M-estimateurs semi-paramétriques dans les modèles à direction révélatrice unique

Michel Delecroix Marian Hristache

Abstract

We consider single-index regression models, for which $E(Y_i|X_i) = \varphi(X_i'\alpha)$. We show the a.s. consistency and asymptotic normality for a class of semiparametric M-estimators of the parameter α. We investigate the efficiency of these estimators. We construct an estimator which achieves the semiparametric asymptotic efficiency bound for these models.

Résumé

Nous considérons des modèles de régression à direction révélatrice unique, pour lesquels $E(Y_i|X_i) = \varphi(X_i'\alpha)$. Nous montrons la convergence presque sûre et la normalité asymptotique d'une classe de M-estimateurs semi-paramétriques du paramètre α. Nous étudions l'efficacité de ces estimateurs. Nous construisons un estimateur qui atteint la borne d'efficacité asymptotique dans ces modèles.

1 Introduction

Nous étudions dans ce papier les propriétés d'une classe d'estimateurs dans les modèles dits "à direction révélatrice unique" ("single-index models"). Les estimateurs sont construits par maximisation d'une fonction objectif Ψ (M-estimateurs). La méthode proposée permet d'uniformiser et généraliser des résultats antérieurs de Ichimura (1993) (moindres carrés semi-paramétriques dans les "single-index models"), Klein et Spady (1993) (quasi-maximum de vraisemblance dans les modèles dichotomiques)

Received by the editors January 1997.

Communicated by M. Hallin.

Key words and phrases : semiparametric models, M-estimators, efficiency bounds, single-index models, pseudo-maximum likelihood.

et Bonneu, Delecroix et Hristache (1995) (quasi-maximum de vraisemblance dans les modèles linéaires généralisés semi-paramétriques).

Dans les modèles "à direction révélatrice unique", on suppose les observations constituées d'une suite de variables aléatoires (X_i', Y_i) à valeurs dans $\mathbb{R}^d \times \mathbb{R}$, indépendantes et de même loi, et l'existence d'une "direction révélatrice", c'est-à-dire un élément α d'un sous-ensemble Θ^* de $\mathbb{R}^d$ tel que :

$$E\left(Y_i|X_i = x\right) = E\left(Y_i|X_i'\alpha = x'\alpha\right) = \varphi\left(x'\alpha\right), \tag{1}$$

ce qui peut encore s'écrire

$$Y_i = \varphi\left(X_i'\alpha\right) + \varepsilon_i, \qquad E\left(\varepsilon_i|X_i\right) = 0 \tag{2}$$

($x'\alpha$ représente évidemment le produit scalaire usuel dans $\mathbb{R}^d$ des vecteurs x et α).

Comme plusieurs auteurs l'ont remarqué, dans un modèle défini par (1) il n'est possible d'identifier que la direction de α, c'est-à-dire que α est connu à une constante près. Pour resoudre ce problème, on va contraindre la dernière composante de α à être égale à 1, comme dans Sherman (1994). Par conséquent, Θ^*, l'espace des paramètres possibles, va être identifié à un sous-ensemble compact Θ de $\mathbb{R}^{d-1}$: $\Theta^* = \left\{(\theta, 1) \ : \ \theta \in \Theta \subset \mathbb{R}^{d-1}\right\}$.

En fait, la forme la plus générale des "single-index models" (cf. Newey (1990) ou Ichimura (1993)) suppose que l'équation (1) s'écrive sous la forme

$$E\left(Y_i|X_i = x\right) = \varphi\left(v\left(x, \alpha\right)\right), \tag{3}$$

où $v(x,\alpha)$ est une fonction connue (appelée "aggregator" ou "index"). Le choix $v(x,\alpha) = x'\alpha$ (le cas le plus fréquent en pratique) nous permettra une écriture plus simple des formules, qui restent valables dans le cas général, sous certains hypothèses sur v. Nous l'avons donc adopté.

Les exemples usuels de modèle à direction révélatrice unique sont fournis par les modèles CDR ("conditional distribution regression", cf. Newey (1990)), caractérisés par le fait que la distribution conditionnelle de Y_i sachant X_i ne dépende que de $X_i'\alpha$ (ou, plus généralement, de $v(X_i,\alpha)$). Ainsi, le modèle dichotomique classique

$$Y_i = \begin{cases} 1 & \text{si} \quad u_i \geq X_i'\alpha, \\ 0 & \text{sinon,} \end{cases} \tag{4}$$

où les erreurs u_i sont indépendantes des régresseurs X_i, est-il un modèle CDR puisque, dans ce cas, la distribution conditionnelle de $Y_i|X_i$ est la loi $\mathcal{B}\left(1, E\left(Y_i|X_i'\alpha\right)\right)$.

De même, le modèle Tobit censuré

$$Y_i = \begin{cases} X_i'\alpha + u_i & \text{si} \quad u_i \geq -X_i'\alpha, \\ 0 & \text{sinon,} \end{cases} \tag{5}$$

(u_i et X_i indépendants), et le modèle Tobit tronqué, où l'on n'observe Y_i et X_i que si $u_i \geq -X_i'\alpha$, constituent-ils encore deux exemples de modèles qui vérifient (1).

Parmi les modèles CDR, on a beaucoup étudié les modèles linéaires généralisés au sens de McCullagh et Nelder (1989). Ces auteurs supposent exactement que la densité conditionnelle de Y_i sachant que $X_i = x$ appartient à la famille exponentielle, et qu'elle ne dépend de x qu'à travers $\varphi(x'\alpha)$, φ étant connue. Sans souci d'exhaustivité on peut citer comme référence sur le sujet, outre la synthèse de McCullagh et Nelder (1989), celle de Gouriéroux (1989) pour les modèles à variables qualitatives. Le modèle que nous utiliserons généralise évidemment celui de McCullagh et Nelder, puisqu'il ne fixe plus la forme de la densité conditionnelle, ni celle de φ.

Pour estimer le paramètre α intervenant en (2), plusieurs approches ont été successivement proposées. Dans le cas où φ est connue, Gouriéroux, Monfort et Trognon (1984) ont montré que l'estimateur défini par

$$\widetilde{\theta}_n = \arg\max_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} \log f\left(Y_i, \varphi\left(X_i'\theta\right)\right) \tag{6}$$

converge vers α même si f n'est pas la densité conditionnelle de $Y_i|X_i$, à condition que la loi de densité f appartienne à la famille exponentielle linéaire :

$$f(y,r) = \exp\left[A(r) + B(y) + C(r)y\right], \tag{7}$$

r étant la moyenne de cette loi. Dans le cas particulier où $f(y,r) = \exp(-\frac{1}{2}r^2 - \frac{1}{2}y^2 + ry)$, on obtient l'estimateur des moindres carrés.

Lorsque φ est inconnue, quand la loi conditionnelle de $Y_i|X_i = x$ est binomiale (cas du modèle dichotomique), Klein et Spady (1993) ont proposé un estimateur de α basé sur la maximisation d'une vraisemblance où $E\left(Y_i|X_i'\theta\right)$ est remplacé par une estimation non paramétrique. Bonneu, Delecroix et Hristache (1995) ont généralisé cette démarche au cas des modèles CDR pour lesquels la densité conditionnelle de $Y_i|X_i = x$ s'écrit sous la forme $f\left(y, r_\alpha\left(x'\alpha\right)\right)$, avec

$$r_\theta\left(t\right) = E\left(Y_i|X_i'\theta = t\right), \quad \theta\in\Theta, \tag{8}$$

f étant connue (cas des modèles linéaires généralisés). L'estimateur proposé, dont ils démontrent la convergence presque sûre et la normalité asymptotique, est défini par

$$\widehat{\theta}_n = \arg\max_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} \log f\left(Y_i, \widehat{r}_\theta\left(X_i'\theta\right)\right), \tag{9}$$

où $\widehat{r}_\theta\left(X_i'\theta\right)$ est l'estimateur du noyau de Nadaraya-Watson de la régression $r_\theta\left(X_i'\theta\right)$ de Y_i sur $X_i'\theta$:

$$\widehat{r}_\theta\left(X_i'\theta\right) = \left\{\sum_{j\neq i} Y_j K\left(\frac{X_i'\theta - X_j'\theta}{a_n}\right)\right\} / \left\{\sum_{j\neq i} K\left(\frac{X_i'\theta - X_j'\theta}{a_n}\right)\right\}. \tag{10}$$

De même, Ichimura (1993) a étudié un estimateur de α basé sur la minimisation de $\sum_{i=1}^{n}\left[Y_i - E\left(Y_i|X_i'\theta\right)\right]^2$:

$$\widetilde{\theta}_n = \arg\min_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} \left[Y_i - \widehat{r}_\theta\left(X_i'\theta\right)\right]^2. \tag{11}$$

Pour unifier les idées des méthodes précédentes, on est amené naturellement à considérer pour le modèle à direction révélatrice unique donné par (1), les estimateurs définis par :

$$\widehat{\theta}_n = \arg\max_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} \Psi\left(Y_i, \widehat{r}_\theta\left(X_i'\theta\right)\right), \tag{12}$$

où Ψ est une fonction vérifiant certaines conditions de régularité, et $\widehat{r}_\theta\left(X_i'\theta\right)$ est donné par la relation (10). L'idée est de maximiser $\frac{1}{n}\sum\limits_{i=1}^{n}\Psi\left(Y_i, E\left(Y_i|X_i'\theta\right)\right)$ par rapport à θ, en substituant à $E\left(Y_i|X_i'\theta\right)$ une estimation obtenue à partir des observations. L'étude de ce type d'estimateurs "semi-paramétriques" est l'objet de ce papier.

Une telle généralisation des estimateurs de Ichimura (1993) et Klein et Spady (1993) a déjà été considérée par Sherman (1994) et Newey (1994). En utilisant des inégalités maximales pour des U-processus, Sherman (1994) obtient la loi asymptotique des estimateurs définis par (12), sous des conditions assez lourdes concernant la fonction Ψ. Sous les hypothèses qu'il utilise, on peut également appliquer les résultats plus généraux de Newey (1994) qui remarque, entre autres, que toute fonction objectif Ψ égale à la log-vraisemblance d'une densité du modèle exponentiel linéaire peut être utilisée pour définir un estimateur convergent de α défini par (1) dans la mesure où, pour une telle fonction, on peut écrire :

$$\alpha = \arg\max_{\theta\in\Theta} E\left[\Psi\left(Y_i, E\left(Y_i|X_i'\theta\right)\right)\right]. \tag{13}$$

Le deuxième paragraphe de ce papier est dès lors consacré à la question qui se pose naturellement compte tenu de la remarque précédente : existe-t-il d'autres fonctions Ψ vérifiant la même propriété ? Peut-on caractériser l'ensemble des fonctions Ψ qui la vérifie ? Nous verrons que sous une condition un peu plus restrictive que (13) mais déjà introduite par Sherman (1994), les seules fonctions Ψ acceptables en un certain sens sont bien les log-vraisemblances d'une densité du modèle exponentiel linéaire.

