

REVUE DE STATISTIQUE APPLIQUÉE

M. BONNEU

M. GBA

Comparaison d'estimateurs semi-paramétriques en régression : étude simulée

Revue de statistique appliquée, tome 47, n° 4 (1999), p. 25-32

http://www.numdam.org/item?id=RSA_1999__47_4_25_0

© Société française de statistique, 1999, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/legal.php>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

COMPARAISON D'ESTIMATEURS SEMI-PARAMÉTRIQUES EN RÉGRESSION : ÉTUDE SIMULÉE

M. Bonneu¹ et M.Gba¹

IRMAR, UMR n° 6625

Université de Rennes 1, Campus de Beaulieu, 35042 Rennes

RÉSUMÉ

Soit $(Y, X) \in \mathbb{R} \times \mathbb{R}^p$ un couple de variables aléatoires. Nous considérons le modèle semi-paramétrique de Y sur X , où l'espérance et la variance conditionnelle de Y à X , dépendent de X par l'intermédiaire de $\langle X, \theta \rangle$. Nous présentons deux estimateurs du paramètre θ . Le premier est un estimateur itératif à un pas utilisant le quasi-score. Le second est un estimateur de type minimax défini à partir d'un estimateur de l'approximation du log du rapport de vraisemblance proposé par B. Li (1996). Des simulations sur deux modèles permettent de comparer les propriétés des estimateurs décrits et de voir leur intérêt.

Mots-clés : *estimateur itératif, estimateur minimax, régression semi-paramétrique, quasi-vraisemblance, simulations de Monte Carlo.*

ABSTRACT

Let the stochastic bivariate $(Y, X) \in \mathbb{R} \times \mathbb{R}^p$. We consider the semiparametric regression model of Y over X , where the conditional average and variance of Y given X link to X through $\langle X, \theta \rangle$. We describe two estimators of the parameter θ . The first is a one step iterative estimator which uses the quasi-score. The second is a minimax point of an estimator of the approximation of the log likelihood ratio proposed by B. Li (1996). Some simulations performed on two models lead to compare the properties of the proposed estimators and derive their interest.

Keywords : *iterative estimator, semiparametric regression, minimax estimator, quasi-likelihood, Monte Carlo simulations.*

¹ adresse actuelle : Équipe GRIMM, Université de Toulouse 2, 5 Allées A. Machado, 31058 Toulouse cedex.

1. Introduction

Soit une suite $(Y_i, X_i)_{i=1, \dots, n} \in \mathbb{R} \times \mathbb{R}^p$ de données statistiques, indépendantes, identiquement distribuées et générées par le couple (Y, X) de loi \mathbb{P}_0 . En régression, le paramètre important dont on suppose l'existence est la fonction de régression de Y en X , soit $g_0(x) = E_0(Y|_{X=x})$. La méthodologie très générale non-paramétrique, qui considère seulement des hypothèses de régularité sur la fonction g_0 , n'est pas sans inconvénient. Le premier inconvénient est le «le fléau de la dimension». Quand le nombre des régresseurs p est élevé, la vitesse de convergence de l'estimateur non paramétrique de g_0 , supposé r fois différentiable, qui est un $O\left(n^{-\frac{r}{2r+p}}\right)$, est moins rapide. Le deuxième inconvénient est lié au manque de moyens pour quantifier l'effet de chaque régresseur. Pour pallier à ces inconvénients une approche alternative est naturellement fournie par la modélisation semi-paramétrique qui suppose

$$E_0(Y|X) = E_0(Y|_{\langle X, \theta_0 \rangle}). \quad (1)$$

Les modèles définis par (1), connus sous le nom des modèles à indice simple et très utilisés en économétrie et biologie ont fait l'objet de plusieurs travaux ces dernières années. Dans cet article nous présentons les récents estimateurs étudiés par Bonneau, Gba (1998) et Gba (1998) dans les modèles définis par (1) où la variance conditionnelle de Y à X n'est pas une fonction de la moyenne.

