Revue de statistique appliquée

M. Bonneu

Lien non spécifié dans les modèles linéaires généralisés

Revue de statistique appliquée, tome 43, nº 2 (1995), p. 49-66

<http://www.numdam.org/item?id=RSA_1995__43_2_49_0>

Rev. Statistique Appliquée, 1995, **XLIII** (2), 49-66

LIEN NON SPÉCIFIÉ DANS LES MODÈLES LINÉAIRES GÉNÉRALISÉS

M. Bonneu

GREMAQ Université des Sciences Sociales
Place Anatole France Toulouse Cedex France
LSP URA D0745 Université Paul Sabatier
118 route de Narbonne 31062 Toulouse France

RÉSUMÉ

Dans cet article, nous synthétisons les méthologies qui répondent au problème de la non spécification de la fonction de lien dans les modèles linéaires généralisés. Nous privilégions l'approche semi paramétrique où la fonction de lien est estimée par un estimateur à noyau de régression. On définit alors un estimateur des coefficients de régression, par maximum d'une pseudo vraisemblance. Nous précisons les propriétés de cet estimateur et l'algorithme de résolution numérique. Une comparaison est faite, avec d'autres estimateurs semi ou non paramétriques. Dans le cas du modèle dichotomique, les résultats obtenus par simulation permettent d'étudier les performances de l'estimateur du maximum de pseudo vraisemblance.

Mots-clés : *Modèle linéaire généralisé – Fonction de lien inconnue – Estimation du maximum de pseudo-vraisemblance – Estimation semi paramétrique – Estimateur à noyau de régression.*

SUMMARY

This paper is a survey of the methods dealing with missing link problems in generalized linear models. We specialize in the nonparametric approach when the link function is estimated by a kernel regression estimator. We then define an estimator for the regression coefficients, using pseudo maximum likelihood methods. We derive the properties of this estimator and we present a numerical algorithm for its computation. We perform numerical comparisons of these estimators with semi or nonparametric estimators. In a binary response model, the simulation results give some insight into the performances of the pseudo maximum likelihood estimator.

Keywords : *Generalized linear model - Link missing - Pseudo maximum likelihood estimation - Semi parametric estimation - Kernel regression estimator.*

1.Introduction

Les modèles linéaires généralisés (MLG) introduits par Nelder et Wedderburn (1972) sont une extension du modèle linéaire gaussien car ils permettent de considérer une loi de probabilité autre que la loi gaussienne et un lien autre que l'identité.

Les données statistiques sont les n observations (y_i, x_i) pour $i = 1, n$.

La variable à expliquer Y est un vecteur aléatoire de $\mathbb{R}^q$. Le vecteur x des p régresseurs $x = (x^1, \cdots, x^p)$, est, soit un vecteur aléatoire de $\mathbb{R}^p$, soit un vecteur non aléatoire $(p < n)$. On considérera, dans ce travail la loi de Y conditionnellement à $X = x$, caractérisée par l'existence de la densité conditionnelle de Y/X par rapport à une mesure σ-finie, non concentrée en un point.

Nous rappelons brièvement dans le paragraphe 1, les définitions et les résultats classiques du MLG. Dans le paragraphe 2, nous présentons le problème du lien non spécifié et les approches de solution paramétriques ou semi paramétriques. Dans le cadre des MLG semi paramétriques, nous proposons dans le paragraphe 3 un estimateur du type maximum de pseudo vraisemblance pour lequel on précise les propriétés de convergence *p.s.*. Cet estimateur fournit une estimation de la fonction de lien. Nous rappelons la définition d'autres estimateurs semi paramétriques. Pour des données simulées dans le modèle dichotomique, une comparaison empirique souligne les performances de l'estimateur proposé.

1.1 Définition des Modèles Linéaires Généralisés

Les MLG ont été largement développés : Mc Cullagh, Nelder (1983), Aitkin *et al* (1990), Lindsey (1989).