Dans le troisème paragraphe nous énoncerons les théorèmes de convergence de $\hat{\theta}_n$ vers α, pour les fonctions Ψ choisies selon ce critère.

Les estimateurs obtenus n'étant pas forcément efficaces, le quatrième paragraphe du papier sera consacré à une discussion de ce problème : nous montrerons comment on peut modifier l'estimateur donné par (12) pour obtenir un estimateur efficace de α, ce qui est, à notre connaissance, nouveau. Les démonstrations des résultats obtenus se trouvent dans l'appendice.

Par souci d'exhaustivité, on peut enfin noter, quoique ceci sorte du cadre de cette étude, que des estimateurs purement non paramétriques de α ont été développés par Stoker (1986), Powell, Stock et Stoker (1989), Härdle et Stoker (1989), notamment dans le cadre des modèles dichotomiques. Pour une comparaison avec un estimateur semi-paramétrique voir Bonneu, Delecroix et Malin (1993).

2 Choix des fonctions Ψ

Pour obtenir la convergence presque sûre de $\widehat{\theta}_n$ vers α, on doit démontrer essentiellement que :

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}\Psi\left(Y_i, \widehat{r}_\theta\left(X_i'\theta\right)\right) - E\left[\Psi\left(Y_i, r_\theta\left(X_i'\theta\right)\right)\right]\right| \xrightarrow{p.s.} 0. \tag{14}$$

(cf. Klein et Spady (1993), Ichimura (1993), Sherman (1994), Bonneu, Delecroix et Hristache (1995)). Le résultat est obtenu en supposant que :

H1 Les variables (X_i', Y_i) sont indépendantes, de même loi.

H2 Θ est un compact de $\mathbb{R}^d$.

On peut alors conclure si, en plus, on suppose que $E\left[\Psi\left(Y_i, r_\theta\left(X_i'\theta\right)\right)\right]$ atteigne son maximum en α défini par (1). Cela amène à imposer les hypothèses suivantes :

H3 Il existe un unique élément α appartenant à l'intérieur de Θ qui vérifie (1).

H4 α est une solution du problème de maximisation

$$\max_{\theta\in\Theta} E\left[\Psi\left(Y_i, E\left(Y_i|X_i'\theta\right)\right)|X_i = x\right], \tag{15}$$

pour tout x dans le support de X_i.

Le théorème 4.1 de Ichimura (1993) donne des conditions suffisantes pour que l'hypothèse H3 soit vérifiée. H4 assure que le problème limite $\max\limits_{\theta\in\Theta} E\left[\Psi\left(Y_i, r_\theta\left(X_i'\theta\right)\right)\right]$ a une solution (unique) donnée par $\theta = \alpha$, ce qui permet d'estimer α par un M-estimateur défini en (12), c'est-à-dire que la limite de $\widehat{\theta}_n$ est égale à la direction révélatrice α définie en (1).

Comme le paramètre d'intérêt n'intervient que dans la loi conditionnelle de Y_i sachant X_i, la loi marginale de X_i n'apporte aucune information supplémentaire sur θ. Il semble donc naturel d'imposer que α soit solution du problème limite pour la loi conditionnelle de Y_i sachant $X_i = x$, pour tout x. Cela est assuré par l'hypothèse H4. Notons que cette condition est trivialement vérifiée pour les fonctions Ψ utilisées dans Ichimura (1993) et Klein et Spady (1993). Elle l'est aussi, plus généralement, pour tout estimateur de la forme (12) où Ψ est la log-vraisemblance d'une densité exponentielle linéaire (voir Newey (1994)).

Sous les conditions générales H1-H4 et des conditions techniques additionnelles, on peut alors démontrer le théorème de convergence p.s. de $\widehat{\theta}_n$ vers α. Sa preuve exige cependant d'imposer à Ψ un lourd ensemble de conditions. Il apparaît que ces conditions peuvent être largement simplifiées quand Ψ est la log-vraisemblance d'une densité exponentielle linéaire. Or ce choix ne fait perdre, *a priori*, aucune généralité au résultat obtenu, si l'on considère la proposition qui suit, inspirée d'un résultat similaire de Gouriéroux, Monfort et Trognon (1984) (voir aussi la proposition 8.52 de Gouriéroux et Monfort (1989)) :

Proposition 1. *Soit* $\Psi : \mathbb{R}^2 \longrightarrow \mathbb{R}$ *une fonction telle que* $\partial_2\Psi(y,r)$ *existe pour tout* $(y,r) \in \mathbb{R}^2$. *Alors l'hypothèse H4 est vérifiée pour toute loi du couple* (Y_i, X_i) *satisfaisant à la condition* (1) *si et seulement s'il existe deux fonctions dérivables* A, $C : \mathbb{R} \longrightarrow \mathbb{R}$ *vérifiant :*

i) $A'(r) + C'(r)\, r \equiv 0$,

ii) $C'(r) \geq 0, \quad \forall r \in \mathbb{R}$,

telles que

$$\Psi(y,r) = A(r) + C(r)\, y + B(y), \tag{16}$$

où $B : \mathbb{R} \longrightarrow \mathbb{R}$ *est une fonction quelconque.*

Preuve :

Condition suffisante. Supposons l'hypothèse H4 vraie pour pour toute loi de (Y_i, X_i). Choisissons $X_i \sim N(0, I_d)$ et $Y_i = X_i'\alpha + \varepsilon_i$, où ε_i vérifie $E(\varepsilon_i|X_i) = 0$. On a $E(Y_i|X_i'\theta) = E[E(Y_i|X_i)|X_i'\theta] = E(X_i'\alpha|X_i'\theta) = \dfrac{\alpha'\theta}{\|\theta\|^2}X_i'\theta$, donc α est solution du problème

$$\max_{\theta\in\Theta} E\left[\Psi\left(Y_i, \frac{\alpha'\theta}{\|\theta\|^2}x'\theta\right)|X_i = x\right],$$

pour tout $x \in \mathbb{R}^d$. Mais α est un point intérieur de Θ et Ψ est dérivable par rapport au deuxième argument, donc

$$E\left[\frac{\partial}{\partial\theta}\Psi\left(Y_i, \frac{\alpha'\theta}{\|\theta\|^2}x'\theta\right)\bigg|_{\theta=\alpha}|X_i = x\right] = 0, \quad \forall x \in \mathbb{R}^d,$$

ce qui donne

$$E[\partial_2\Psi(Y_i, x'\alpha)|X_i = x]\left(x - \alpha\frac{x'\alpha}{\|\alpha\|^2}\right) = 0, \quad \forall x \in \mathbb{R}^d,$$

et par conséquent

$$E[\partial_2\Psi(Y_i, x'\alpha)|X_i = x] = 0, \quad \forall x \neq c\alpha \ (c \in \mathbb{R}).$$

En particulier, cela est vrai si la loi de $Y_i|X_i = x$ est discrète, le support étant constituée de deux points $y_1(x'\alpha)$ et $y_2(x'\alpha)$ tels que $y_1(r) < r = x'\alpha < y_2(r)$. Alors,

$$\begin{cases} p_1(r) + p_2(r) = 1 \\ p_1(r)y_1(r) + p_2(r)y_2(r) = r \end{cases} \Rightarrow\ p_1(r)\partial_2\Psi(y_1(r), r) + p_2(r)\partial_2\Psi(y_2(r), r) = 0,$$

pour tout $y_1(r) < r < y_2(r)$, ce qui entraîne :

$$(y_2(r) - r)\partial_2\Psi(y_1(r), r) + (r - y_1(r))\partial_2\Psi(y_2(r), r) = 0, \qquad \forall\, y_1(r) < r < y_2(r).$$

On obtient donc

$$\frac{\partial_2\Psi(y_1(r), r)}{y_1(r) - r} = \frac{\partial_2\Psi(y_2(r), r)}{y_2(r) - r}, \quad \begin{cases} \forall\ r \in \mathbb{R}, \\ \forall\ y_1(r) < r < y_2(r). \end{cases}$$

On en déduit l'existence, pour tout $r \in \mathbb{R}$, d'un réel $\lambda = \lambda(r)$ tel que $\partial_2\Psi(y, r) = \lambda(r)(y - r)$ pour tout $y \in \mathbb{R}$. Il suffit alors de choisir les fonctions A et C telles que $A'(r) = -\lambda(r)r$ et $C'(r) = \lambda(r)$. Le fait que $E[\Psi(Y_i, r)|X_i = x]$ ait un maximum en $r = x'\alpha$ entraîne $\lambda(r) \geq 0$, donc les conditions i) et ii) sont satisfaites.

Condition nécessaire. Soit $\Psi(y, r) = A(r) + C(r)y + B(y)$, avec A et C vérifiant i) et ii). Une conséquence immédiate de ces deux conditions est que la fonction

$$r \mapsto A(r) + C(r)s$$

a un maximum en $r = s$. Supposons que $E(Y_i|X_i) = E(Y_i|X_i'\alpha)$. Alors

$$\begin{aligned} &E[\Psi(Y_i, E(Y_i|X_i'\theta))|X_i = x] \\ &= A(E(Y_i|X_i'\theta = x'\theta)) + C(E(Y_i|X_i'\theta = x'\theta))E(Y_i|X_i'\alpha = x'\alpha) + E(B(Y_i|X_i = x)) \\ &\leq A(E(Y_i|X_i'\alpha = x'\alpha)) + C(E(Y_i|X_i'\alpha = x'\alpha))E(Y_i|X_i'\alpha = x'\alpha) + E(B(Y_i|X_i = x)), \end{aligned}$$

donc α est solution du problème de maximisation (15). ∎

La proposition précédente montre que si Ψ correspond à une log-vraisemblance déduite d'une densité du modèle exponentiel linéaire, sous l'unique condition que le modèle vérifie (1), H4 est vérifiée, ce qui permet l'estimation effective de la direction révélatrice α. Si on ne dispose pas d'information supplémentaire sur le modèle défini par (1), au niveau de la loi des (X_i, Y_i), il est de plus naturel de choisir Ψ qui vérifie H4 *quelle que soit* cette loi, ce qui amène d'après la proposition 1 à se limiter au seul choix des fonctions Ψ correspondant à la log-vraisemblance d'une densité exponentielle linéaire. Nous démontrerons donc le théorème de convergence pour ce choix particulier de la fonction Ψ, qui de plus exige peu d'hypothèse techniques supplémentaires.

D'autres choix de fonctions Ψ sont cependant possibles, qui permettent de vérifier H4, pour certaines lois particulières du couple (X_i, Y_i) (mais évidemment pas pour toutes). Si donc on dispose d'une indication particulière sur la loi des observations du modèle étudié, on peut bien sûr déterminer ces autres fonctions Ψ "cas par cas". On pourra voir Sherman (1994) pour cette approche (par exemple si la loi de Y_i est symétrique). Evidemment, pour ces choix "non universels" de Ψ, on pourra obtenir par nos techniques de démonstration le même type de théorème général de convergence de $\widehat{\theta}_n$ que Sherman (1994), sous des conditions plus faibles que les siennes (*cf.* remarques du paragraphe III). L'intérêt de ces choix est qu'ils incluent une information supplémentaire sur le modèle (la connaissance de la forme de la vraie densité conditionnelle, par exemple), ce qui permet comme nous le verrons d'améliorer leurs efficacité asymptotique.

