimateurs

C'est une forme alternative des M-estimateurs, utilisant la régression pondérée développée par Holland et Welsch (1977) et définie par :

$$\mathrm{Min}\ Q(\underline{\beta}, d) = \sum_{i=1}^{T} w_i(.)(y_i - \underline{x}'_i\underline{\beta})^2 \tag{1.8}$$

$w(.)$ est une fonction des écarts définie par : $w(t) = \psi(t)/t$. Si l'on dispose déjà d'une estimation de la dispersion $\tilde{d}$, il suffit de résoudre :

$$\sum_{i=1}^{T} w_i(y_i, \underline{x}'_i, \tilde{d})(y_i - \underline{x}'_i\underline{\beta})x_{ij} = 0, j = 1..., K \tag{1.9}$$

$w_i(.)$ est une fonction de pondérations, donnant un poids w_i voisin de zéro pour les observations ayant des résidus importants, et un poids proche de un pour les observations ne présentant pas d'anomalies. Cette fonction est définie par :

$$w_i(y_i,\ \underline{x}'_i,\ \tilde{d}) = \frac{\psi\left[\left(y_i - \underline{x}'_i\underline{\beta}\right)/\tilde{d}\right]}{\left[\left(y_i - \underline{x}'_i\underline{\beta}\right)/\tilde{d}\right]}, i = 1...., T \tag{1.10}$$

$\psi(.)$ étant un fonction objectif quelconque.

Les fonctions robustes utilisées dans ce papier, sont celles de Huber (cf. formule (1.3)) d'Andrews et de Fair définis par :

$$\begin{aligned} \text{Andrews}: \psi(t) &= \sin(t/c_1) \quad |t| \leq \pi\, c_1 && (1.11)\\ &= 0 \qquad \text{ailleurs} \\ \text{Fair}: \psi(t) &= t(1 + |t|/c_2)^{-1} && (1.12) \end{aligned}$$

c_1 et c_2 désignent des constantes à fixer par l'utilisateur.

L'expression (1.8) est celle d'une régression pondérée; il suffira d'avoir un programme de régression pondérée et d'itérer (1.9) jusqu'à la convergence.

L'estimateur robuste de la dispersion le plus utilisé est défini par :

$$d = \operatorname{Med}_i \left| (y_i - \underline{x}_i'\underline{\beta}) - \operatorname{Med}(y_i - \underline{x}_i'\underline{\beta}) \right| / 0.6745 \qquad (1.13)$$

Holland et Welsch (1977) montrent que pour la fonction de Huber (1.3) on peut faire des itérations sur $\underline{\beta}$ et d en même temps, c'est à dire qu'on part d'une première estimation de $\underline{\beta}$, on calcule d dans (1.13), ensuite on calcule les $w_i(.)$ par (1.10) et on résoud à nouveau (1.9). A partir de cette nouvelle estimation de $\underline{\beta}$, on calcule à nouveau d, les $w_i(.)$ et ainsi de suite jusqu'à la convergence de d et $\underline{\beta}$.

Pour les fonctions $\rho(.)$ non convexes, et pour des raisons de convergence, les auteurs ne conseillent pas d'itérer $\underline{\beta}$ et d en même temps, mais de choisir une estimation de d et d'itérer seulement sur $\underline{\beta}$. cette méthode est souvent appelée la méthode des MC repondérés et itérés MCRI.

L'estimateur de la matrice des variances covariances du vecteur $\underline{\tilde{\beta}}$ solution de (1.8) sera :

$$\operatorname{Var}(\underline{\tilde{\beta}}) = \sigma_w^2 \left(\sum_{i=1}^{T} w_i(.) \underline{x}_i \underline{x}_i' \right)^{-1} \qquad (1.14)$$

avec

$$\sigma_w^2 = (1/(T-K)) \sum_{i=1}^{T} w_i(.)(y_i - \underline{x}_i'\underline{\tilde{\beta}})^2$$

$w_i(.)$ sont les pondérations calculées par (1.10) à la convergence de l'estimateur.

Il faut noter que dans le cadre des W-estimateurs, on n'estime pas simultanément $\underline{\beta}$ et d comme dans (1.5) ou (1.6); en effet on part généralement d'une mesure de dispersion robuste, (1.13) par exemple et on fait des itérations pour calculer $\underline{\beta}$. La procédure de calcul des W-estimateurs a attiré beaucoup d'auteurs, Holland et Welsch (1977), Asselin et al (1980); or il se trouve que les estimateurs robustes cherchés ne sont plus indépendants de l'échelle, comme on peut le constater sur l'exemple que nous traiterons à la fin de ce papier.