Dans le cadre de ce travail, nous définissons un modèle linéaire généralisé comme un modèle paramétrique défini par les 3 hypothèses suivantes :

H 1) Les v.a. $(Y_i, X_i)_{i=1,n}$ sont i.i.d.. La densité conditionnelle de $Y_i \mid X_i$ appartient à la famille naturelle exponentielle et s'écrit :

$$f(y_i; \theta_i, \varphi) = c(y, \varphi)\exp[(y \cdot \theta_i - b(\theta_i))/a(\varphi)] \tag{1}$$

pour tout $i = 1, n$ et tout y appartenant à un ouvert $I \subset \mathbb{R}^q$, qui est le support de f.

Les fonctions $a(\cdot)$, $b(\cdot)$ et $c(\cdot)$ sont des fonctions définies et connues ($a(\varphi) > 0$, $c(y, \varphi) > 0$). La fonction $b()$ est strictement concave, deux fois différentiable.

Le produit scalaire des vecteurs y et θ_i de $\mathbb{R}^q$ est noté $y \cdot \theta_i$. Les paramètres inconnus θ_i sont appelés paramètres naturels, alors que φ est un paramètre réel de nuisance.

H 2) Le vecteur x_i intervient dans l'explication de y_i à travers le prédicteur linéaire :

$$\gamma_i = \chi_i \beta \qquad (i = 1, n)$$

où le paramètre inconnu β est de dimension $p(p < n)$. La matrice χ_i de dimension $(q \times p)$ est déterminée par une fonction connue des valeurs $x_i^1, \cdots, x_i^p$.

H 3) Le prédicteur linéaire γ_i est lié à l'espérance $E(Y_i \mid X_i = x_i)$ notée μ_i ou $\mu_i(\theta_i)$, par une fonction vectorielle r :

$$\mu_i = r(\gamma_i) \qquad (i = 1, n)$$

La fonction r, deux fois continûment différentiable, sera considérée comme la fonction de lien.

La fonction g de lien au sens de Nelder, définie par $g(\mu_i) = \gamma_i$, n'est autre que l'inverse de r, dans le cas où g est biunivoque.

Remarque :

L'expression (1) permet d'exprimer l'espérance conditionnelle de $Y_i \mid X_i$ en fonction de θ_i :

$$\mu_i = b'(\theta_i)$$

La fonction $b'(\cdot)$ est strictement croissante et il est possible de reparamétrer la densité en fonction du paramètre μ_i qui appartient à un ouvert M de $\mathbb{R}$. La variance conditionnelle de $Y_i \mid X_i$, égale à $b''(\theta) \cdot a(\varphi)$ est donc fonction de μ_i et du paramètre de nuisance φ.

La fonction de lien naturelle ou canonique correspond à $r_0 = b'$. Le paramètre naturel θ_i s'écrit donc :

$$\theta_i = r_0^{-1}(r(\chi_i\beta))$$

1.2 Exemples

Dans le cas unidimensionnel ($q = 1$), les modèles MLG permettent de traiter une classe très large de modèles de régression quand la densité (1) est celle d'une loi gaussienne, binomiale, Poisson, gamma, ou inverse gaussienne; par exemple, pour la densité gaussienne et la fonction de lien Identité, on obtient le modèle classique linéaire gaussien.

Dans le cas de variables entières ($I = \mathbb{N}$), il est naturel de considérer la loi de Poisson et pour la fonction de lien ln $(\cdot)$, on obtient le modèle loglinéaire. Le cas des tables de contingence peut être traité par un tel modèle loglinéaire, en remarquant que la distribution de v.a. indépendantes de loi de Poisson conditionnellement à leur somme est multinomiale. Dans le cas de variables positives ($I =]0, +\infty[$), l'analyse statistique des durées de vie peut être réalisée par des MLG en considérant la loi gamma.

Un exemple de MLG très utilisé, est le modèle dichotomique ($I = \{0, 1\}$) dans le cas d'expériences indépendantes de Bernoulli, défini classiquement à partir d'une variable latente y^* non observable :

$$y_i = \mathbb{I}_{\{y_i^* < 0\}} = \mathbb{I}_{\{\varepsilon_i < x_i\beta\}}$$

où la v.a. ε_i est le résidu d'une régression linéaire de la variable latente y_i^* sur le vecteur x_i des p régresseurs tel que $y_i^* = \varepsilon_i - x_i\beta$. Si on suppose que les v.a. ε_i sont i.i.d., l'espérance conditionnelle de y_i s'écrit :

$$\mu_i = r(x_i\beta)$$

La fonction de lien r représente alors la fonction de répartition de ε_i. Si on choisit $r = \Phi$ (fonction de répartition de la loi gaussienne réduite), le MLG correspondant est le modèle probit. Si on choisit $r(t) = (1 - e^{-t})^{-1}$ il s'agit du modèle logit.