Remarquons enfin (preuve de la proposition précédente) qu'en fait on peut remplacer l'hypothèse H4 par l'hypothèse suivante :

H4' α donnée par l'hypothèse H3 vérifie

$$E\left[\partial_2 \Psi\left(Y_i, E\left(Y_i | X_i'\alpha\right)\right) | X_i'\alpha\right] = 0. \tag{17}$$

L'hypothèse 11(i) de Sherman (1994), qui joue un rôle clé dans sa démarche, demande que la relation (17) soit vérifiée pour tout $\theta \in \Theta$, et apparaît donc plus forte que H4. En fait, cette condition équivaut à H4', lorsque Ψ a la forme donnée en (16) : l'équation (17) est alors vérifiée pour tout $\theta \in \Theta$. En conclusion, les hypothèses 11(i) de Sherman (1994), H4 et H4' sont équivalentes si on les suppose vérifiées *pour toute loi* du couple (Y_i, X_i), et chacune d'entre elles implique l'hypothèse 3 de Sherman (1994), à savoir que α défini par (1) est solution du problème de maximisation

$$\max_{\theta \in \Theta} E\left[\Psi\left(Y_i, E\left(Y_i | X_i'\theta\right)\right)\right]. \tag{18}$$

La discussion sur le choix de Ψ achevée, nous pouvons passer à la démonstration des théorèmes de convergence de $\widehat{\theta}_n$ défini par (12) et (10), dans le cas où Ψ est donnée par (16).

3 Théorèmes de convergence

Pour obtenir (14) il est nécessaire de démontrer d'abord la convergence des estimateurs $\widehat{r}_\theta\left(X_i'\theta\right)$ vers $r_\theta\left(X_i'\theta\right)$ uniformément en θ et en X_i, ainsi que celles de leurs dérivées. C'est l'objet du lemme qui suit, dont la démonstration figure dans Bonneu, Delecroix et Hristache (1995). Le résultat est obtenu sous les hypothèses qui suivent :

H5 Soit $h_\theta(t)$ la densité de $X_i'\theta$. On suppose que le support S de X vérifie :

1) S est compact ;
2) $\inf\limits_{x\in S} h_\theta(x'\theta) > 0$;
3) $h_\theta(t)$ et $r_\theta(t) = E(Y_i \,|X_i'\theta = t)$ sont trois fois continûment différentiables par rapport à t, et les dérivées d'ordre trois sont lipschitziennes sur $T_\theta(S) = \{t \in \mathbb{R};\ t = x'\theta,\ x \in S\}$, uniformément par rapport à $\theta \in \Theta$;
4) $R(x,\theta) = r_\theta(x'\theta)$ est une fonction deux fois continûment différentiable par rapport à θ sur $S \times \Theta$;
5) $D(x,\theta) = h_\theta(x'\theta)$ est continue sur $S \times \Theta$.

H6 1) $E(|Y_i|^m) < \infty,\ m \geq 4$;

2) $var(Y_i \,|X_i = x)$ est bornée et bornée inférieurement par une constante strictement positive sur S.

H7 – $K(s) = \dfrac{1}{\sqrt{2\pi}} \exp\left(-\dfrac{s^2}{2}\right)$, $s \in \mathbb{R}$, ou bien

– 1) $K : \mathbb{R} \to \mathbb{R}$ est deux fois continûment différentiable et K'' est lipschitzienne ;
2) $\int K(s)\,ds = 1$;
3) $\int sK(s)\,ds = 0$;
4) $\mathrm{supp}K \subseteq [-1,1]$.

H8 La suite $\{a_n\}_{n=1}^{\infty}$ est choisie telle $a_n = cn^{-\gamma}$, avec $c > 0$ et $\gamma \in \left(\frac{1}{8}, \frac{1}{7}\right)$.

On obtient :

Lemme 1. *(Bonneu, Delecroix et Hristache (1995))*
Sous les hypothèses H1–H2 et H5-H8, on a

$$\sup_{(x,\theta)\in S\times\Theta} |\widehat{r}_\theta(x'\theta) - r_\theta(x'\theta)| \underset{n\to\infty}{\overset{p.s.}{\to}} 0, \tag{19}$$

$$n^{1/4} \cdot \sup_{(x,\theta)\in S\times\Theta} |\widehat{r}_\theta(x'\theta) - r_\theta(x'\theta)| \underset{n\to\infty}{\overset{P}{\to}} 0, \tag{20}$$

$$n^{1/4} \cdot \sup_{(x,\theta)\in S\times\Theta} \left| \frac{\partial \widehat{r}_\theta(x'\theta)}{\partial\theta} - \frac{\partial r_\theta(x'\theta)}{\partial\theta} \right| \underset{n\to\infty}{\overset{P}{\to}} 0, \tag{21}$$

$$\sup_{(x,\theta)\in S\times\Theta} \left| \frac{\partial^2 \widehat{r}_\theta(x'\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2 r_\theta(x'\theta)}{\partial\theta\partial\theta'} \right| \underset{n\to\infty}{\overset{P}{\to}} 0. \tag{22}$$

■

Posons

$$r_1 = \inf_{(x,\theta)\in S\times\Theta} r_\theta(x'\theta) \quad \text{et} \quad r_2 = \sup_{(x,\theta)\in S\times\Theta} r_\theta(x'\theta).$$

$\Psi : \mathbb{R} \times [r_1, r_2] \to \mathbb{R}$ étant de la forme

$$\Psi(y, r) = A(r) + C(r)\, y, \tag{23}$$

nous supposons que A et C vérifient l'hypothèse technique :

H9 1) A, $C : [r_1, r_2] \to \mathbb{R}$ sont deux fois continûment différentiables ;
2) $A'(r) + C'(r)\, r = 0$, $\forall r \in [r_1, r_2]$;
3) $C'(r) > 0$, $\forall r \in [r_1, r_2]$;
4) A'' et C'' satisfont à une condition de Lipschitz.

On obtient le théorème :

Théorème 1. *Sous les hypothèses H1–H9, $\widehat{\theta}_n$ défini par*

$$\widehat{\theta}_n = \arg\max_{\theta\in\Theta} \frac{1}{n} \sum_{i=1}^{n} \left[A\left(\widehat{r}_\theta\left(X_i'\theta\right)\right) + C\left(\widehat{r}_\theta\left(X_i'\theta\right)\right) Y_i\right] \tag{24}$$

converge presque sûrement vers α.

Preuve : Il s'agit d'une adaptation directe des resultats de Bonneu, Delecroix et Hristache (1995). ■

Pour obtenir la normalité asymptotique de $\widehat{\theta}_n$, on supposera enfin vérifiée une hypothèse habituelle de régularité :

H10 La matrice

$$M = E\left[C'\left(\varphi\left(X_i'\alpha\right)\right) \left.\frac{\partial r_\theta\left(X_i'\theta\right)}{\partial\theta}\right|_{\theta=\alpha} \left.\frac{\partial r_\theta\left(X_i'\theta\right)}{\partial\theta'}\right|_{\theta=\alpha}\right] \tag{25}$$

est définie positive.

Cela permet d'énoncer :

Théorème 2. *Sous les hypothèses H1–H10*

$$\sqrt{n}\left(\widehat{\theta}_n - \alpha\right) \xrightarrow{L} N(0, \Sigma), \tag{26}$$

avec $\Sigma = M^{-1} V M^{-1}$, *où*

$$V = E\left\{\left[C'\left(\varphi\left(X_i'\alpha\right)\right)\right]^2 \Omega(X_i) \left.\frac{\partial r_\theta\left(X_i'\theta\right)}{\partial\theta}\right|_{\theta=\alpha} \left.\frac{\partial r_\theta\left(X_i'\theta\right)}{\partial\theta'}\right|_{\theta=\alpha}\right\}, \tag{27}$$

et $\Omega(X_i) = var(Y_i|X_i)$.

Preuve : Voir appendice. ■

Remarques

1. Un cas particulier d'utilisation des théorèmes 1 et 2 correspond au cas de l'estimateur SLS de Ichimura (1993). On choisit alors pour Ψ une densité gaussienne. On obtient directement :

Corollaire 1. *Si*

$$\widehat{\theta}_n = \arg\min_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n}\left[Y_i - \widehat{r}_\theta\left(X_i'\theta\right)\right]^2, \tag{28}$$

et si les hypothèses H1–H3 et H5-H8 sont vérifiées, alors :
- $\widehat{\theta}_n \xrightarrow{p.s.} \alpha$,
- $\sqrt{n}\left(\widehat{\theta}_n - \alpha\right) \xrightarrow{L} N(0,\Sigma)$, $\Sigma = M^{-1}VM^{-1}$, *où*

$$\begin{aligned} M &= E\left[\left.\frac{\partial r_\theta\left(X_i'\theta\right)}{\partial\theta}\right|_{\theta=\alpha} \left.\frac{\partial r_\theta\left(X_i'\theta\right)}{\partial\theta'}\right|_{\theta=\alpha}\right], \\ V &= E\left\{\left.\frac{\partial r_\theta\left(X_i'\theta\right)}{\partial\theta}\right|_{\theta=\alpha} \left[var\left(Y_i\left|X_i\right.\right)\right]^{-1} \left.\frac{\partial r_\theta\left(X_i'\theta\right)}{\partial\theta'}\right|_{\theta=\alpha}\right\}. \end{aligned} \tag{29}$$

2. Les théorèmes 1 et 2 restent toujours valables pour une fonction Ψ qui n'a pas la forme donnée en (16), si l'on remplace les hypothèses H9 et H10 par :

H9' 1) Ψ est deux fois continûment différentiable par rapport à r ;

2) Il existe une fonction $F:\mathbb{R}\to(0,\infty)$ telle que

i) $E\left(F^4\left(|Y_i|\right)\right) < \infty$;

ii) $\left|\frac{\partial^2}{\partial r^2}\Psi(y,r) - \frac{\partial^2}{\partial r^2}\Psi(y,r')\right| \le F(|y|)\,|r-r'|$, $\forall r,r'\in[r_1,r_2]$, $\forall y \in \mathbb{R}$;

iii) $\sup_{r\in[r_1,r_2]}\left|\frac{\partial^2}{\partial r^2}\Psi(y,r)\right| \le F(|y|)$;

3) $E\left[\left.\frac{\partial^2}{\partial r^2}\Psi(Y_i,r)\right|_{r=r_\alpha\left(X_i'\alpha\right)} |X_i\right] = E\left[\left.\frac{\partial^2}{\partial r^2}\Psi(Y_i,r)\right|_{r=r_\alpha\left(X_i'\alpha\right)} |X_i\alpha\right]$.