Notons par ailleurs que les M-estimateurs peuvent être considérés comme des W estimateurs. En effet, on peut exprimer facilement l'estimateur de Huber par exemple, sous la forme d'un estimateur des MC pondérés (cf Asselin de Beauville et al (1980)).

I.3. Sur l'efficacité de l'estimateur de Huber

La fonction objectif (1.3) de Huber dépend d'un paramètre H à fixer par l'utilisateur. Huber propose une valeur H = 1.345, "passe partout" qui assure une perte d'efficacité de l'ordre de 5 % si la distribution de l'erreur dans le modèle suit une loi normale centrée réduite. En particulier, Huber suppose que la distribution de l'erreur dans (1.1) est du type normale perturbée de la forme :

$$F(t) = (1 - \varepsilon)N(0,1) + \varepsilon G(t) \tag{1.15}$$

ϵ désigne le degré de perturbation compris entre 0 et 1 et G(t) une fonction de répartition arbitraire.

Huber (1981) a montré que pour ce type de distribution (1.15), la variance d'un estimateur de la classe M défini pour une fonction $\rho(.)$ convexe est :

$$\mathrm{Var}(F,\psi) = E\left[\psi^2(.)\right]/[E\psi'(.)]^2 \tag{1.16}$$

Pour la fonction (1.3) de Huber, on peut vérifier que pour une distribution des erreurs non perturbée, $\varepsilon = 0$ dans (1.15), on a :

$$E\left[\psi^2(.)\right] = g_1(H) = \mathrm{H}^2 + (1-\mathrm{H}^2)(2\Phi(H)-1) - 2H\emptyset(H) \tag{1.17}$$

et

$$E[\psi'(.)] = g_2(H) = 2\Phi(H) - 1 \tag{1.18}$$

où $\Phi(.)$ désigne la fonction de répartition d'une distribution normale centrée réduite et $\emptyset(.)$ sa densité.

Dans ces conditions, si l'on compare l' estimateur de Huber $\tilde{\beta}$ solution de (1.6) et l'estimateur des MC, à l'aide du critère de l'inefficacité relative de $\tilde{\beta}$ par rapport à $\hat{\beta}$ on aura :

$$e = E(\hat{\beta} - \beta)(\hat{\beta} - \beta)'/E(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)' \tag{1.19}$$

Sachant que les deux estimateurs sont sans biais, les u_i suivent une distribution N(0,1), (1.19) se ramène au rapport des traces des matrices des variances covariances des estimateurs; soit après simplification on aura :

$$e = 1/Var(F,\psi) = (g_2(H))^2/g_1(H) \tag{1.20}$$

Il suffit donc de se fixer le niveau d'efficacité désiré, et de calculer H à partir de (1.20). Ci dessous on donne les valeurs de H pour différentes valeurs de e d'après Krasker et Welsch (1982).

TABLEAU 1
Inefficacité de l'estimateur de Huber par rapport à l'estimateur des MC lorsque les u_i suivent une distribution N(0,1).

e	0.8	0.85	0.90	0.95	0.975	0.99	1
H	0.529	0.732	0.982	1.345	1.655	2.01	∞

Par ailleurs, le paramètre H est lié au degré de perturbation ε inconnu dans le modèle, (si l'on suppose que la distribution de l'erreur est du type (1.15). Pour la fonction (1.3) Huber (1981) montre que :

$$(2\emptyset(H)/H) - 2\Phi(-H) = \varepsilon/(1 - \varepsilon) \tag{1.21}$$

et pour H=1.345, ceci va correspondre à un degré de perturbation de l'ordre de 5,7 % .

I.4. L'estimateur adaptatif de Huber

La relation (1.21), indique que l'on doit fixer H en pratique sur la base d'une certaine connaissance du degré de perturbation ε inconnu dans l'échantillon.

Récemment, Samarov (1985) a proposé dans le cadre des MG-estimateurs de lier la constante H de la fonction de Huber (1.3) au degré de perturbation ε dans l'échantillon, en minimisant le maximum que peut atteindre l'erreur quadratique moyenne des MG-estimateurs.

Nous proposons dans ce travail pour la fonction (1.3) de Huber, de choisir la valeur de H qui minimise la variance estimée dans l'échantillon. En effet, une estimation de la matrice des variances covariances de l'estimateur de Huber solution de (1.6) est d'après (1.7) :

$$\text{Var}(\underline{\beta}) = \delta_H^{\,2} \left(\sum_{i=1}^{T} \underline{x}_i \underline{x}_i' \right)^{-1} \tag{1.22}$$

$$\delta_H^{\,2} = (\mathrm{T}^2/(T-K))\mathrm{d}^2 \, \frac{\sum_{i=1}^{T} \psi^2{}_H \left[(y_i - \underline{x}_i'\underline{\beta})) /d \right]}{\left(\sum_{i=1}^{T} \psi'_H \left[(y_i - \underline{x}_i'\underline{\beta}) /d \right] \right)^2} \tag{1.23}$$

Etant donné que la variance estimée de l'erreur $\delta^2{}_H$ dépend de H dans (1.23), nous proposons de calculer cette quantité pour différentes valeurs de H; disons en faisant varier H de 1 à 2 par exemple et de choisir la valeur optimale de H qui minimisera $\delta^2{}_H$.