2. Le modèle et les estimateurs

Les hypothèses que l'on formule sur \mathbb{P}_0 résolvent les problèmes précités et d'autre part précisent l'existence de fonctions r_0 et w_0 liant les deux premiers moments conditionnels de Y à X au prédicteur linéaire $\langle X, \theta_0 \rangle$, que l'on notera dans la suite $X \cdot \theta_0$. Le modèle que nous considérons est donc défini par les deux hypothèses (2) et (3) suivantes

$$E_0(Y|X) = r_0(X \cdot \theta_0), \quad (2)$$

$$\text{Var}_0(Y|X) = w_0(X \cdot \theta_0), \quad (3)$$

où les fonctions r_0 et w_0 supposées inconnues, appartiennent à l'ensemble des fonctions réelles deux fois continûment dérivables. Le paramètre θ_0 est supposé appartenir à l'ensemble Θ des vecteurs de \mathbb{R}^p dont la première composante est égale à 1, ce qui assure l'identifiabilité du modèle.

Pour tout θ appartenant à Θ , on considère les estimateurs à noyau de la régression de Y_i en $X_i \cdot \theta$ et de Y_i^2 en $X_i \cdot \theta$, que l'on notera respectivement $\hat{r}_i(\theta)$ et $\hat{w}_i(\theta)$,

$$\hat{r}_i(\theta) = \frac{\sum_{j=1, j \neq i}^n Y_j \cdot K_{a_n}(X_i \cdot \theta - X_j \cdot \theta)}{\sum_{j=1, j \neq i}^n K_{a_n}(X_i \cdot \theta - X_j \cdot \theta)}, \quad (4)$$

$$\hat{w}_i(\theta) = \frac{\sum_{j=1, j \neq i}^n Y_j^2 \cdot K_{a_n}(X_i \cdot \theta - X_j \cdot \theta)}{\sum_{j=1, j \neq 1}^n K_{a_n}(X_i \cdot \theta - X_j \cdot \theta)} - \hat{r}_i(\theta)^2, \quad (5)$$

qui sont les estimateurs de Nadaraya-Watson, pour le noyau gaussien et pour une fenêtre a_n , définis par $K_{a_n}(u) = \frac{\sqrt{1/2\pi}}{a_n} \exp(-u^2/2a_n^2)$. On considère alors l'estimateur $\tilde{\theta}$ semi-paramétrique des moindres carrés ordinaires

$$\tilde{\theta} = \arg \min_{\theta \in \Theta} \sum_{i=1}^n [Y_i - \hat{r}_i(\theta)]^2, \quad (6)$$

dont H. Ichimura (1993) a étudié les propriétés de convergence en probabilité et en loi. Cet estimateur convergent est préférable, pour un grand échantillon, à tout estimateur paramétrique où la fonction r_0 serait incorrectement choisi. Cependant cet estimateur ne tenant pas compte de la variance conditionnelle de Y à X , la variance de loi asymptotique de $n^{1/2}(\tilde{\theta} - \theta_0)$ n'est pas égale à l'inverse de la borne d'efficacité semi-paramétrique. En effet pour tout estimateur régulier³ du paramètre θ_0 , solution d'une fonction estimante linéaire en $Y_i - E_0(Y_i|X_i\theta)^4$, dans un modèle vérifiant les hypothèses (2) et (3), la borne d'efficacité semi-paramétrique est égale à l'inverse de la matrice W ,

$$W(\theta_0) = E_0 \left[\frac{1}{V_0(Y|X)} \cdot r^*(X, \theta_0) \cdot {}^t r^*(X, \theta_0) \right], \quad (7)$$

où $r^*(X, \theta) = \frac{\partial E_0(Y|X \cdot \theta)}{\partial \theta}$. Les justificatifs de cette borne et les détails des calculs sont explicités dans les travaux de W. Newey (1990), W. Wefelmeyer (1996).

Notre premier estimateur, noté $\hat{\theta}_1$, est basé sur la méthode de Raphson Newton à un pas. Il s'exprime en fonction de $\tilde{\theta}$ et du quasi-score estimé \hat{S} , par

$$\hat{\theta}_1 = \tilde{\theta} + [\hat{M}(\tilde{\theta})]^{-1} \hat{S}(\tilde{\theta}), \quad (8)$$

³ Un estimateur $\tilde{\theta}_n$ est dit régulier si pour toute suite, $\theta_n = \theta_0 + \frac{h}{\sqrt{n}}$, $h \in \mathbb{R}^p$, la distribution de $\sqrt{n}(\tilde{\theta}_n - \theta_n)$ converge faiblement vers une loi Gaussienne $L(h, W^{-1}(\theta_0))$.