H10' La matrice

$$M = E\left[-\left.\frac{\partial^2}{\partial\theta\partial\theta'}\Psi\left(Y_i, r_\theta\left(X_i'\theta\right)\right)\right|_{\theta=\alpha}\right]$$

est définie positive.

On obtient :

Théorème 3. *Sous les hypothèses H1–H8 et H9'-H10', $\widehat{\theta}_n$ défini par*

$$\widehat{\theta}_n = \arg\max_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n}\Psi\left(Y_i, \widehat{r}_\theta\left(X_i'\theta\right)\right) \tag{30}$$

converge presque sûrement vers α. De plus, on a

$$\sqrt{n}\left(\widehat{\theta}_n - \alpha\right) \xrightarrow{L} N\left(0,\Sigma^*\right), \tag{31}$$

avec $\Sigma = M^{-1}VM^{-1}$, *où*

$$V = E\left[\frac{\partial}{\partial\theta}\Psi\left(Y_i, r_\theta\left(X_i'\theta\right)\right)\Big|_{\theta=\alpha}\frac{\partial}{\partial\theta'}\Psi\left(Y_i, r_\theta\left(X_i'\theta\right)\right)\Big|_{\theta=\alpha}\right]. \tag{32}$$

Un exemple important d'utilisation de ce dernier résultat, qui permet de retrouver l'estimateur de Klein et Spady (1993) comme cas particulier, est constitué par les modèles CDR dans lesquels la densité conditionnelle de Y_i sachant que $X_i = x$ ne dépend que de $\varphi\left(x'\alpha\right)$, quand la forme de cette densité est connue du statisticien. Cela permet d'utiliser pour Ψ la vraie log-vraisemblance conditionnelle de l'échantillon, même lorsqu'elle ne provient pas d'une loi exponentielle linéaire.

3. Les conditions H5-H8 sont classiques dans ce contexte. On peut remarquer que si l'on utilise un noyau d'ordre m, alors dans l'hypothèse H6 γ peut être choisi dans l'intervalle $\left(\frac{1}{4m}, \frac{1}{7}\right)$.

4. On peut s'affranchir des hypothèses H5-1) et 2) en utilisant une technique de élagage sur les observations. Cela revient finalement à remplacer $\widehat{r}_\theta$ par

$$\widehat{r}_\theta\left(t\right) = \left\{\sum_{j\neq i} Y_j K\left(\frac{t - X_j'\theta}{a_n}\right)\ I_{\{X_j\in S\}}\right\} \Big/ \left\{\sum_{j\neq i} K\left(\frac{X_i'\theta - X_j'\theta}{a_n}\right)\ I_{\{X_j\in S\}}\right\},$$

et $\widehat{\theta}_n$ par

$$\widehat{\theta}_n = \underset{\theta\in\Theta}{\arg\max}\ \frac{1}{n}\sum_{i=1}^{n}\Psi\left(Y_i, \widehat{r}_\theta\left(X_i'\theta\right)\right)\ I_{\{X_j\in S\}},$$

en supposant que le compact S choisi vérifie H5-2).

On verra Bonneu, Delecroix et Hristache (1995) pour voir que dans ce cas, la convergence uniforme p.s. de $\widehat{r}_\theta$ vers r_θ reste assurée. Cette technique a été utilisée par Härdle, Hall et Ichimura (1993) et Ichimura (1993). Nous avons cependant choisi de garder l'hypothèse H5 pour simplifier la rédaction de l'article et permettre une meilleure prehension des concepts.

4 Etude de l'efficacité de $\widehat{\theta}_n$

Dans les modèles définis par (3), Newey (1990) montre que la borne d'efficacité prend la forme $\Sigma_0 = \left[E\left(SS'\right)\right]^{-1}$, avec

$$S = \frac{1}{\Omega\left(X_i\right)}\left[Y_i - \varphi\left(v\left(X_i, \alpha\right)\right)\right]\varphi'\left(v\left(X_i, \alpha\right)\right)\left[v_\alpha - E^\Omega\left(v_\alpha|v\right)\right], \tag{33}$$

où :

- $\Omega\left(X_i\right) = var\left(Y_i|X_i\right)$
- $v_\alpha = \dfrac{\partial v\left(X_i, \theta\right)}{\partial\theta}\Big|_{\theta=\alpha}$
- $E^\Omega\left(v_\alpha|v\right) = \dfrac{E\left(\Omega^{-1}v_\alpha|v\right)}{E\left(\Omega^{-1}|v\right)}$.

Pour $v\left(x, \theta\right) = x'\theta$ on obtient

$$\Sigma_0^{-1} = E\left\{\frac{1}{\Omega\left(X_i\right)}r_\alpha'^2\left(X_i'\alpha\right)\left[X_i - E^\Omega\left(X_i|X_i'\alpha\right)\right]\left[X_i - E^\Omega\left(X_i|X_i'\alpha\right)\right]'\right\}. \tag{34}$$

Dans le cas particulier où la variance conditionnelle ne dépend que de $X_i'\alpha$, cette relation devient :

$$\begin{aligned}\Sigma_0^{-1} &= E\left\{\frac{1}{\Omega(X_i)} r_\alpha'^2(X_i'\alpha)\left[X_i - E(X_i|X_i'\alpha)\right]\left[X_i - E(X_i|X_i'\alpha)\right]'\right\} \\ &= E\left\{\frac{1}{\Omega(X_i)} \left.\frac{\partial r_\theta(X_i'\theta)}{\partial\theta}\right|_{\theta=\alpha} \left.\frac{\partial r_\theta(X_i'\theta)}{\partial\theta'}\right|_{\theta=\alpha}\right\}.\end{aligned} \tag{35}$$

Si l'on compare cette matrice à la matrice de variance-covariance de l'estimateur du pseudo-maximum de vraisemblance (théorème 2), on s'aperçoit qu'elles coïncident si $\Omega(X_i) = [C'(\varphi(X_i'\alpha))]^{-1}$, c'est-à-dire si la vraie variance conditionnelle de Y_i sachant X_i est égale à la variance $\Delta(X_i) = [C'(\varphi(X_i'\alpha))]^{-1}$ associée à la pseudo-vraisemblance utilisée. Mais comme $\Omega(X_i)$ est inconnue, il y a évidemment une perte d'efficacité quand on estime α par cette méthode, perte donnée par le rapport $\Delta(X_i)/\Omega(X_i)$. Plus ce rapport est proche de 1, plus l'estimateur basé sur le pseudo-maximum de vraisemblance est efficace, la borne d'efficacité étant atteinte lorsqu'il est égal à 1.

Dans le cas où la variance conditionnelle est fonction de r_α (voir Carroll, Fan, Gijbels and Wand (1995) pour une autre approche dans le cadre de cette hypothèse), on a :

$$\Omega(X_i) = v_0(r_\alpha(X_i'\alpha)), \qquad (v_0 \text{ fonction connue}),$$

et on peut obtenir un estimateur efficace de α . Cela a été déjà fait par Härdle, Hall et Ichimura (1993), mais par une approche en deux étapes, en utilisant les moindres carrés pondérés. On peut l'obtenir directement en choisissant comme pseudo-vraisemblance f l'unique densité appartenant à la famille exponentielle linéaire (voir (7)) qui vérifie $C'(r) = [v_0(r)]^{-1}$. Alors $\widehat{\theta}_n$ est efficace, car $\Delta(X_i) = \Omega(X_i) = v(r_\alpha(X_i'\alpha))$ entraîne l'égalité des deux matrices M et V données par (25) et (27), ce qui fait que la matrice de variance-covariance asymptotique de $\widehat{\theta}_n$ coïncide avec Σ_0 donnée par (35). On obtient donc :

Proposition 2. *Dans la classe de modèles à direction révélatrice unique tels que*

$$var(Y_i|X_i) = v_0(r_\alpha(X_i'\alpha)) \tag{36}$$

où v est une fonction connue, sous les conditions H1-H10,

$$\widehat{\theta}_n = \arg\max_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^n \left[A(\widehat{r}_\theta(X_i'\theta)) + C(\widehat{r}_\theta(X_i'\theta))Y_i\right] \tag{37}$$

avec $C'(r) = [v_0(r)]^{-1}$, est un estimateur asymptotiquement efficace de α.

Notons que si la vraie densité conditionnelle $f(y,r)$ de Y_i sachant X_i est une exponentielle linéaire, l'estimateur défini par $\Psi(y,r) = \log f(y,r)$ est également efficace dans la classe des modèles CDR.

Finalement, deux cas restent à considérer, celui où la fonction v_0 est inconnue du statisticien, et celui où on ne dispose d'aucune information sur Ω. On peut encore dans chaque cas obtenir un estimateur efficace. C'est l'objet des deux paragraphes qui suivent.

4.1 Cas où $\Omega(X_i)$ est fonction de $X_i'\alpha$.

Pour construire un estimateur efficace de α, il suffit de modifier l'estimateur usuel, en multipliant le critère habituel Ψ par un estimateur $\widehat{a}$ d'une fonction a ($a(x)$ sera par la suite essentiellement $\Delta(x)/\Omega(x)$). Supposons vérifiées les conditions suivantes :

C1) $a \in C(S, \mathbb{R}_+)$, où $C(S, \mathbb{R}_+) = \{f : S \to \mathbb{R}_+,\ f \text{continue}\}$;

C2) $0 \leq m_n \leq n, \quad m_n \underset{n\to\infty}{\longrightarrow} \infty, \quad \dfrac{m_n}{n} \underset{n\to\infty}{\longrightarrow} 0;$

C3) $\widehat{a}$ est une application mesurable de $\mathbb{R}^{(d+1)m_n}$ dans $C(S, \mathbb{R}_+)$, calculée sur les premiers m_n couples (X_i', Y_i). On note

$$\widehat{a}_{m_n} = \widehat{a}(X_1, Y_1, \ldots, X_{m_n}, Y_{m_n});$$

C4) $\sup_{x\in S} |\widehat{a}_{m_n}(x) - a(x)| \xrightarrow[m_n\to\infty]{p.s.} 0;$

C5) $m_n^{1/4} \sup_{x\in S} |\widehat{a}_{m_n}(x) - a(x)| \xrightarrow[m_n\to\infty]{P} 0;$

C6) $\dfrac{1}{\sqrt{n-m_n}} \displaystyle\sum_{i=m_n+1}^{n} \left.\frac{\partial \Psi(Y_i, r_\theta(X_i'\theta))}{\partial\theta}\right|_{\theta=\alpha} [\widehat{a}_{m_n}(X_i) - a(X_i)] \xrightarrow[n\to\infty]{P} 0;$

C7) $E\left[a(X_i) \left.\dfrac{\partial r_\theta(X_i'\theta)}{\partial\theta}\right|_{\theta=\alpha} |X_i'\alpha\right] = 0.$

On obtient alors :