Prenons le cas par exemple, où l'erreur dans le modèle suit une loi double exponentielles G(.) de densité $g(u) = 0.5 \ e^{-|u|}$. La variance de l'estimateur de

Huber sera à partir de (1.16) :

$$\mathrm{Var}(G,\psi) = \frac{2 - 2(1 + H)e^{-H}}{(1 - e^{-H})^2} \tag{1.24}$$

On peut remarquer aisément que cette variance est une fonction croissante de H. Pour ce modèle, il vaut mieux faire tendre H vers 0 et utiliser en conséquence l'estimateur qui minimise la somme des valeurs absolues des écarts MSAE[(1)].

Par ailleurs, si le praticien utilise automatiquement la fonction objectif de Huber (1.3), et le paramètre H=1.345, la variance de l'estimateur de Huber sera égale à 1.42 d'après (1.24) et il va enregistrer une perte d'efficacité de l'ordre de 30 % par rapport à l'estimateur MSAE.

II. La régression robuste généralisée

Les observations aberrantes en régression peuvent se présenter au niveau de la variable endogène et souvent au niveau des variables exogènes. Or les M-estimateurs et les W-estimateurs ne peuvent limiter que les mauvais effets des points aberrants qui se situent au niveau de la variable endogène. On propose dans la littérature les MG-estimateurs ; une classe d'estimateurs qui tient compte des deux types d'aberrations et nous proposerons une généralisation des W-estimateurs.

II.1. Les MG-estimateurs

La généralisation des M-estimateurs va tenir compte de la nature différente des points aberrants en régression. On distingue dans la littérature l'estimateur de Mallows (1975)[(2)], de Schweppe (1975)[(3)] et l'estimateur de Krasker et Welsch (1982). Les deux derniers estimateurs sont les plus utilisés en pratique.

La forme générale des MG-estimateurs est :

$$\text{Min } S(\underline{\beta},d) = \text{ad} + \sum_{i=1}^{T} v^2(x_i)\rho_H\left[(y_i - \underline{x}_i'\underline{\beta})/(dv(x_i))\right] d \tag{2.1}$$

a désigne une constante pour que l'estimateur de la dispersion d soit un estimateur efficace de σ lorsque l'erreur suit une distribution normale, $\rho_H(.)$ représente la fonction de Huber (1.3) et $v(x_i)$ une fonction de pondération dans l'espace des variables explicatives.

Schweppe (1975)[(3)] propose comme fonction de pondération :

$$v(x_i) = (1 - h_{ii})^{1/2} \tag{2.2}$$

(1) qui est l'estimateur du maximum de vraisemblance dans ce cas, de variance $[2g(0)]^{-2}$, d'après Basset et Koenker (1978).

(2) dans Peters et al (1982) ou Denby et Larsen (1977)

(3) dans Handschin, Kohlas et al (1975)

où h_{ii} est le $i^{ème}$ élément de la diagonale de la matrice de projection $\underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'$, qui permet de détecter les points aberrants sur les variables explicatives[1] ,$\underline{X}$ étant la matrice (TxK)des variables exogènes dans le modèle. On peut vérifier que les valeurs estimées des y_i par MC s'écrivent :

$$\hat{y}_i = \underline{x}_i' \left(\sum_{t=1}^{T} \underline{x}_t \underline{x}_t'\right)^{-1} \underline{x}_i y_i + \sum_{\substack{j=1 \\ j \neq i}}^{T} \underline{x}_i' \left(\sum_{t=1}^{T} \underline{x}_t \underline{x}_t'\right)^{-1} \underline{x}_j y_j \qquad (2.3)$$

lorsque h_{ii} va tendre vers 1, la valeur prévue sera fortement influencée par y_i, mais n'aura pas de mauvais effets sur l'ajustement dans (2.1) car $v^2(x_i)$ tendra vers 0 si on emploie (2.2).