D'autre fonctions de lien $r(\cdot)$ dissymétriques généralisant le modèle logit ont été étudiées (Pregibon (1981)). Le choix de la fonction r n'est pas toujours justifié et peut conduire à une mauvaise spécification du modèle. C'est ce point de vue qui sera développé dans l'application du paragraphe 3.

Dans le cas multidimensionnel, le cas le plus naturellement considéré est le modèle multinomial logit ou probit des catégories ordonnées ou non. Pour illustrer ce cas, nous considérons une application sur des données de distributions de revenus de communes, quand on dispose de la répartition empirique pour des données groupées en classes.

On dispose de n communes et pour chaque commune, de L classes identiques de revenus de ménages. Soit $y_i = (n_{i1}, \cdots, n_{iL})$ le vecteur des effectifs $n_{i\ell}$ correspondant à la ℓ^{eme} classe de revenu et à la i^{eme} commune ($\ell = 1, \cdots, L$ et $i = 1, \cdots, n$). Soit $x_i = (x_i^1, \cdots, x_i^p)$ le vecteur de p régresseurs caractérisant la commune. On note $p_{i\ell}$ la probabilité qu'un ménage j de la commune i, ait un revenu R_{ij} se trouvant dans la classe ℓ ($j = 1, \cdots, n_i$ avec $n_i = \sum_{\ell=1}^{L} n_{i\ell}$)

Si on considère le modèle de régression pour la variable latente R_{ij} :

$$\mathrm{Log} R_{ij} = x_i\beta + \varepsilon_{ij} \qquad \text{où les } v.a.\ \varepsilon_{ij} \text{ sont centrées i.i.d.}$$

La probabilité $p_{i\ell}$ s'exprime en fonction des logarithmes des bornes de la classe ℓ, notées : $[b_{\ell-1}, b_\ell[$, c'est à dire :

$$p_{i\ell} = F(\ln b_\ell - x_i\beta) - F(\ln b_{\ell-1} - x_i\beta)$$

où F est la fonction de répartition des v.a. ε_{ij}. Les modèlisations classiques considèrent le modèle probit $F = \Phi$ (ou dans le cas logit : $F(t) = (1 - e^{-t})^{-1}$), ce qui revient à supposer que les v.a. ε_{ij} sont de même loi gaussienne. Le paramètre β est alors déterminé en maximisant la vraisemblance de $Y_1, \cdots, Y_I$ quand la distribution conditionnelle de Y_i sachant $X_i = x_i$ est une loi multinomiale de paramètre n_i et $(p_{i1}, \cdots, p_{iL})$.

Si on se place dans le cadre semi paramétrique la fonction F sera un paramètre fonctionnel inconnu non spécifié.

1.3 Estimateur du Maximum de Vraisemblance (EMV)

On considère dans la suite de ce travail, le cas unidimensionnel ($q = 1$). On peut reparamétriser la densité de $Y_i \mid X_i$, soit en fonction de μ_i, soit en fonction de β.

Le paramètre de nuisance n'intervient que dans certains MLG. Sans perdre de généralité, on considérera dans la suite $a(\varphi) = 1$, ce qui est vérifié pour le cas binomial et Poisson. Pour les autres cas, le cadre de ce travail reste encore valable, si on dispose d'un estimateur, convergent en probabilité du paramètre φ.

Comme $q = 1$, l'hypothèse $H2$ peut s'écrire sans perte de généralité :
$\gamma_i = x_i\beta = \sum_{j=1}^{p} x_i^j \beta_j$.