Proposition 3. *Sous les hypothèses H1–H8, et H9'-H10', si les conditions C1–C7 sont vérifiées, alors*

$$\widehat{\theta}_n = \arg\max_{\theta\in\Theta} \sum_{i=m_n+1}^{n} \hat{a}_{m_n}(X_i)\, \Psi(Y_i, \widehat{r}_\theta(X_i'\theta)) \tag{38}$$

converge presque sûrement vers α. De plus,

$$\sqrt{n}\left(\widehat{\theta}_n - \alpha\right) \xrightarrow{L} N(0, \Sigma_a), \qquad \Sigma_a = M_a^{-1} V_a M_a^{-1}, \tag{39}$$

où

$$\begin{aligned} M_a &= E\left[-a(X_i) \frac{\partial^2}{\partial\theta\partial\theta'} \Psi(Y_i, r_\theta(X_i'\theta))\Big|_{\theta=\alpha}\right], \\ V_a &= E\left[a^2(X_i) \frac{\partial}{\partial\theta} \Psi(Y_i, r_\theta(X_i'\theta))\Big|_{\theta=\alpha} \frac{\partial}{\partial\theta'} \Psi(Y_i, r_\theta(X_i'\theta))\Big|_{\theta=\alpha}\right]. \end{aligned} \tag{40}$$

Preuve : Il s'agit d'une adaptation directe de la démonstration des théorèmes 1 et 2. Un résumé de la preuve pour la normalité asymptotique est donné dans l'appendice. ■

Il en résulte que, en utilisant un estimateur $\hat{a}_{m_n}(x)$ vérifiant les conditions C3-C7 et convergeant presque sûrement et uniformément sur S vers $\Delta(x)/\Omega(x)$, on obtient un estimateur efficace de α :

Proposition 4. *Supposons les hypothèses H1-H10 satisfaites. Supposons de plus que* $\sup_{x\in S}\left|\hat{a}_{m_n}(x) - \dfrac{\Delta(x)}{\Omega(x)}\right| \xrightarrow[m_n\to\infty]{p.s.} 0$ *et que* $\hat{a}$ *vérifie les conditions C3-C7. Alors*

$$\widehat{\theta}_n = \arg\max_{\theta\in\Theta} \sum_{i=m_n+1}^{n} \widehat{a}(X_i)\left[A\left(\widehat{r}_\theta(X_i'\theta)\right) + C\left(\widehat{r}_\theta(X_i'\theta)\right)Y_i\right], \tag{41}$$

vérifie :

- $\widehat{\theta}_n \xrightarrow{p.s.} \alpha$,
- $\sqrt{n}\left(\widehat{\theta}_n - \alpha\right) \xrightarrow{L} N(0,\Sigma)$, *où*

$$\Sigma = \left\{E\left[\Omega(X_i)^{-1}\left.\frac{\partial r_\theta(X_i'\theta)}{\partial\theta}\right|_{\theta=\alpha}\left.\frac{\partial r_\theta(X_i'\theta)}{\partial\theta'}\right|_{\theta=\alpha}\right]\right\}^{-1}. \tag{42}$$

Preuve : Directe. ■

Une méthode d'estimation efficace de α est donnée par une approche en deux étapes. On note $n_2 = n - m_n$. En une première étape, on détermine un estimateur convergent $\widetilde{\theta}_n$ donnée par l'équation (24), basé sur les premières m_n observations :

$$\widetilde{\theta}_n = \arg\max_{\theta\in\Theta} \sum_{i=1}^{m_n}\left[A\left(\widehat{r}_\theta^{(1),i}(X_i'\theta)\right) + C\left(\widehat{r}_\theta^{(1),i}(X_i'\theta)\right)Y_i\right], \tag{43}$$

où

$$\widehat{r}_\theta^{(1),i}(t) = \frac{\sum_{\substack{j=1\\ j\neq i}}^{m_n} Y_j K\left(\dfrac{t - X_j'\theta}{a_{m_n}}\right)}{\sum_{\substack{j=1\\ j\neq i}}^{m_n} K\left(\dfrac{t - X_j'\theta}{a_{m_n}}\right)} \tag{44}$$

(a_n représente dans cette formule la suite des fenêtres utilisées et définie par H8, la confusion avec la fonction $a(\cdot)$ paraîssant exclue). L'estimateur obtenu n'est pas efficace. Ensuite, à l'aide de $\widetilde{\theta}_n$ on construit un estimateur $\hat{a}_{m_n}(X_i)$ de $\Delta(X_i)/\Omega(X_i)$, pour $m_n < i \leq n$:

$$\hat{a}_{m_n}(X_i) = \left[C'\left(\widehat{r}_{\widetilde{\theta}_n}^{(1)}\left(X_i'\widetilde{\theta}_n\right)\right)\widehat{\omega}_{\widetilde{\theta}_n}\left(X_i'\widetilde{\theta}_n\right)\right]^{-1}, \tag{45}$$

où

$$\widehat{r}_\theta^{(1)}(t) = \frac{\sum_{j=1}^{m_n} Y_j K\left(\dfrac{t - X_j'\theta}{a_{m_n}}\right)}{\sum_{j=1}^{m_n} K\left(\dfrac{t - X_j'\theta}{a_{m_n}}\right)} \tag{46}$$

et

$$\widehat{\omega}_\theta (t) = \frac{\sum_{j=1}^{m_n} Y_j^2 K\left(\frac{t - X_j'\theta}{a_{m_n}}\right)}{\sum_{j=1}^{m_n} K\left(\frac{t - X_j'\theta}{a_{m_n}}\right)} - \left[\widehat{r}_\theta^{(1)} (t)\right]^2 . \tag{47}$$

Dans la deuxième étape, on définit $\widehat{\theta}_n$ par

$$\widehat{\theta}_n = \arg\max_{\theta\in\Theta} \sum_{i=m_n+1}^{n} \hat{a}_{m_n} (X_i) \left[A\left(\widehat{r}_\theta^{(2),i} (X_i'\theta)\right) + C\left(\widehat{r}_\theta^{(1),i} (X_i'\theta)\right) Y_i\right], \tag{48}$$

où

$$\widehat{r}_\theta^{(2),i} (t) = \frac{\sum_{\substack{j=m_n+1 \\ j\neq i}}^{n} Y_j K\left(\frac{t - X_j'\theta}{a_{n_2}}\right)}{\sum_{\substack{j=m_n+1 \\ j\neq i}}^{n} K\left(\frac{t - X_j'\theta}{a_{n_2}}\right)} \qquad (n_2 = n - m_n) \tag{49}$$

et $\hat{a}_{m_n}$ est défini en (45) .

Le résultat qui suit, dont la démonstration est donnée dans l'appendice, montre que $\hat{a}$ vérifie les conditions demandées dans la proposition précédente.

Lemme 2. *Sous les hypothèses H1–H10 (avec $s \geq 6$), si de plus $v_\theta (t) = E\left(Y_i^2 | X_i'\theta = t\right)$ est lipschitzienne sur $T_\theta (S)$, uniformément par rapport à $\theta \in \Theta$, et $V (x, \theta) = v_\theta (x'\theta)$ est continue sur $S \times \Theta$, alors $\hat{a}_{m_n} (X_i)$ donnée par* (45) *vérifie les conditions C3-C7, avec $a (x) = \Delta (x)\ \ [\Omega (x)]^{-1}$.*

Finalement, il en résulte que $\widehat{\theta}_n$ donné par (48) fournit un exemple d'estimateur efficace de α dans la classe des modèles à direction révélatrice unique vérifiant $var\,(Y_i|X_i) = var\,(Y_i|X_i'\alpha)$. Il constitue aussi un exemple d'estimateur efficace dans la classe des modèles CDR dans laquelle la densité conditionnelle de Y_i sachant X_i est une exponentielle linéaire (cas des modèles linéaires généralisés), quand la forme exacte de cette densité est inconnue. En effet, comme un modèle CDR est aussi un modèle à direction révélatrice unique, on peut déterminer pour un tel modèle deux bornes d'efficacité : la première, quand on le regarde comme modèle CDR, est donnée par

$$\Sigma_{CDR}^{-1} = E\left[\frac{\partial}{\partial\theta} \log f\left(Y_i, r_\theta\left(X_i'\theta\right)\right)\Big|_{\theta=\alpha} \frac{\partial}{\partial\theta'} \log f\left(Y_i, r_\theta\left(X_i'\theta\right)\right)\Big|_{\theta=\alpha}\right], \tag{50}$$

et la deuxième lorsqu'on le considère en tant que modèle à direction révélatrice unique, est donnée par (35). En général, ces deux bornes sont différentes. Une condition suffisante pour qu'elles coïncident est

$$var\left(\partial_r \log f\left(Y_i, r_\alpha\left(X_i'\alpha\right)\right) | X_i'\alpha\right) = \frac{1}{var\,(Y_i|X_i'\alpha)}, \tag{51}$$

condition satisfaite si $f (y, r)$ appartient à la famille exponentielle linéaire.

4.2 Cas général.

On suppose ici de ne plus disposer d'informations supplémentaires sur la variance conditionnelle $\Omega(X_i)$. Il nous faudra à la fois modifier le critère Ψ comme auparavant et l'estimateur $\widehat{r}_\theta$ pour obtenir une estimation efficace.

Pour $\widehat{r}_\theta$, supposons donnée une fonction réelle b et une suite d'estimateurs $\widehat{b} = \widehat{b}_{m_n}$ bâtis sur les m_n premiers éléments de l'échantillon :

C8) $b \in C(S, \mathbb{R}_+)$;

C9) $\widehat{b}$ est une application mesurable de $\mathbb{R}^{(d+1)m_n}$ dans $C(S, \mathbb{R}_+)$. On note

$$\widehat{b}_{m_n} = \widehat{b}(X_1, Y_1, \ldots, X_{m_n}, Y_{m_n});$$

C10) $\sup_{x \in S} \left| \widehat{b}_{m_n}(x) - b(x) \right| \xrightarrow[m_n \to \infty]{p.s.} 0.$

On obtient :

Lemme 3. *Sous les hypothèses H1–H2 et H5-H8, si les conditions C2 et C8–C10 sont vérifiées, alors*

$$\sup_{(x,\theta) \in S \times \Theta} \left| \widehat{r}_\theta^{\widehat{b},i}(x'\theta) - r_\theta^b(x'\theta) \right| \underset{n \to \infty}{\overset{p.s.}{\to}} 0, \tag{52}$$

$$n^{1/4} \sup_{(x,\theta) \in S \times \Theta} \left| \widehat{r}_\theta^{\widehat{b},i}(x'\theta) - r_\theta^b(x'\theta) \right| \underset{n \to \infty}{\overset{P}{\to}} 0, \tag{53}$$

$$n^{1/4} \sup_{(x,\theta) \in S \times \Theta} \left| \frac{\partial \widehat{r}_\theta^{\widehat{b},i}(x'\theta)}{\partial \theta} - \frac{\partial r_\theta^b(x'\theta)}{\partial \theta} \right| \underset{n \to \infty}{\overset{P}{\to}} 0, \tag{54}$$

$$\sup_{(x,\theta) \in S \times \Theta} \left| \frac{\partial^2 \widehat{r}_\theta^{\widehat{b},i}(x'\theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 r_\theta^b(x'\theta)}{\partial \theta \partial \theta'} \right| \underset{n \to \infty}{\overset{P}{\to}} 0, \tag{55}$$

avec

$$\widehat{r}_\theta^{\widehat{b},i}(t) = \frac{\sum_{\substack{j=m_n+1 \\ j \neq i}}^n \widehat{b}_{m_n}(X_j) Y_j K\left(\dfrac{t - X_j'\theta}{a_{n_2}}\right)}{\sum_{\substack{j=m_n+1 \\ j \neq i}}^n \widehat{b}_{m_n}(X_j) K\left(\dfrac{t - X_j'\theta}{a_{n_2}}\right)}, \quad (n_2 = n - m_n) \tag{56}$$

et

$$r_\theta^b(t) = \frac{E[b(X_i) Y_i | X_i'\theta = t]}{E[b(X_i) | X_i'\theta = t]}. \tag{57}$$

Preuve : Directe.