Remarquons, que si l'on ignore le terme $v^2(x_i)$,la fonction qui minimise les écarts dans (2.1) est celle d'un M-estimateur de Huber (1.6), de paramètre $H' = Hv(x_i)$. Par conséquent, pour chaque observation $(y_i, \underline{x}_i')$ on a une fonction de pondération $\rho_{H'}(.)$, ce qui n'était pas le cas avec l'estimateur de Huber (1.6). En particulier, lorsque les observations sont uniformément distribuées dans l'espace des variables explicatives, tous les h_{ii} seront égaux à K/T, et dans ce cas l'estimateur défini en (2.1) est un M-estimateur de Huber, et la fonction de Huber (1.3) est de paramètre $H(1 - K/T)^{1/2}$ qui est inférieur à H; ceci va se traduire par une perte d'efficacité de l'estimateur de Huber (voir Tableau 1) par rapport à l'estimateur efficace des MC. C'est pour cela en fait que les MG-estimateurs sont moins efficaces que les M-estimateurs en cas d'absence de points aberrants sur les variables explicatives dans le modèle.

Krasker et Welsch (1982), ont défini leur MG-estimateur en étudiant l'influence de chaque observation sur l'ajustement; ces auteurs arrivent à la même expression que (2.1) avec :

$$v^2(x_i) = (\underline{x}_i'\underline{A}^{-1}\underline{x}_i) \qquad (2.4)$$

où $\underline{A}$est une matrice (KxK) qui dépend des variables explicatives du modèle et dont l'estimation vérifie :

$$\tilde{\underline{A}} = (1/T) \sum_{i=1}^{T} g_1 \left(\frac{H}{\left(\underline{x}_i' \tilde{\underline{A}}^{-1} \underline{x}_i\right)^{1/2}} \right) \underline{x}_i \underline{x}_i' \qquad (2.5)$$

$g_1(.)$ est définie dans (1.17) et H représente le paramètre de la fonction de Huber (1.3).

Signalons au préalable, que la démarche de ces auteurs est différente de celle de Huber, puisqu'ils considèrent que l'hypothèse de normalité est vérifiée pour le "modèle central" et qu'il définissent leur estimateur à partir des fonctions d'influence de Hampel (1974)[2] Aussi la détermination de H est différente de (1.20), voir Krasker et Welsch (1982) ou Peters et al (1982).

---

(1) Hoaglin et Welsch (1978)

(2) voir ausssi Deniau, C, dans Astérique (1977)

Notons pour conclure, que toutes les généralisations possibles des M-estimateurs, minimisent (2.1) et utilisent uniquement la fonction objectif de Huber (1.3). Ces estimateurs consituent en fait une généralisation de l'estimateur de Huber.

II.2. Les WG-estimateurs

Nous proposons d'étendre l'approche de Holland et Welsch (1977) des MCRI (moindres carrés repondérés et itérés) aux MG-estimateurs définis en (2.1); nous obtenons des estimateurs robustes qui vont tenir compte des observations aberrantes sur les variables endogènes et exogènes du modèle, estimateurs que nous appellerons WG-estimateurs. L'intérêt de cette généralisation, c'est qu'on ne se limitera plus à la fonction de Huber d'une part, et qu'on peut calculer ces estimateurs avec un programme de MC pondérés d'autre part.

Si l'on dispose d'une estimation de la dispersion $\tilde{d}$, la condition nécessaire du minimim dans (2.1) s'écrit :

$$\sum_{i=1}^{T} w_i(.)(y_i - \underline{x}_i'\underline{\beta})x_{ij} = 0 \qquad j = 1...,K \tag{2.6}$$

avec

$$w_i(.) = \frac{\psi\left[(y_i - \underline{x}_i'\underline{\beta})/(\tilde{d}\; v(x_i))\right]}{\left[(y_i - \underline{x}_i'\underline{\beta})/(\tilde{d}\; v(x_i))\right]} \tag{2.7}$$

$\psi(.)$ désigne la dérivée d'une fonction objectif quelconque, $v(x_i)$ une fonction de pondération dans l'espace des variables explicatives et $\tilde{d}$ une mesure de dispersion robuste.

Nous pouvons prendre par exemple l'estimateur MSAE, de fonction objectif $\rho(t) = |t|$; cet estimateur est plus robuste que l'estimateur des MC[1] , mais perd en robustesse s'il y a des points aberrants au niveau des variables exogènes du modèle.

Pour éliminer les mauvais effets de ces observations aberrantes on cherchera $\underline{\beta}$ solution de :

$$\text{Min} \sum_{i=1}^{T} \left|y_i - \underline{x}_i'\underline{\beta}\right| \; v(x_i) \tag{2.8}$$

C'est comme si on avait à estimer le modèle suivant :

$$y_i^* = \underline{x}_i' * \underline{\beta} + u_i * \quad i = 1...,T \tag{2.9}$$

avec

$$y_i * = v(x_i).y_i \;,\; \underline{x}_i' * = v(x_i).\underline{x}_i' \;\; et \;\; u_i * = v(x_i).u_i$$

(1) Blattberg et Sargent (1971)

Evidemment la variance des termes d'erreur n'est plus constante dans (2.9); or en présence de points aberrants sur les variables explicatives l'estimation des coefficents est souvent incertaine[1] . En particulier lorsqu'il n'y a pas de points aberrants, et si les h_{ii} sont distribués uniformément, la fonction de pondération $v(x_i)$ donnée par (2.2) est constante pour $i = 1, ..., T$ et on retrouve l'estimateur MSAE sur le modèle (1.1.).