Compte tenu des remarques précédentes, l'estimateur du maximum de vraisemblance $\hat{\beta}_{MV}$ du paramètre β, est déterminé par :

$$\hat{\beta}_{MV} = \arg\max_{\beta \in B} L_n(\beta) \tag{2}$$

$$\text{avec } L_n(\beta) = \sum_{i=1}^{n} f^*(y_i, r(x_i\beta))$$

$$f^*(y_i, r(x_i\beta)) = [y_i A(r(x_i\beta)) - B(r(x_i\beta))]$$

L'expression de f^* se déduit de (1), en considérant le logarithme de f et les termes qui dépendent de β. Les fonctions A et B définies sur $M \subset \boldsymbol{R}$, sont donc : $A = b'^{-1}; B = b \circ b'^{-1}$. Elles vérifient notamment : $B'(\mu) = \mu A'(\mu)$

Le tableau 1 ci-dessous précise pour les lois usuelles les expressions des fonctions A et B, ainsi que leurs domaines de définition M et les fonctions de lien canonique r_0 définies sur $A(M)$:

TABLEAU 1

	$A(\mu)$	$B(\mu)$	M	Fonction $r_0(\gamma)$ dans le cas du lien canonique
Loi normale	μ	$\mu^2/2$	$\boldsymbol{R}$	γ
Loi Bernoulli	logit μ	$-\ln(1-\mu)$	$]0, 1[$	$(1 - e^{-\gamma})^{-1}$
Loi Poisson	$\ln \mu$	μ	$]0, +\infty[$	e^{γ}
Loi Gamma	$-\frac{1}{\mu}$	$\ln \mu$	$]0, +\infty[$	$-\frac{1}{\gamma}$

L'estimateur $\hat{\beta}_{MV}$, solution de $\frac{\partial L_n(\beta)}{\partial_\beta} = 0$, vérifie les équations normales :

$$^tXW_\beta(y - \mu_\beta) = 0 \tag{3}$$

où X est la matrice $(n \times p)$ des p régresseurs, W_β la matrice diagonale d'ordre n de terme diagonal $A'(r(x_i\beta)) \times r'(x_i\beta)$. Les vecteurs y et μ_β sont des vecteurs colonnes de composantes respectivement y_i et $r(x_i\beta)$. Le terme $A'(r(x_i\beta))$ n'est autre que l'inverse de la variance conditionnelle de $Y_i \mid X_i$. L'algorithme de résolution de $\hat{\beta}_{MV}$, mis en œuvre dans le logiciel $GLIM$, est un algorithme de moindres carrés pondérés avec itération. En effet, on considère le changement de variable défini pour le vecteur z_β de composante :

$$z_i = x_i\beta + (y_i - r(x_i\beta))/r'(x_i\beta) \qquad (i = 1, n)$$

Les équations normales (3) s'écrivent alors :

$$ {}^{t}XV_{\beta}^{-1}(z_{\beta} - X\beta) = 0 $$

où V_{β} est la matrice de variance de z_{β} :

$$ V_{\beta}^{-1} = \text{ diag } (A'(r(x_i\beta)) \times (r'(x_i\beta))^2) $$

Pour les MLG où la fonction de lien n'est pas nécessairement canonique, les théorèmes de convergence de $\hat{\beta}_{MV}$ et la théorie des tests asymptotiques ont été établis par Mathieu (1981), Fahrmeir, Kaufmann (1985), Fahrmeir (1987). L'unicité de $\hat{\beta}_{MV}$ n'est pas garantie pour une fonction de lien non canonique. Ces propriétés de convergence ne sont plus valables si le MLG est mal spécifié : densité ou fonction de lien incorrecte, régresseurs omis. Nous considérons dans ce travail, le cas d'une mauvaise spécification de la fonction de lien. Le choix de la fonction de lien est lié à la sélection des régresseurs, qui peut être différente suivant la spécification de la fonction de lien. Dans le cas d'une mauvaise spécification de la densité, on peut considérer la quasi vraisemblance (Mc Cullagh (1983), Fahrmeir (1990)).