Proposition 5. *Sous les hypothèses H1–H8, et H9'-H10', si les conditions C1–C10 sont vérifiées (avec r_θ remplaçé par r_θ^b) , alors*

$$\widehat{\theta}_n = \arg\max_{\theta \in \Theta} \sum_{i=m_n+1}^n \hat{a}_{m_n}(X_i) \Psi\left(Y_i, \widehat{r}_\theta^{\widehat{b},i}(X_i'\theta)\right) \tag{58}$$

converge presque sûrement vers α. *De plus,*

$$\sqrt{n}\left(\widehat{\theta}_n-\alpha\right)\xrightarrow{L}N\left(0,\Sigma_{ab}\right),\qquad \Sigma_{ab}=M_{ab}^{-1}V_{ab}M_{ab}^{-1},\tag{59}$$

où

$$\begin{aligned} M_{ab}&=E\left[-a\left(X_i\right)\frac{\partial^2}{\partial\theta\partial\theta'}\Psi\left(Y_i,r_\theta^b\left(X_i'\theta\right)\right)\Big|_{\theta=\alpha}\right],\\ V_{ab}&=E\left[a^2\left(X_i\right)\frac{\partial}{\partial\theta}\Psi\left(Y_i,r_\theta^b\left(X_i'\theta\right)\right)\Big|_{\theta=\alpha}\frac{\partial}{\partial\theta'}\Psi\left(Y_i,r_\theta^b\left(X_i'\theta\right)\right)\Big|_{\theta=\alpha}\right].\end{aligned}\tag{60}$$

Preuve : Directe.

Dans le cas particulier de la méthode basée sur le pseudo-maximum de vraisemblance, on obtient :

Corollaire 2. *Sous les hypothèses H1-H10, si les conditions C1–C10 sont satisfaites, alors*

$$\widehat{\theta}_n=\arg\max_{\theta\in\Theta}\sum_{i=m_n+1}^{n}\widehat{a}_{m_n}\left(X_i\right)\left[A\left(\widehat{r}_\theta^{b,i}\left(X_i'\theta\right)\right)+C\left(\widehat{r}_\theta^{b,i}\left(X_i'\theta\right)\right)Y_i\right],\tag{61}$$

vérifie :

- $\widehat{\theta}_n\xrightarrow{p.s.}\alpha$,
- $\sqrt{n}\left(\widehat{\theta}_n-\alpha\right)\xrightarrow{L}N\left(0,\Sigma_{ab}\right)$, $\Sigma_{ab}=M_{ab}^{-1}V_{ab}M_{ab}^{-1}$, *où*

$$M_{ab}=E\left[a\left(X_i\right)\frac{\partial r_\theta^b\left(X_i'\theta\right)}{\partial\theta}\Big|_{\theta=\alpha}\Delta^b\left(X_i\right)^{-1}\frac{\partial r_\theta^b\left(X_i'\theta\right)}{\partial\theta'}\Big|_{\theta=\alpha}\right],$$

$$V_{ab}=E\left[a^2\left(X_i\right)\frac{\partial r_\theta^b\left(X_i'\theta\right)}{\partial\theta}\Big|_{\theta=\alpha}\Delta^b\left(X_i\right)^{-1}\;\Omega\left(X_i\right)\;\Delta^b\left(X_i\right)^{-1}\frac{\partial r_\theta^b\left(X_i'\theta\right)}{\partial\theta'}\Big|_{\theta=\alpha}\right],$$

avec $\Omega\left(X_i\right)=var\left(Y_i\left|X_i\right.\right)$ *et* $\Delta^b\left(X_i\right)=\left[C'\left(r_\alpha^b\left(X_i'\alpha\right)\right)\right]^{-1}$.

Si l'on compare Σ_{ab} avec Σ_0 donnée par (34) et si l'on tient compte du fait que

$$\frac{\partial r_\theta^b\left(X_i'\theta\right)}{\partial\theta}\Big|_{\theta=\alpha}=\dot{r}_\alpha^b\left(X_i'\alpha\right)\left[X_i-\frac{E\left[b\left(X_i\right)X_i|X_i'\alpha\right]}{E\left[b\left(X_i\right)|X_i'\alpha\right]}\right],\tag{62}$$

on s'aperçoit que pour avoir l'égalité des deux matrices il suffit de prendre

$$b\left(X_i\right)=\left[\Omega\left(X_i\right)\right]^{-1}\quad\text{et}\quad a\left(X_i\right)=\left[C'\left(r_\alpha^b\left(X_i'\alpha\right)\right)\Omega\left(X_i\right)\right]^{-1}.\tag{63}$$

En conséquence, on pourrait obtenir un estimateur efficace de α si l'on disposait des estimateurs $\widehat{a}$ de a et $\widehat{b}$ de b qui vérifient les conditions C1–C10.

On peut alors envisager encore une fois une procédure en deux étapes pour estimer α d'une manière efficace. On divise l'échantillon $\left(X_i',Y_i\right)_{i=1}^n$ en deux sous-échantillons. On utilise le premier sous-échantillon, de taille m_n, pour obtenir en même temps un estimateur (non efficace) $\widetilde{\theta}_n$ de α et un estimateur $\widehat{\Omega}(X_i)$ de la variance conditionnelle $\Omega(X_i) = var(Y_i|X_i)$. Cela nous permet de définir des estimateurs $\widehat{a}$ et $\widehat{b}$ de a et b qui ne dépendent pas des observations du deuxième sous-échantillon. Dans un deuxième temps, on utilise le deuxième sous-échantillon pour déterminer un estimateur efficace $\widehat{\theta}_n$ de α, en utilisant les expressions de $\widehat{a}$ et $\widehat{b}$ obtenues dans la première étape.

Notons que si l'on prend comme pseudo-vraisemblance la densité gaussienne (donc $C(r) = r$), il est suffisant d'obtenir à l'issue du premier pas un estimateur de la variance conditionnelle, car dans ce cas a et b donnés par (63) coïncident.

Appendice

Pour la démonstration du théorème 2 nous aurons besoin des quatre lemmes qui suivent. Les lemmes 5-7 sont similaires aux lemmes 5, 7 and 8 de Bonneu, Delecroix et Hristache (1995).

Lemme 4 (Ichimura (1993)). *Soit $J_n(\theta)$ deux fois continûment différentiable. Supposons que les conditions suivantes sont satisfaites :*

(1) $\widehat{\theta}_n$ converge en probabilité vers α et

$$\widehat{\theta}_n = \arg\min_{\theta\in\Theta} J_n(\theta).$$

(2) Il existe un vecteur aléatoire Δ_n qui converge en loi vers un vecteur aléatoire normal d'espérance nulle et de matrice de variance-covariance V, tel que

$$\left|\sqrt{n}\frac{\partial J_n(\alpha)}{\partial\theta} - \Delta_n\right| = o_p(1).$$

(3) Il existe une matrice M définie positive telle que

$$\left|\frac{\partial^2 J_n(\alpha)}{\partial\theta\partial\theta'} - M\right| = o_p(1).$$

(4) Pour tout $\varepsilon > 0$, il existe un voisinage U_ε de α tel que

$$P\left(\sup_{\theta\in U_\varepsilon}\left|\frac{\partial^2 J_n(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2 J_n(\alpha)}{\partial\theta\partial\theta'}\right| > \varepsilon\right) \underset{n\to\infty}{\longrightarrow} 0.$$

Alors $\sqrt{n}\left(\widehat{\theta}_n - \alpha\right)$ converge en loi vers un vecteur normal d'espérance 0 et matrice de variance-covariance $M^{-1}VM^{-1}$.

Lemme 5. *Sous les hypothèses H1–H6 et H9',*

$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial^2}{\partial\theta_l\partial\theta_k}\Psi(Y_i, \widehat{r}_\theta(X_i'\theta)) - \frac{\partial^2}{\partial\theta_l\partial\theta_k}\Psi(Y_i, r_\theta(X_i'\theta))\right] \xrightarrow[n\to\infty]{P} 0,$$

uniformément par rapport à $\theta \in \Theta$.

Lemme 6. *Sous les hypothèses H1–H10,*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{\partial}{\partial\theta}\Psi\left(Y_i,\widehat{r}_\theta\left(X_i'\theta\right)\right)\bigg|_{\theta=\alpha}-\frac{\partial}{\partial\theta}\Psi\left(Y_i,r_\theta\left(X_i'\theta\right)\right)\bigg|_{\theta=\alpha}\right]\xrightarrow[n\to\infty]{P}0.$$

Lemme 7. *Sous les hypothèses H1–H6 et H9', pour tout $\varepsilon>0$, il existe un voisinage U_ε de α tel que*

$$P\left(\sup_{\theta\in U_\varepsilon}\left|\frac{\partial^2}{\partial\theta\partial\theta'}\Psi\left(Y_i,r_\theta\left(X_i'\theta\right)\right)-\frac{\partial^2}{\partial\theta\partial\theta'}\Psi\left(Y_i,r_\theta\left(X_i'\theta\right)\right)\bigg|_{\theta=\alpha}\right|>\varepsilon\right)\xrightarrow[n\to\infty]{}0.$$

Preuve du théorème 2 : Il suffit de montrer que

$$J_n(\theta)=-\frac{1}{n}\sum_{i=1}^{n}\Psi\left(Y_i,\widehat{r}_\theta\left(X_i'\theta\right)\right)$$

vérifie les conditions (1)–(4) du lemme 4.

(1) $\widehat{\theta}_n$ converge presque sûrement vers α, conformément au théorème 1.