L'estimateur que nous avons défini en (2.8) est un WG-estimateur de la forme définie en (2.6) avec :

$$w_i(.) \ = \ v(x_i) / \left| y_i \ - \ \underline{x}_i' \underline{\beta} \right| \tag{2.10}$$

Si l'on note $\underline{\tilde{\beta}}$ la solution de (2.8) on estimera l'échelle robuste par :

$$\tilde{d} \ = \ \underset{i}{\text{Min}} \left| y_i \ - \ \underline{x}_i' \underline{\tilde{\beta}} \right| / 0.6745 \tag{2.11}$$

C'est cette estimation de la dispersion que nous retiendrons pour le calcul des WG-estimateurs, d'Andrews et de Fair; nous avons retenu aussi la fonction $v(x_i)$ définie en (2.2); la résolution de (2.6) peut être résumée en cinq étapes.

1ère étape :

Calculer la fonction de pondération dans l'espace des variables explicatives; nous retenons $v(x_i) \ = \ (1 \ - \ h_{ii})^{1/2}$.

2ème étape :

Donner les valeurs initiales de $\underline{\beta}$ et d ; nous retenons $\underline{\tilde{\beta}}$ et $\underline{\tilde{d}}$ solutions de (2.8) et (2.11).

3ème étape :

Calculer la fonction empirique de pondération $W_i^{(0)}$ définie en (2.7) pour chaque observation $i \ = \ 1, ..., T$ en remplaçant $\underline{\beta}$ et d par leurs expression trouvées dans l'étape précédente.

4ème étape :

Utiliser alors un programme de régression pondérée pour calculer une nouvelle estimation de $\underline{\beta}$ soit :

$$\underline{\beta}^{(1)} \ = \ \left(\sum_{i=1}^{T} W_i^{(0)} \underline{x}_i \underline{x}_i' \right)^{-1} \sum_{i=1}^{T} W_i^{(0)} \underline{x}_i y_i$$

5ème étape :

Aller en 3 et ainsi de suite jusqu'à un seuil de convergence approprié.

Nous avons considéré dans notre travail la fonction $v(x_i)$ comme étant une fonction des h_{ii}; notre approche s'applique facilement à d'autres types de fonctions comme celle de Krasker et Welsch (1982) par exemple.

---

(1) Asselin de Beauville et al (1980)

Notons que Peters et al (1982) ont développé une procédure de calcul des M-estimateurs et des MG-estimateurs, avec pour seule fonction objectif, la fonction (1.3) de Huber. Notre généralisation des W-estimateurs conviendra donc à toutes les fonctions robustes, Andrews, Fair ....

III. Résultats numériques

Nous reprenons le modèle de Henderson et Velleman (1981), qui explique la mortalité dans 60 villes américaines. Les variables explicatives retenues sont, la moyenne annuelle des précipitations RAIN, l'éducation EDUC, la densité de la population POPDEN, le pourcentage de la population non noire NONW et le logarithme d'une variable mesurant la pollution LOGS02. Les données se trouvent dans l'ouvrage de Hoaglin Mosteller et Tukey (1983).

Nous comparerons les estimateurs de Huber, d'Andrews et de Fair dans un premier temps et leurs généralisations, les MG-estimateurs et les WG-estimateurs dans un second temps.

III.1 M-estimateurs et W-estimateurs

Le tableau 2, présente les estimateurs de Huber (H=1.345), Andrews ($c_1 = 1.339$), et Fair ($c_2 = 1.4$). Les chiffres entre parenthèses sont les seuils à partir desquels l'estimateur pondère les observations; ces valeurs assurent une perte d'efficacité de 5 % si la distribution de l'erreur dans le modèle est gaussienne (cf Holland et Welsch 1977).