⁴ c'est-à-dire une fonction de la forme $\sum_{i=1}^n \alpha_i(Y_i - E_0(Y_i|X_i\theta))$, où α_i est une variable aléatoire mesurable en X_i

avec

$$\hat{S}(\theta) = \sum_{i=1}^n \left(\frac{Y_i - \hat{r}_i(\theta)}{\hat{w}_i(\theta)} \right) \frac{\partial \hat{r}_i(\theta)}{\partial \theta} \text{ et } \hat{M}(\theta) = \sum_{i=1}^n (1/\hat{w}_i(\theta)) \cdot \left(\frac{\partial \hat{r}_i(\theta)}{\partial \theta} \right) \cdot t \left(\frac{\partial \hat{r}_i(\theta)}{\partial \theta} \right).$$

La quasi-vraisemblance, au sens de Mac Cullagh (1983) n'étant plus définie dans le cas de l'hypothèse (3), il n'est pas opportun de considérer l'estimateur de quasi-score estimé, déterminé par $\hat{S}(\theta) = 0$, car il converge vers une valeur qui n'est pas forcément la valeur θ_0 qui vérifie (2) et (3). Ceci explique la nécessité de la recherche d'une nouvelle méthodologie d'estimation autre que la maximisation de la quasi-vraisemblance.

Une autre alternative au problème d'existence de la quasi-vraisemblance, est d'utiliser l'approximation du log du rapport de vraisemblance, proposée par D. Li (1996). Dans notre cas, cette approximation dépend des valeurs des fonctions r_0 et w_0 au point $X_i \cdot \theta$. On considère son estimation, notée \hat{R} qui sera défini en (10), obtenue en remplaçant r_0 et w_0 par leurs estimateurs donnés en (4) et (5). Le second estimateur, noté $\hat{\theta}_2$ est dont l'estimateur minimax de \hat{R} ,

$$\sup_{\eta \in \Theta} \hat{R}(\hat{\theta}_2, \eta) = \inf_{\theta \in \Theta} \sup_{\eta \in \Theta} \hat{R}(\theta, \eta) \quad (9)$$

avec

$$\hat{R}(\theta, \eta) = (1/2) \sum_{i=1}^n \left[\frac{Y_i - \hat{r}_i(\theta)}{\hat{w}_i(\theta)} + \frac{Y_i - \hat{r}_i(\eta)}{\hat{w}_i(\eta)} \right] (\hat{r}_i(\eta) - \hat{r}_i(\theta)). \quad (10)$$

Cet estimateur a été étudié par M. Gba (1998) dans le cadre plus général des modèles dynamiques.

Sous certaines hypothèses (cf. M. Bonneau, M.Gba (1998)), ces deux estimateurs de θ_0 sont convergents et asymptotiquement efficaces, ce qui se traduit par $\sqrt{n}(\hat{\theta}_j - \theta_0) \rightarrow N_p(0, W^{-1}(\theta_0))$, quand $n \rightarrow \infty$ et pour $j = 1, 2$. L'objectif du paragraphe suivant est le calcul de ces estimateurs, afin de vérifier leurs propriétés de convergence.

3. Validation empirique

Pour étudier les performances à taille d'échantillon finie des deux estimateurs $\hat{\theta}_1$ ou $\hat{\theta}_2$, définis respectivement en (8) et (9), nous avons choisi deux modèles dont la loi \mathbb{P}_0 est telle que la densité conditionnelle de Y à $X = x$ est paramétré par $x \cdot \theta_0$. Pour le premier, la variance est une fonction de la moyenne. Les simulations sont réalisées pour différentes valeurs de n , variant de 50 à 400. Chaque simulation est répétée 100 fois afin d'obtenir des estimations du biais et de la variance à partir des statistiques empiriques.

La loi \mathbb{P}_0 du premier modèle (désigné par (a)) définie par

(a) pour tout $i = 1, \dots, n$, $Y_i = r_0(X_i\theta_0) + r_0(X_i\theta_0) \cdot \varepsilon_i$, où ε_i est une v.a. Gaussienne centrée réduite indépendante de X_i . La vraie valeur de la fonction r_0 est $r_0(t) = t + (1/2) \ln(1+t)$ et celle du paramètre $\theta_0 = (\theta_0^1, \theta_0^2) = (1, 2)$. Le couple de v.a. $X_i = (X_{i,1}, X_{i,2})$ est indépendant et de loi respectivement du chi-deux de paramètre 1 et exponentielle de paramètre 3.