(2) Soit $\Delta_n=-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\Psi\left(Y_i,r_\theta\left(X_i'\theta\right)\right)\Big|_{\theta=\alpha}$. Alors :

i) $\Delta_n\xrightarrow{L}N\left(0,V\right)$, car

$$\alpha=\ \arg\max_{\theta\in\Theta}E\left[\Psi\left(Y_i,r_\theta\left(X_i'\theta\right)\right)\right]\quad\Longrightarrow\quad E\left[\frac{\partial}{\partial\theta}\Psi\left(Y_i,r_\theta\left(X_i'\theta\right)\right)\bigg|_{\theta=\alpha}\right]=0.$$

ii) $\left|\sqrt{n}\frac{\partial J_n(\alpha)}{\partial\theta}-\Delta_n\right|$

$$=\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{\partial}{\partial\theta}\ \Psi\left(Y_i,\widehat{r}_\theta\left(X_i'\theta\right)\right)\big|_{\theta=\alpha}-\frac{\partial}{\partial\theta}\ \Psi\left(Y_i,r_\theta\left(X_i'\theta\right)\right)\big|_{\theta=\alpha}\right]\right|$$

converge en probabilité vers 0 quand $n\to\infty$ (lemme 6).

(3) $$\frac{\partial^2 J_n(\alpha)}{\partial\theta\partial\theta'}=-\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta\partial\theta'}\Psi\left(Y_i,\widehat{r}_\theta\left(X_i'\theta\right)\right)\bigg|_{\theta=\alpha}$$

$$\xrightarrow[n\to\infty]{P}E\left[-\frac{\partial^2}{\partial\theta\partial\theta'}\Psi\left(Y_i,r_\theta\left(X_i'\theta\right)\right)\bigg|_{\theta=\alpha}\right]=M\quad\text{(lemme 5)},$$

et M est une matrice définie positive.

(4) Pour tout $\varepsilon>0$, il existe un voisinage U_ε de α tel que

$$P\left(\sup_{\theta\in U_\varepsilon}\left|\frac{\partial^2 J_n(\theta)}{\partial\theta\partial\theta'}-\frac{\partial^2 J_n(\alpha)}{\partial\theta\partial\theta'}\right|>\varepsilon\right)$$

$$=P\left(\sup_{\theta\in U_\varepsilon}\left|\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial^2}{\partial\theta\partial\theta'}\Psi\left(Y_i,\widehat{r}_\theta\left(X_i'\theta\right)\right)-\frac{\partial^2}{\partial\theta\partial\theta'}\Psi\left(Y_i,\widehat{r}_\theta\left(X_i'\theta\right)\right)\Big|_{\theta=\alpha}\right]\right|>\varepsilon\right)$$

$$\leq 2\,P\left(\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial^2}{\partial\theta\partial\theta'}\Psi\left(Y_i,\widehat{r}_\theta\left(X_i'\theta\right)\right)-\frac{\partial^2}{\partial\theta\partial\theta'}\Psi\left(Y_i,r_\theta\left(X_i'\theta\right)\right)\right]\right|>\frac{\varepsilon}{3}\right)$$

$$P\left(\sup_{\theta\in U_\varepsilon}\left|\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial^2}{\partial\theta\partial\theta'}\Psi\left(Y_i,r_\theta\left(X_i'\theta\right)\right)-\frac{\partial^2}{\partial\theta\partial\theta'}\Psi\left(Y_i,r_\theta\left(X_i'\theta\right)\right)\Big|_{\theta=\alpha}\right]\right|>\frac{\varepsilon}{3}\right)\longrightarrow 0,$$

conformément aux lemmes 5 et 7.

■

Preuve de la proposition 3. Afin d'obtenir la normalité asymptotique, on doit vérifier les quatre conditions du lemme 4 pour

$$J_n(\theta)=-\frac{1}{n_2}\sum_{i=m_n+1}^{n}\widehat{a}_{m_n}(X_i)\;\;\Psi\left(Y_i,\widehat{r}_\theta\left(X_i'\theta\right)\right),\tag{64}$$

où $n_2=n-m_n$. On se résume à vérifier seulement la deuxième condition, puisque les preuves pour les trois autres sont tout à fait similaires à celles du théorème 1 (pour la première condition) et des lemmes 5 et 7 (pour les deux dernières conditions).

Soit

$$\begin{aligned}\Delta_n &= -\frac{1}{\sqrt{n_2}}\sum_{i=m_n+1}^{n}a(X_i)\;\;\partial_\theta\Psi\left(Y_i,r_\alpha\left(X_i'\alpha\right)\right)\\ &= -\frac{1}{\sqrt{n_2}}\sum_{i=m_n+1}^{n}a(X_i)\;\;\partial_r\Psi\left(Y_i,r_\alpha\left(X_i'\alpha\right)\right)\;\;\partial_\theta r_\alpha\left(X_i'\alpha\right).\end{aligned}\tag{65}$$

En utilisant une décomposition de la forme

$$\begin{aligned}\hat a\hat b\hat c-abc &= bc\left(\hat a-a\right)+ac\left(\hat b-b\right)+ab\left(\hat c-c\right)\\ &\quad+a\left(\hat b-b\right)\left(\hat c-c\right)+b\left(\hat a-a\right)\left(\hat c-c\right)+c\left(\hat a-a\right)\left(\hat b-b\right)\\ &\quad+\left(\hat a-a\right)\left(\hat b-b\right)\left(\hat c-c\right),\end{aligned}$$

il est alors suffisant de montrer la convergence vers zéro en probabilité des termes suivants :

$$W_1=\frac{1}{\sqrt{n_2}}\sum_{i=m_n+1}^{n}\partial_r\Psi\left(Y_i,r_\alpha\left(X_i'\alpha\right)\right)\;\;\partial_\theta r_\alpha\left(X_i'\alpha\right)\;\;\left[\widehat{a}_{m_n}(X_i)-a(X_i)\right],$$

$$W_2=\frac{1}{\sqrt{n_2}}\sum_{i=m_n+1}^{n}a(X_i)\;\;\partial_\theta r_\alpha\left(X_i'\alpha\right)\;\;\left[\partial_r\Psi\left(Y_i,\widehat{r}_\alpha\left(X_i'\alpha\right)\right)-\partial_r\Psi\left(Y_i,r_\alpha\left(X_i'\alpha\right)\right)\right],$$

$$W_3 = \frac{1}{\sqrt{n_2}} \sum_{i=m_n+1}^{n} a(X_i) \ \partial_r \Psi(Y_i, r_\alpha(X_i'\alpha)) \ [\partial_\theta \widehat{r}_\alpha(X_i'\alpha) - \partial_\theta r_\alpha(X_i'\alpha)],$$
$$W_4 = \frac{1}{\sqrt{n_2}} \sum_{i=m_n+1}^{n} a(X_i) \ [\partial_r \Psi(Y_i, \widehat{r}_\alpha(X_i'\alpha)) - \partial_r \Psi(Y_i, r_\alpha(X_i'\alpha))] \times \ [\partial_\theta \widehat{r}_\alpha(X_i'\alpha) - \partial_\theta r_\alpha(X_i'\alpha)],$$
$$W_5 = \frac{1}{\sqrt{n_2}} \sum_{i=m_n+1}^{n} \partial_r \Psi(Y_i, r_\alpha(X_i'\alpha)) \ [\widehat{a}_{m_n}(X_i) - a(X_i)] \ [\partial_\theta \widehat{r}_\alpha(X_i'\alpha) - \partial_\theta r_\alpha(X_i'\alpha)],$$
$$W_6 = \frac{1}{\sqrt{n_2}} \sum_{i=m_n+1}^{n} \partial_\theta r_\alpha(X_i'\alpha) [\widehat{a}_{m_n}(X_i) - a(X_i)] [\partial_r \Psi(Y_i, \widehat{r}_\alpha(X_i'\alpha)) - \partial_r \Psi(Y_i, r_\alpha(X_i'\alpha))],$$
$$W_7 = \frac{1}{\sqrt{n_2}} \sum_{i=m_n+1}^{n} [\widehat{a}_{m_n}(X_i) - a(X_i)] \ [\partial_r \Psi(Y_i, \widehat{r}_\alpha(X_i'\alpha)) - \partial_r \Psi(Y_i, r_\alpha(X_i'\alpha))] \times \times \ [\partial_\theta \widehat{r}_\alpha(X_i'\alpha) - \partial_\theta r_\alpha(X_i'\alpha)].$$

Alors :

- la condition C6 équivaut à $W_1 = o_p(1)$;
- la condition C7 permet d'utiliser le lemme 6 de Bonneu, Delecroix et Hristache (1995) pour montrer que $W_2 = o_p(1)$;
- l'hypothèse H8, le lemme 6 de Bonneu, Delecroix et Hristache (1995) et le lemme 1 entraînent $W_3 = o_p(1)$;
- le lemme 1, l'hypothèse H9' et les conditions C1 et C4 donnent $W_4 = o_p(1)$ et $W_7 = o_p(1)$;
- la condition C5 et le lemme 1 entraînent $W_5 = o_p(1)$, $W_6 = o_p(1)$.

■

Preuve du lemme 2. En utilisant le lemme 1 et l'hypothèse H3-4), on obtient :

$$\sup_{x\in S} \left| \hat{r}_{\tilde{\theta}}^{(1)}\left(x'\tilde{\theta}\right) - r_\alpha(x'\alpha) \right| \leq \sup_{x\in S} \left| \hat{r}_{\tilde{\theta}}^{(1)}\left(x'\tilde{\theta}\right) - r_{\tilde{\theta}}\left(x'\tilde{\theta}\right) \right| + \sup_{x\in S} \left| r_{\tilde{\theta}}\left(x'\tilde{\theta}\right) - r_\alpha(x'\alpha) \right|$$
$$\leq \sup_{(x,\theta)\in S\times\Theta} \left| \hat{r}_{\theta}^{(1)}(x'\theta) - r_\theta(x'\theta) \right| + \sup_{x\in S} \left| R\left(x,\tilde{\theta}\right) - R(x,\alpha) \right| \xrightarrow[m_n\to\infty]{p.s.} 0,$$

ce qui, en tenant compte du fait que C' est une fonction continue, donne

$$\sup_{x\in S} \left| C'\left(\hat{r}_{\tilde{\theta}}^{(1)}\left(x'\tilde{\theta}\right)\right) - C'(r_\alpha(x'\alpha)) \right| \xrightarrow[m_n\to\infty]{p.s.} 0. \tag{66}$$

De la même façon que pour $\hat{r}_\theta(x'\theta)$ on peut montrer (voir lemme 1) que

$$\sup_{(x,\theta)\in S\times\Theta} \left| \hat{v}_\theta(x'\theta) - v_\theta(x'\theta) \right| \xrightarrow[m_n\to\infty]{p.s.} 0, \tag{67}$$

où

$$\hat{v}_\theta(t) = \frac{\sum_{j=1}^{m_n} Y_j^2 K\left(\frac{t - X_j'\theta}{a_{m_n}}\right)}{\sum_{j=1}^{m_n} K\left(\frac{t - X_j'\theta}{a_{m_n}}\right)}. \tag{68}$$

Par un raisonnement similaire à celui qui précède on démontre que

$$\sup_{x\in S} \left| \hat{v}_{\tilde{\theta}}\left(x'\tilde{\theta}\right) - v_\alpha(x'\alpha) \right| \xrightarrow[m_n\to\infty]{p.s.} 0, \tag{69}$$

ce qui entraîne

$$\sup_{x\in S}\left|\widehat{\omega}_{\tilde{\theta}}\left(x'\tilde{\theta}\right)-\Omega\left(x\right)\right|\xrightarrow[m_n\to\infty]{p.s.}0. \tag{70}$$

On obtient alors C4 comme une conséquence immédiate de (66), (70) et de l'hypothèse H4-2).