Le calcul de ces estimateurs nécessite l'intégration d'un facteur d'échelle d (inconnu); nous considérons alors deux cas :

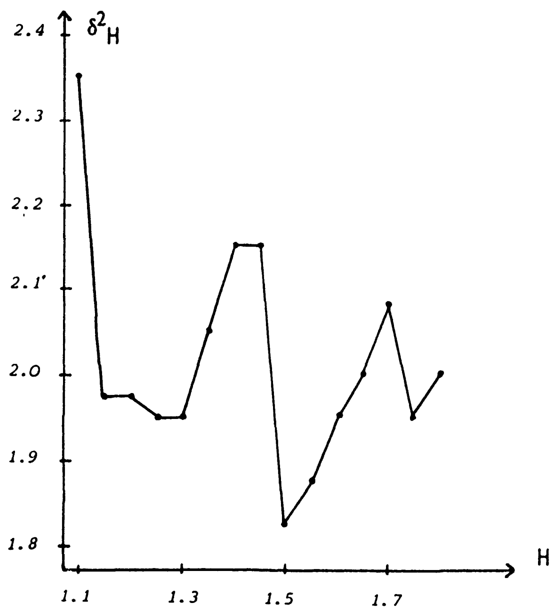
- L'échelle robuste est calculée à partir du vecteur résiduel de l'estimateur MSAE; nous noterons ces estimateurs par Andrews 1, Huber 1 et Fair 1.
- Dans le second cas l'échelle robuste est calculée à partir du vecteur résiduel obtenu à la convergence de l'estimateur de Huber solution de (1.6); ces estimateurs seront notés par Andrews 2, Huber 2 et Fair 2. Tous les estimateurs du tableau 2 ont été calculés par l'approche des MCRI.

Pour comparer ces estimateurs, nons considérons trois critères : la sensibilité des estimateurs robustes à l'échelle, l'écart type des coefficients estimés et une mesure de dispersion robuste RS qui est la médiane de l'écart absolu par rapport à la médiane du vecteur résiduel à la convergence (cf formule (1.13)). Nous pouvons alors constater à partir du tableau 2 la forte sensibilité de l'estimateur d'Andrews à l'échelle (Andrews 1 et Andrews 2), la moyenne sensibilité de l'estimateur de Huber (Huber 1 et Huber 2); l'estimateur de Fair est le plus stable (Fair 1 et Fair 2). Si l'on note $>$ une relation de préférence qui tient compte de ces trois critères à la fois, nous aurons : Fair $>$ Huber $>$ Andrews.

TABLEAU 2
Coefficients estimés W-estimateurs

Estimateurs	cste	NONW	EDUC	POPDEN	RAIN	LOGS02	RS
MC	930.086 (96.24)	3.345 (0.59)	−13.28 (6.98)	0.0028 (0.0037)	1.64 (0.62)	13.80 (3.82)	–
Andrews1 $c_1 = 1.339$	938.750 (67.73)	2.39 (0.40)	−16.18 (5.00)	0.009 (0.003)	2.07 (0.415)	12.44 (2.92)	29.72
Huber1 $H = 1.345$	922.99 (77.36)	2.91 (0.48)	−13.94 (5.67)	0.004 (0.003)	1.93 (0.48)	14.58 (3.18)	27.86
Fair1 $c_2 = 1.4$	925.47 (75.75)	3.09 (0.47)	−13.99 (5.52)	0.0041 (0.003)	1.85 (0.47)	14.66 (3.16)	27.99
Andrews2 $c_1 = 1.339$	897.24 (78.45)	2.61 (0.48)	−11.35 (5.73)	0.002 (0.003)	1.86 (0.48)	17.93 (3.24)	29.55
Huber2 $H = 1.345$	915.69 (83.57)	2.84 (0.51)	−12.93 (6.10)	0.004 (0.0034)	1.87 (0.52)	14.91 (3.40)	26.77
Fair2 $c_2 = 1.4$	924.20 (78.70)	3.10 (0.49)	−13.73 (5.73)	0.0038 (0.003)	1.83 (0.49)	14.67 (3.26)	28.28

Nous avons aussi considéré pour le cas de la fonction de Huber de lier la constante H au degré de perturbation dans l'échantillon; nous avons alors étudié les variations de la variance des résidus (cf. formule 1.22) en faisant varier H de 1.1 à 1.8, nous avons alors obtenu le graphique suivant :



Nous pouvons alors constater que pour ce modèle, la variance estimée de l'estimateur de Huber est plus faible pour H=1.5 que pour H=1.345.

III.2 Les MG-estimateurs et les WG-estimateurs

Dans la classe des MG estimateurs, nous comparerons les estimateurs de Schweppe (H=1.417) et l'estimateur de Krasker Welsch (H=4.2). Concernant le premier estimateur, la valeur de H assure une perte d'efficacité de 5 % , lorsque la distribution des erreurs est gaussienne et lorsque les h_{ii} sont uniformément distribués. Pour le second estimateur, nous garderons la valeur de H=4.2 proposée par Krasker et Welsch (1982) pour ce modèle.

Par ailleurs, nous avons proposé dans ce travail (cf II.2), une généralisation des W estimateurs. Nous comparerons alors les estimateurs, Huber G(H=1.417), Andrews G($c_1 = 1.339$) et Fair G($c_2 = 1.4$).L'indice G indique que ces estimateurs appartiennent à la classe des WG estimateurs.