La log-vraisemblance conditionnelle de Y_i à X_i dans le modèle (a) s'écrit, à une constante près

$$L_a(\theta) = (-1/2) \sum_{i=1}^n \left(\frac{Y_i - r_0(X_i \cdot \theta)}{r_0(X_i \cdot \theta)} \right)^2 - \sum_{i=1}^n \ln(r_0(X_i \cdot \theta)). \quad (11)$$

Le second modèle (b) est un modèle tobit dont la loi \mathbb{P}_0 est déterminée par

(b) pour tout $i = 1, \dots, n$, la v.a. $Y_i = Y_i^* 1_{\{Y_i^* > 0\}}$, avec $Y_i^* = X_i \cdot \theta_0 + \varepsilon_i$, où

$\theta_0 = (\theta_0^1, \theta_0^2) = (1, 2)$, ε_i est une v.a. gaussienne centrée réduite et indépendante de $X_i = (X_{i,1}, X_{i,2})$ qui un couple de v.a. indépendantes de loi respectivement du chi-deux de paramètre 1 et de loi gaussienne centrée de variance égale à 2.

Dans le modèle (b), la variance n'est plus fonction de la moyenne. Cependant les hypothèses (2) et (3) sont vérifiées pour les fonctions $w_0(t) = -r_0(t)^2 + tr_0(t) + \Phi(t)$ et $r_0(t) = t\Phi(t) + \phi(t)$, où Φ est la fonction de répartition de la loi gaussienne centrée réduite et ϕ sa dérivée⁵. Le logarithme de la vraisemblance conditionnelle s'écrit, à une constante près,

$$L_b(\theta) = (-1/2) \sum_{i=1/Y_i > 0} (Y_i - X_i \cdot \theta)^2 + \sum_{i=1/Y_i = 0} \ln \Phi(-X_i \cdot \theta), \quad (12)$$

Dans les deux modèles (a) et (b) qui vérifient (2) et (3), on supposera qu'en effet les vraies valeurs des fonctions r_0 et w_0 sont inconnues du statisticien. Pour le modèle (b) par exemple cela supposerait que le statisticien ne connaisse pas la loi de la perturbation ε_i dont la vraie distribution est une Gaussienne. On considère donc les estimateurs semi-paramétriques $\hat{\theta}$, $\hat{\theta}_1$ et $\hat{\theta}_2$, définis en respectivement en (6), (8) et (9). Les résultats obtenus sont synthétisés dans le tableau 1 et sont comparés à l'estimateur semi-paramétrique du maximum de vraisemblance (SMV) pour le modèle (a) qui maximise la pseudo vraisemblance

$$SL_a(\theta) = (-1/2) \sum_{i=1}^n \left(\frac{Y_i - \hat{r}_i(\theta)}{\hat{r}_i(\theta)} \right)^2 - \sum_{i=1}^n \ln(\hat{r}_i(\theta)), \quad (13)$$

⁵ voir Annexe

obtenue en remplaçant dans la vraisemblance $L_a(\theta)$ calculée en (11), le terme $r_0(X_i \cdot \theta)$ par $\hat{r}_i(\theta)$ défini en (4). On définit l'estimateur SMV pour le modèle (b), comme l'estimateur du maximum de la pseudo vraisemblance

$$SL_b(\theta) = \sum_{i=1/Y_i > 0} \frac{-1}{2} (Y_i - \hat{r}_i(\theta))^2 + \sum_{i=1/Y_i = 0} \ln \Phi(-\hat{r}_i(\theta)), \quad (14)$$

où $\hat{r}_i(\theta)$ est l'estimateur de $E(Y_i^* | X_i \theta)$. Nous rappelons que le tableau suivant seuls les estimateurs de la deuxième composante θ_0^2 de θ_0 sont présentés, la première étant toujours supposée égale à 1 pour rendre les modèles identifiables.