En effet l'écriture des équations normales (3) permet de généraliser les MLG au modèles de quasi vraisemblance (Mc Cullagh – 1983) définis par :

$$ \mu_i = E(Y_i \mid X_i = x) = r(x_i\beta) \qquad V(Y_i \mid X_i = x) = \xi V(\mu_i) $$

où r et V sont des fonctions connues. L'estimateur de quasi vraisemblance est solution des équations :

$$ \sum_{i=1}^{n} \frac{\partial r(x_i\beta)}{\partial\beta} \cdot \frac{(y_i - r(x_i\beta))}{\xi V(\mu_i)} = 0 $$

L'estimateur de β par quasi vraisemblance consiste à "approcher" les observations ci-dessus par un modèle linéaire généralisé tel que la fonction de lien au sens de Nelder et la fonction variance soient respectivement r^{-1} et V.

Le cas où on ne spécifie pas la fonction de lien, que nous considérons dans les paragraphes suivants pour les MLG, pourrait se généraliser aux modèles de quasi-vraisemblance.

2. Fonction de lien inconnue

Quand, on ne dispose d'aucune information sur la fonction de lien r, les réponses au problème du choix de r, diffèrent selon que l'on considère les méthodologies en statistique paramétrique ou non paramétrique.

2.1 Choix de la fonction de lien dans le cas paramétrique

– Une première approche consiste à construire une famille paramétrique de fonctions de lien qui contienne la fonction de lien canonique. Cette famille de fonctions dépend d'un paramètre ψ de dimension finie, appelé paramètre de lien. Dans le cas du modèle dichotomique, des familles généralisant le modèle logit ont été proposées par Atkinson (1982), Pregibon (1981), Stukel (1988). L'introduction de ce paramètre supplémentaire ψ a l'inconvénient d'augmenter la variance estimée de l'estimateur MV du paramètre d'intérêt β, défini en (2). Des familles de fonction de lien standardisées, réduisant cette inflation de variance, ont été introduites dans le cas général du MLG (Czado(1992)), avec pour autre objectif de palier aux problèmes d'instabilité numérique dans le calcul de l'EMV pour une fonction de lien non naturelle :

Soit $\gamma \in$ vect $\{x^1, \cdots, x^p\}$, sous espace vectoriel engendré par les p composantes de x et soit $\psi \in \boldsymbol{R}$. Si l'on considère la fonction de lien canonique $r_0(\gamma) = b'(\gamma)$ (*cf.* tableau 1), la famille de fonctions $r(\cdot, \psi)$ est alors définie par :

$$r(\gamma, \psi) = r_0(h_{\gamma_0}(\gamma, \psi))$$

où la fonction $h_{\gamma_0}(\gamma, \psi)$ caractérisée pour un paramètre γ_0 approprié est définie par :

$$h_{\gamma_0}(\gamma, \psi) = \left(\gamma_0 + \frac{(\gamma - \gamma_0 + 1)^\psi - 1}{\psi}\right) I\!I_{\{\gamma \geqslant \gamma_0\}}(\gamma) + \gamma I\!I_{\{\gamma < \gamma_0\}}(\gamma)$$

$$\text{ou par} \quad h_{\gamma_0}(\gamma, \psi) = \gamma I\!I_{\{\gamma \geqslant \gamma_0\}}(\gamma) + \left(\gamma_0 - \frac{(-\gamma + \gamma_0 + 1)^\psi - 1}{\psi}\right) I\!I_{\{\gamma < \gamma_0\}}(\gamma)$$

Pour $\psi = 1$, $h_{\gamma_0}(\gamma, 1)$ est l'identité et dans ce cas la fonction $r(\gamma, 1)$ n'est autre que la fonction de lieu canonique $r_0(\gamma)$. La fonction $h_{\gamma_0}(\gamma, \psi)$ est dite γ_0 standardisé, au sens de la propriété suivante :

$$h_{\gamma_0}(\gamma_0, \psi) = \gamma_0 \qquad \text{et} \qquad \left[\frac{\partial}{\partial \psi} h_{\gamma_0}(\gamma, \psi)\right]_{\gamma = \gamma_0} \quad \text{indépendant de } \psi.$$

– Une deuxième approche est développée dans le cas de tests non emboîtés. Dans un catalogue de fonctions de lien, il est alors possible de considérer deux fonctions r_1 et r_2, et les modèles linéaires généralisés associés : M_1 défini pour r_1 et M_2 défini pour r_2.