TABLEAU 3
Coefficients estimés MG-estimateurs et WG-estimateurs

Estimateurs	cste	NONW	EDUC	POPDEN	RAIN	LOGS02	RS
Schweppe $H = 1.417$	917.99 (91.09)	2.77 (0.52)	−13.37 (7.05)	0.005 (0.006)	1.90 (0.49)	14.33 (5.16)	26.93
Krasker – Welsch $H = 4.2$	915.04 (81.09)	2.60 (0.68)	−13.66 (6.13)	0.007 (0.005)	2.01 (0.44)	13.63 (4.10)	28.58
AndrewsG $c_1 = 1.339$	925.12 (73.56)	2.42 (0.44)	−15.60 (5.44)	0.008 (0.003)	1.96 (0.45)	13.76 (3.13)	30.93
HuberG $H = 1.417$	917.105 (82.14)	2.75 (0.49)	−13.48 (6.02)	0.006 (0.004)	1.92 (0.51)	13.92 (3.38)	27.11
FairG $c_2 = 1.4$	925.12 (77.02)	3.04 (0.48)	−14.11 (5.63)	0.005 (0.003)	1.86 (0.48)	14.00 (3.23)	27.97

Nous avons dans un premier temps, utilisé le logiciel TROLL pour calculer les estimateurs de Schweppe et de Krasker Welsch, ensuite nous avons calculé une mesure de dispersion robuste d (cf formule (1.13)) à partir du vecteur des résidus de l'estimateur de Schweppe. Ce facteur d'échelle sera retenu pour calculer par un programme de MCRI, les WG estimateurs.. Les résultats sont présentés dans le tableau 3 ; les chiffres entre parenthèses représentent les écarts types des coefficients estimés.

Notons que l'écart type des coefficients estimés des MG-estimateurs et des WG-estimateurs n'est pas défini de la même façon. (voir Peters et al (1982) pour l'expression de la matrice des variances covariances des MG estimateurs).

On peut d'abord remarquer que dans la classe des WG-estimateurs, l'estimateur Andrews G nous donne les plus faibles écarts types des coefficients estimés ;

ensuite on trouve l'estimateur Fair G, et enfin en dernière position l'estimateur de Huber G. Or étant donné la forte sensiblilité de l'estimateur d'Andrews à l'échelle (tableau 2), la généralisation de la fonction Fair est la plus intéressante.

Dans la classe des MG-estimateurs, l'estimateur de Krasker Welsch est plus intéressant que l'estimateur de Schweppe; la trace de la matrice des variances covariances est de 6630,631 pour le premier estimateur et de 8374,226 pour le second.

Si l'on compare les résultats d'estimation des W-estimateurs et des WG-estimateurs on peut constater que l'écart type des coefficients estimés a légèrement diminué si l'on compare les résultats d'estimation Andrews 2, Huber 2 et Fair 2 (tableau 2) à Andrews G, Huber G et Fair G (tableau 3) respectivement.

Conclusion

La fonction robuste de Fair semble être intéressante en pratique; celle-ci n'a pas été couramment utilisée dans les applications empiriques comme la fonction de Huber ou d'Andrews. L'intérêt de cette fonction réside dans sa faible sensibilité à l'échelle, et l'écart type des coefficients estimés est plus faible (dans l'exemple traité) pour la fonction Fair que pour la fonction de Huber. Ceci nous a amené à proposer une généralisation des W-estimateurs, car les généralisations possibles des M-estimateurs, (l'estimateur de Schweppe et l'estimateur de Krasker et Welsch) utilisent la fonction robuste de Huber.

Demeure le problème du choix des paramètres; pour ces fonctions robustes nous n'avions considéré que le cas de la fonction de Huber; notre approche adaptative lie le paramètre de la fonction de Huber au degré de perturbation dans l'échantillon. Des expériences de simulation restent nécessaires pour étudier sérieusement les propriétés en petit échantillon de ces estimateurs.

Pour conclure, nous reprenons cette remarque de Tukey (dans Launer et Wilkinson (1979)) "it is perfectly proper to use both classical and robust resistant methods routinely, and only worry when they differ enough to matter. But When they differ you should think HARD".

Le problème que nous rencontrons malheureusement en pratique, c'est que même au niveau des méthodes robustes, les résultats différent sensiblement, selon la fonction robuste choisie, le facteur d'échelle retenu, et les paramètres à fixer par l'utilisateur pour ces fonctions; mais la fonction Fair semble être la plus stable !