TABLEAU 1
Moyennes et écart-types empiriques des estimateurs
semi-paramétriques SMV (maximum de vraisemblance),
 $\tilde{\theta}^2$ (moindres carrés ordinaires),
 $\hat{\theta}_1^2$ et $\hat{\theta}_2^2$, pour différentes tailles n .

	Modèle (a)				Modèle (b)		
	S.M.V.	$\tilde{\theta}^2$	$\hat{\theta}_1^2$	$\hat{\theta}_2^2$	S.M.V.	$\hat{\theta}_1^2$	$\hat{\theta}_2^2$
$n = 50$	1.835 (0.852)	1.755 (1.445)	1.630 (2.714)	2.112 (0.467)	2.17 (0.381)	1.604 (1.071)	2.320 (0.764)
$n = 100$	1.833 (0.427)	1.915 (0.808)	1.532 (1.062)	2.191 (.401)	1.922 (0.378)	1.853 (0.610)	1.820 (0.564)
$n = 200$	1.817 (0.413)	1.824 (0.778)	1.770 (0.906)	2.149 (0.370)	2.013 (0.156)	1.955 (0.202)	1.950 (0.423)
$n = 400$	2.001 (0.366)	1.556 (0.562)	1.815 (0.385)	1.981 (0.254)	1.988 (0.115)	1.965 (0.140)	2.020 (0.239)

Dans le tableau 1, l'estimateur semi-paramétrique de maximum de vraisemblance (SMV) est le meilleur de tous les estimateurs dans les deux modèles. Ceci n'est pas surprenant car dans la construction du (SMV), on dispose d'une information supplémentaire qui est la loi conditionnelle de Y à X . A l'opposé l'estimateur des moindres carrés basé uniquement sur le premier moment est moins bon. La performance de cet estimateur étant déjà connue de tous, il est d'intérêt moindre par rapport aux autres estimateurs, c'est donc simplement pour cette raison qu'on ne l'a pas présenté pour le modèle (b) dans le tableau 1. Pour le modèle (a), à partir de $n = 400$, les variances des estimateurs SMV et $\hat{\theta}_1$ sont très proches. L'estimateur $\hat{\theta}_1$ est préférable à $\tilde{\theta}$, ce qui montre l'intérêt d'introduire l'estimation de w_0 . Cette remarque est d'autant plus vraie pour certains modèles dont la variance des observations apporte plus d'informations que la moyenne (voir Crowder, M. 1986).

Cette bonne performance de $\hat{\theta}_1$ est également mise en évidence pour le modèle (b). La variance empirique de l'estimateur $\hat{\theta}_2$ est faible pour des petites tailles

d'échantillon ($n = 50$). Par contre, bien que la vitesse de convergence en \sqrt{n} soit la même pour les 3 estimateurs SMV, $\hat{\theta}_1$ et $\hat{\theta}_2$ on peut remarquer que la convergence de l'estimateur $\hat{\theta}_2$ est plus lente. Les résultats obtenus confirment les bonnes performances des estimateurs $\hat{\theta}_1$ et $\hat{\theta}_2$, quand les seules hypothèses sur \mathbb{P}_0 sont celles considérées en (2) et (3), pour des fonctions r_0 et w_0 inconnues. Les calculs réalisés avec le logiciel Matlab, nécessitent des programmes d'optimisation qui dépendent de valeurs initiales. Si nous n'avons pas rencontré de problème de calcul pour ces données simulées, il est évident que la programmation est plus délicate pour des données réelles, à cause du problème d'optimisation. Le calcul de l'estimateur en deux étapes $\hat{\theta}_1$ est sensible à l'estimation préliminaire $\tilde{\theta}$, car il nécessite une optimisation de type moindres carrés pour la première étape, le calcul analytique pour la seconde ne posant pas de difficulté. L'estimateur minimax $\hat{\theta}_2$, est plus coûteux en temps, car il nécessite un balayage et une discrétisation, qui est plus ou moins longue, selon la procédure minimax choisie.

En conclusion, les résultats obtenus dans ces simulations confirment les bonnes performances des estimateurs $\hat{\theta}_1$ et $\hat{\theta}_2$, quand les seules hypothèses sur P_0 sont celles considérées en (2) et (3), pour des fonctions r_0 et w_0 inconnues. A partir de $n = 400$, l'utilisation de ces estimateurs est satisfaisante pour nos exemples où le nombre de variables explicatives est 2. Il conviendrait d'augmenter la taille de l'échantillon si on augmente le nombre de variables explicatives.