On note $\ell_i(\hat\beta_1, \hat\beta_2) = f^*(y_i, r_1(x_i\hat\beta_1)) - f^*(y_i, r_2(x_i\hat\beta_2))$ le logarithme du rapport des densités estimées ($\hat\beta_1$ et $\hat\beta_2$ sont les EMV respectifs dans les modèles M_1 et M_2).

Dans un cadre plus général, Vuong (1989) propose un test non emboîté, permettant de sélectionner l'un des 2 modèles M_1 ou M_2.

Le test est basé sur la propriété de convergence en loi qui s'écrit sous l'hypothèse nulle d'égalité des 2 fonctions r_1 et r_2 :

$$\frac{1}{n} \sum_{i=1}^{n} \ell_i(\hat\beta_1, \hat\beta_2)/s \xrightarrow[n \to +\infty]{\text{Loi}} \mathcal{N}(0,1)$$

$$\text{où} \;\; s = [\tfrac{1}{n} \sum_{i=1}^{n} \{\ell_i(\hat\beta_1, \hat\beta_2) - \frac{1}{n} \sum_{i=1}^{n} \ell_i(\hat\beta_1, \hat\beta_2)\}^2]^{1/2}$$

Mais ce test ne permet pas de répondre au problème de spécification de la fonction de lien, il ne permet pas notamment de choisir r_1 contre r_2.

– Une troisième approche est celle des techniques de choix de modèles.

La méthodologie de choix de modèles est présentée dans un cadre général dans le livre de Linhart et Zucchini (1986). Elle consiste à définir, au préalable, une mesure de dissemblance entre un vrai modèle et une famille de pseudo modèles. L'objectif est alors de choisir le pseudo modèle qui minimise une fonction de risque basée sur cette mesure de dissemblance. Il est nécessaire, en général, d'estimer cette fonction de risque qui dépend de paramètres inconnus.

Dans le cas des MLG, la mesure de dissemblance est la mesure d'information de Kullback calculée entre le modèle saturé (qui contient autant de paramètres naturels θ_i que d'observations) et le MLG ayant pour fonction de lien r. La fonction de risque est définie par :

$$E_\theta(\mathcal{J}(\theta, \hat{\beta}_{MV}))$$

$$\text{où}\quad \mathcal{J}(\theta,\beta) = -\int \ln \prod_{i=1}^{n} f(y_i \mid x_i; b'(r(x_i\beta))) \times \prod_{i=1}^{n}(f(y_i \mid x_i; \theta_i)d\sigma(y_i))$$

mesure de dissemblance déterminée à partir de l'expression (1) de f pour une fonction de lien r donnée. L'espérance mathématique E_θ est prise, relativement au vrai modèle qui est le modèle saturé. Cette fonction de risque se calcule analytiquement dans le modèle linéaire gaussien, mais doit être estimée dans les autres cas, soit par estimation simulée du bootstrap, soit par estimation asymptotique analogue au critère d'Akaike, adapté aux MLG. (Bonneu (1988), Bonneu, Milhaud (1993)).

Dans un catalogue de fonction de lien r, il est alors possible de spécifier le lien en choisissant la fonction r qui correspond au minimum de l'estimation de $E(\mathcal{J}(\theta, \hat{\beta}_{MV}))$.

2.2 Modèles linéaires généralisés semi paramétriques

Dans le cadre semi paramétrique, on considère la fonction de lien r comme un paramètre fonctionnel inconnu. L'hypothèse H_1 est toujours vérifiée, a savoir que la loi conditionnelle de la $v.a.r$ Y_i sachant X_i appartient à la famille exponentielle naturelle, de densité explicitée en (1) pour $\varphi = 1$.

Soit l'hypothèse $H2'$:

H 2') il existe $\alpha \in \mathcal{B} = \{(\beta_1, \cdots, \beta_p) \in \mathbb{R}^p / \beta_1 = 1\}$ tel que :

$$m(x) = E(Y_i \mid X_i = x) = E(Y_i \mid X_i\alpha = x\alpha) \tag{4}$$

où α est la vraie valeur du paramètre représentant les coefficients de régression.