Ensuite, on peut écrire :

$$\begin{aligned}\widehat{a}\left(x\right)-a\left(x\right) &= \left[C'\left(\hat{r}_{\tilde{\theta}}^{(1)}\left(x'\tilde{\theta}\right)\right)\right]^{-1}\left\{\hat{v}_{\tilde{\theta}}\left(x'\tilde{\theta}\right)-\left[\hat{r}_{\tilde{\theta}}^{(1)}\left(x'\tilde{\theta}\right)\right]^{2}\right\}^{-1}\\ &\quad -\left[C'\left(r_{\alpha}\left(x'\alpha\right)\right)\right]^{-1}\left\{v_{\alpha}\left(x'\alpha\right)-\left[r_{\alpha}\left(x'\alpha\right)\right]^{2}\right\}^{-1}.\end{aligned} \tag{71}$$

Pour vérifier la condition C5, il est donc suffisant de montrer que :

$$\sup_{(x,\theta)\in S\times\Theta} m_n^{1/4}\left|C'\left(\hat{r}_{\tilde{\theta}}^{(1)}\left(x'\tilde{\theta}\right)\right)-C'\left(r_{\alpha}\left(x'\alpha\right)\right)\right|\xrightarrow[m_n\to\infty]{P}0, \tag{72}$$

$$\sup_{(x,\theta)\in S\times\Theta} m_n^{1/4}\left|\hat{v}_{\tilde{\theta}}\left(x'\tilde{\theta}\right)-v_{\alpha}\left(x'\alpha\right)\right|\xrightarrow[m_n\to\infty]{P}0, \tag{73}$$

et

$$\sup_{(x,\theta)\in S\times\Theta} m_n^{1/4}\left|\left[\hat{r}_{\tilde{\theta}}^{(1)}\left(x'\tilde{\theta}\right)\right]^{2}-\left[r_{\alpha}\left(x'\alpha\right)\right]^{2}\right|\xrightarrow[m_n\to\infty]{P}0. \tag{74}$$

Vérification de (72) *:*

On peut écrire

$$\begin{aligned}&\sup_{x\in S}\left|\hat{r}_{\tilde{\theta}}^{(1)}\left(x'\tilde{\theta}\right)-r_{\alpha}\left(x'\alpha\right)\right|\\ &\qquad\le \sup_{(x,\theta)\in S\times\Theta}\left|\hat{r}_{\theta}^{(1)}\left(x'\theta\right)-r_{\theta}\left(x'\theta\right)\right|+\sup_{x\in S}\left|R\left(x,\tilde{\theta}\right)-R\left(x,\alpha\right)\right|\\ &\qquad\le \sup_{(x,\theta)\in S\times\Theta}\left|\hat{r}_{\theta}^{(1)}\left(x'\theta\right)-r_{\theta}\left(x'\theta\right)\right|+k\,\left|\tilde{\theta}-\alpha\right|,\end{aligned}$$

où k est une constante positive. Alors (72) est vérifiée si l'on prouve que :

$$\sup_{(x,\theta)\in S\times\Theta} m_n^{1/4}\left|\hat{r}_{\theta}^{(1)}\left(x'\theta\right)-r_{\theta}\left(x'\theta\right)\right|\xrightarrow[m_n\to\infty]{P}0, \tag{75}$$

et

$$\sup_{(x,\theta)\in S\times\Theta} m_n^{1/4}\left|\tilde{\theta}-\alpha\right|\xrightarrow[m_n\to\infty]{P}0. \tag{76}$$

La convergence en (75) s'obtient du lemme 1, tandis que (76) est une conséquence immédiate du théorème 2.

La démarche pour obtenir (73) et (74) est analogue et, par conséquent, omise.

Ensuite, on a

$$\begin{aligned}\widehat{a}-a &= \frac{1}{C'\left(\widehat{r}\right)\left(\widehat{v}-\widehat{r}^{2}\right)}-\frac{1}{C'\left(r\right)\left(v-r^{2}\right)}\\ &= \left[\frac{2r}{C'\left(r\right)\left(v-r^{2}\right)^{2}}-\frac{C''\left(r\right)}{C'^{2}\left(r\right)\left(v-r^{2}\right)^{2}}\right]\left(\widehat{r}-r\right)-\frac{1}{C'\left(r\right)\left(v-r^{2}\right)^{2}}\left(\widehat{v}-v\right)\\ &\quad+\ \text{termes d'ordre deux en }\left(\widehat{r}-r\right)\text{ et }\left(\widehat{v}-v\right).\end{aligned}$$

Comme pour (72) et (73), on peut démontrer que le supremum sur $S \times \Theta$ des termes d'ordre deux multipliés par $\sqrt{n_2} = \sqrt{n - m_n}$ tendent vers zéro en probabilité. Afin d'obtenir la condition C6, il reste donc à montrer que :

$$\frac{1}{\sqrt{n_2}} \sum_{i=m_n+1}^{n} q_1 \left(Y_i, X_i'\alpha\right) \left[\hat{r}_{\tilde{\theta}}^{(1)} \left(X_i'\tilde{\theta}\right) - r_\alpha \left(X_i'\alpha\right)\right] \partial_\theta r_\alpha \left(X_i'\alpha\right) \xrightarrow[n\to\infty]{P} 0, \tag{77}$$

et

$$\frac{1}{\sqrt{n_2}} \sum_{i=m_n+1}^{n} q_2 \left(Y_i, X_i'\alpha\right) \left[\hat{v}_{\tilde{\theta}} \left(X_i'\tilde{\theta}\right) - v_\alpha \left(X_i'\alpha\right)\right] \partial_\theta r_\alpha \left(X_i'\alpha\right) \xrightarrow[n\to\infty]{P} 0, \tag{78}$$

où

$$q_1 (y, t) = \left\{ \frac{2r_\alpha (t)}{C' (r_\alpha (t)) \left[v (t) - r_\alpha^2 (t)\right]^2} - \frac{C'' (r_\alpha (t))}{C'^2 (r_\alpha (t)) \left[v (t) - r_\alpha^2 (t)\right]^2} \right\} \partial_r \Psi (y, r_\alpha (t))$$

et

$$q_2 (y, t) = -\frac{1}{C' (r_\alpha (t)) \left[v (t) - r_\alpha^2 (t)\right]^2} \partial_r \Psi (y, r_\alpha (t)) .$$

En tenant compte du fait que, pour $k = 1, 2$,

$$E \left[q_k \left(Y_i, X_i'\alpha\right) \partial_\theta r_\alpha \left(X_i'\alpha\right) | X_i'\alpha\right] = E \left\{ \underbrace{E \left[q_k \left(Y_i, X_i'\alpha\right) | X_i\right]}_{=0} \partial_\theta r_\alpha \left(X_i'\alpha\right) | X_i'\alpha \right\} = 0,$$

car l'hypothèse H4 entraîne $E \left[\partial_r \Psi \left(Y_i, r_\alpha \left(X_i'\alpha\right)\right) | X_i\right] = 0$, une démonstration basée sur le corollaire 1 de Bonneu, Delecroix et Hristache (1995) permet d'obtenir (77) et (78) .

Remerciements. Le papier a bénéficié, lors de son élaboration, des discussions enrichissantes que nous ont accordées O. Bunke, X. Milhaud et A. Monfort. Nous tenons à les en remercier. Evidemment, toutes les erreurs nous restent imputables.

Références

[1] Bonneu M., Delecroix M., Hristache M. (1995) : Semiparametric Estimation of Generalized Linear Models and Related Models, preprint, CREST, Paris.

[2] Bonneu M., Delecroix M., Malin E. (1993) : Semiparametric versus nonparametric estimation in single index regression model : a computational approach, *Computational Statistics*, **8**, 207-222.

[3] Carroll R.J., Fan J., Gijbels I., Wand M.P. (1995) : Generalized Partially Linear Single-Index Models, discussion paper n°9506, Institut de Statistique, Université Catholique de Louvain.

[4] Chamberlain G. (1986) : Asymptotic efficiency in semi-parametric models with censoring, *Journal of Econometrics*, **32**, 189-218.

[5] Cosslett S.R. (1987) : Efficiency bounds for distribution-free estimators of the binary choice and censored regression models, *Econometrica*, **55**, 559-585.

[6] Gourieroux C. (1989) : *Économétrie des variables qualitatives*, Economica, Paris.

[7] Gourieroux C., Monfor, A., Trognon A. (1984) : Pseudo maximum likelihood methods : theory, *Econometrica*, **52**, 681-700.

[8] Härdle W., Hall P., Ichimura H. (1993) : Optimal smoothing in single-index models, *Annals of Statistics*, **21**, 157-178.

[9] Härdle W., Stoker T.M. (1989) : Investigating smooth multiple regression by the method of average derivatives, *Journal of the American Statistical Association*, **84**, 986-995.

[10] Hall P. (1989) : On Projection Pursuit Regression, *Annals of Statistics*, **17**, 573-588.

[11] Ichimura H. (1993) : Semiparametric least squares (SLS) and weighted SLS estimation of single-index models, *Journal of Econometrics*, **58**, 71-120.

[12] Klein R.L., Spady R.H. (1993) : An efficient semiparametric estimator for binary response models, *Econometrica*, **61**, 387-421.

[13] Mc Cullagh P., Nelder J. A. (1989) : *Generalized Linear Models*, Chapman and Hall, London.

[14] Newey W.K. (1990) : Efficient estimation of semiparametric models via moment restrictions, Bellcore Economics Discussion Paper #65.

[15] Newey W.K. (1994) : The asymptotic variance of semiparametric estimators", *Econometrica*, **62**, 1349-1382.

[16] Newey W.K., Stoker T.M. (1993) : "Efficiency of weighted average derivative estimators and index models, *Econometrica*, **61**, 1199-1223.

[17] Powell J.L., Stock J.H., Stoker T.M. (1989) : Semiparametric estimation of index coefficients, *Econometrica*, **57**, 1403-1430.

[18] Sherman R.P. (1994) : U-processes in the analysis of a generalized semiparametric regression estimator, *Econometric Theory*, **10**, 372-395.

[19] Stoker T.M. (1986) : Consistent estimation of scaled coefficients, *Econometrica*, **54**, 1461-1481.

ENSAI et CREST,
Rue Blaise Pascal,
Campus de Ker-Lann,
35170 BRUZ, FRANCE