Bibliographie

[1] ANDREWS, D.F., BICKEL, P.J., HAMPEL, F.R., HUBER, P.J., ROGERS, W.H., TUKEY, J.W. (1972). "Robust estimates of location : Survey and advances". *Princeton University Press.*

[2] ANDREWS, D.F., (1974) "A robust method for multiple linear regression". *Technometrics 16, 523-531.*

[3] ASSELIN de BEAUVILLE, DOLLA, A. (1980) "Une méthode de protection du modèle linéaire" *RSA, Vol. XXVIII, 2,25-43.*

[4] ASTERISQUE (1977) "Théorie de la robustesse et estimation d'un paramètre". *Société mathématique de France. Vol. 43-44.*

[5] BASSET, J., KOENKER, R. (1978) "Asymptotic theory of least absolute error in regression" *JASA 73, 618-622.*

[6] BELSLEY, D.A., KUH, E., WELSCH, R.E. (1980) "Regression diagnostics : identifying influential data and sources of collinearity". *John Wiley.*

[7] BLATTBERG, R. SARGENT, T. (1971) "Regression with Non-Gaussian stable disturbances : Some sampling results" *Econometrica 39, 501-510.*

[8] CHAFFAI, M.E. (1985) "De l'estimation robuste dans les modèles linéaires : Une nouvelle méthode d'estimation optimale, La Lp Trace" *th. Doct.* 3ème *cycle, Aix Marseille III.*

[9] COOK, R.D. (1977) "Detection of influential observations in linear regression". *Technometrics 19, 15-18.*

[10] COOK, R.D., WEISBERG, S. (1980) "Characterisations of an ampirical influence function for detecting influential cases in regression". *Technometrics* 22, 495-508.

[11] DENBY, L., LARSEN, W. (1977) "Robust regression estimators compared via Monte Carlo" *Communications in Statistics A6*, 335-362.

[12] DUTTER, R. (1977) "Numerical solution of robust regression problems : computational aspects a comparison". *J. Stat. Comp. Simul.*, 5, 207-238.

[13] FOX, J. (1984) "Linear statistical models and related methods". John Wiley.

[14] HAMPEL, F.R. (1974) "The influence curve and its role in robust estimation", *JASA 69*, 383-394.

[15] HANDSCHIN, E., KOHLAS, J., FIECHTER, A., SCHWEPPE, F. (1975) "Bad data analysis for power system state estimation", *IEEE Transactions on Power Apparitus and Systems 2*, 329-337.

[16] HNEDERSON, H.V., VELLEMAN, P.F. (1981) "Building multiple regression models interactively" *Biometrics 37*, 391-411.

[17] HOAGLIN, D.C., WELSCH, R.E. (1978) "The Hat Matrix in regression and ANOVA". *The American Statistician 32*, 17-22.

[18] HOAGLIN, D.C., MOSTELLER, F., TUKEY, J.W. (1983) "Understanding robust and exploratory data analysis" *John Wiley.*

[19] HOGG, R.V. (1979) "Statistical robustness : One view of its use in applications today" *The American Statistician 33*, 108-115.

[20] HOLLAND, F.W., WELSCH, R.E. (1977) "Robust regression using iteratively reweighted least squares". *Communications in Statistical A6*, 813-828.

[21] HUBER, P.J. (1964) "Robust estimation of a location parameter". *Ann. Math. Stat. 37*, 73-101.

[22] HUBER, P.J. (1972) "Robust statistics : A review". *The Annals of Statistics 45*, 1041-1067.

[23] HUBER, P.J. (1973) "Robust regression asymptotics conjectures and Monte Carlo" *The Annals of Statistics 5*, 799-821.

[24] HUBER, P.J. (1981) "Robust Statistics", *John Wiley.*

[25] KRASKER, W.S. (1980) "Estimation in linear regression models with disparate data points" *Econometrica 48*, 1333-1346.

[26] KRASKER, W.S., WELSCH, R.E. (1982) "Efficient bounded influence regression estimation", *JASA 77*, 595-604.

[27] LAUNER, R., WILKINSON, G.N. (1979) "Robustness in statistics" *New York Academic Press.*

[28] MALINVAUD, E. (1981) "Méthodes statistiques de l'économétrie" *Paris Dunod.*

[29] MONTGOMERY, D.C., PECK, E.A. (1982) "Introduction to linear regression analysis" *John Wiley.*

[30] MORINEAU, A. (1978) "Régression robuste : méthodes d'ajustement et de validation" *RSA*, vol. XXVI, 3, 5-28.

[31] PETERS, S., SAMAROV, A., WELSCH, R.E. (1982) "Computational procedures for bounded influence and robust regression (TROLL : BIFF and BIFMOD)" *Technical report n° 30, MIT.*

[32] SAMAROV, A.M. (1985) "Bounded influence regression via local Minimax MSE" *JASA 80*, 1032-1040.

[33] WEISBERG, R.E. (1980) "Applied Linear regression" *John Wiley.*