Annexe. Moment de la loi tronquée

Dans les calculs suivants l'espérance est prise par rapport à loi \mathbb{P}_0 .

$$\begin{aligned} E(Y_i/X_i) &= E((X_i\theta_0 + \varepsilon_i)1_{\{X_i\theta_0 + \varepsilon_i > 0\}}|X_i) \\ &= X_i\theta_0 E(1_{\{X_i\theta_0 + \varepsilon_i > 0\}}|X_i) + E(\varepsilon_i 1_{\{X_i\theta_0 + \varepsilon_i > 0\}}|X_i) \\ &= X_i\theta_0 \mathbb{P}(\varepsilon_i > -X_i\theta_0|X_i) + E(\varepsilon_i 1_{\{\varepsilon_i > -X_i\theta_0\}}|X_i). \end{aligned}$$

La loi de ε_i étant une Gaussienne centrée réduite et indépendante de X_i , on trouve finalement

$$\begin{aligned} E(Y_i|X_i) &= X_i\theta_0(1 - \Phi(-X_i\theta_0)) + \int_{-X_i\theta_0}^{+\infty} u\phi(u)du \\ &= X_i\theta_0\Phi(X_i\theta_0) + \phi(X_i\theta_0). \end{aligned}$$

$$\begin{aligned} E(Y_1^2|X_i) &= (X_i\theta_0)^2\mathbb{P}(\varepsilon_i > -X_i\theta_0|X_i) + 2X_i\theta_0 E(\{\varepsilon_i 1_{\{\varepsilon_i > -X_i\theta_0\}}\}|X_i) \\ &\quad + E(\varepsilon_i^2 1_{\{\varepsilon_i > -X_i\theta_0\}}|X_i) \cdot E(\varepsilon_i^2 1_{\{\varepsilon_i > -X_i\theta_0\}}|X_i) \\ &= \int_{-X_i\theta_0}^{+\infty} u^2\phi(u)du \\ &= [-u\phi(u)]_{-X_i\theta_0}^{+\infty} + \int_{-X_i\theta_0}^{+\infty} \phi(u)du = -X_i\theta_0\phi(X_i\theta_0) + \Phi(X_i\theta_0). \end{aligned}$$

Donc

$$E(Y_i^2|X_i) = (X_i\theta_0)^2\Phi(X_i\theta_0) + X_i\theta_0\phi(X_i\theta_0) + \Phi(X_i\theta_0).$$

En conséquence

$$\begin{aligned} \text{Var}(Y_i|X_i) = & -(X_i\theta_0\Phi(X_i\theta_0) + \phi(X_i\theta_0))^2 \\ & + X_i\theta_0(X_i\theta_0\Phi(X_i\theta_0) + \phi(X_i\theta_0)) + \Phi(X_i\theta_0). \end{aligned}$$

Références

- [1] BONNEU M., (1995), Lien non spécifié dans les modèles linéaires généralisés. *Rev. Statistique Appliquées*, **XLIII(2)**, 49-66.
- [2] BONNEU M. & GBA M., (1998), Estimation semi-paramétrique de quasi-score. *Bull. Belg. Math. Soc.* **5**, 693-712.
- [3] CROWDER M., (1986), On consistency and inconsistency of estimating equations. *Econometric theory* **2**, 305-330.
- [4] GBA M., (1998), Minimax semiparametric estimation in dynamic regression. *Soumis à Multivariate Analysis*.
- [5] ICHIMURA H., (1993), Semiparametric least squares (SLS) and weighted (WLS) estimation of single index models. *Journal of Econometrics*, **58**, 71-720.
- [6] LI B., (1996), A minimax approach to consistency and efficiency for estimated equations. *The Annals of Statistics*, **24, 3**, 1283-1297.
- [7] MAC CULLAGH P., (1983), Quasi-likelihood functions. *The Annals of Statistics*, **11**, 59-67.
- [8] NEWBY W., (1990), Semiparametric efficient bound. *Journal of Applied Econometrics*, **5**, 99-135.
- [9] WEFELMEYER W., (1996), Quasi-likelihood models and optimal inference. *The Annals of Statistics*, **24, 1**, 405-422.