L'espace paramétrique $\mathcal{B}$ est ainsi défini, pour que le modèle soit identifiable, ce qui sera justifié dans la propriété suivante.

L'hypothèse H_2' signifie que $m(x)$ ne dépend de x que par l'intermédiaire de $x\alpha$. En effet, α est la forme linéaire sur $\mathbb{R}^p$ telle que : $E(Y_i \mid \alpha X_i = u) = (m \circ \alpha^{-1})(u)$.

Définition 1

On appelle MLG semi paramétrique le modèle défini par les hypothèses H_1 et H_2' ∎

Remarque :

Cette définition différe de l'approche semi paramétrique (Green et Yandell 1985), qui consiste à décomposer $m(x)$, à l'aide d'une fonction g inconnue à valeur réelle :

$$m(x) = r(x^1\alpha + g(x^2)), \text{ où } x = (x^1, x^2)$$

Dans ce cas la fonction r est supposée connue et les paramètres à estimer sont α et g, ce qui n'est pas le cadre de notre étude.

Définition 2

Soit S le support de la densité de X_i, on définit pour tout $\beta \in \mathcal{B}$ la fonction $r_\beta(\cdot)$ qui à $z \in \{x\beta / x \in S, \beta \in \mathcal{B}\}$ associe le réel :

$$r_\beta(z) = E(Y_i \mid X_i\beta = z) \qquad (5)$$ ∎

De cette définition, on déduit que la fonction $r_\beta(\cdot)$ dépend de α, de β et de la loi de x_i. Elle vérifie notamment :

$$m(x) = r_\alpha(x\alpha)$$

Ce qui revient à considérer que la fonction de lien au sens de l'hypothèse H_3 est le paramètre fonctionnel $r_\alpha(\cdot)$ inconnue qui dépend donc du paramètre α de dimension finie.

Propriété : Sous les hypothèses :

(i) $\mathcal{B} = \{(\beta_1, \cdots, \beta_p) \in \mathbb{R}^p / \beta_1 = 1\}$

(ii) Pour tout $\beta \in \mathcal{B}$, la fonction $r_\beta(z)$ est dérivable en tout point $z \in \{x\beta / \, x \in S, \; \beta \in \mathcal{B}\}$.
alors, pour tout $(\beta, \beta^*) \in \mathcal{B}^2$, $\qquad r_\beta(x\beta) = r_{\beta^*}(x\beta^*) \Longrightarrow \beta = \beta^*$. ∎

En effet, d'après (ii), l'égalité $r_{\beta^*}(x\beta) = r_{\beta^*}(x\beta^*)$ implique en dérivant par rapport à x :

$$r'_\beta(x\beta) \cdot \beta = r'_{\beta^*}(x\beta^*) \cdot \beta^*$$

D'après la contrainte $\beta_1 = \beta_1^* = 1$ sur l'espace paramètrique $\mathcal{B}$, on en déduit que $r'_\beta(x\beta) = r'_{\beta^*}(x\beta^*)$, et donc que $\beta = \beta^*$.

Exemple dans le cas gaussien

On se place dans le cas où (Y_i, X_i) est un vecteur gaussien centré : la loi de $Y_i \mid X_i = x$ est gaussienne de moyenne $x\alpha = < x, \alpha >$ et de variance 1. Cette hypothèse n'est pas vérifiée pour toute loi gaussienne puisque cela suppose que la covariance de Y_i avec la 1ère composante de X_i vaut 1. Le vecteur aléatoire X_i est gaussien supposé sans perte de généralité, centré, de matrice variance I_p. Dans ce cas α n'est autre que le vecteur cov (X_i, Y_i) égal à $(1, \alpha_2)$.

Dans ce cas la fonction $r_\alpha(\cdot)$ est l'identité. L'expression de $r_\beta(z) = E(Y_i \mid X_i\beta = z)$ se calcule aussi comme l'espérance linéaire de Y_i sachant $X_i\beta = z$:

$$r_\beta(z) = EL(Y_i \mid X_i\beta = z) = E(Y_i) + \text{cov}\,(Y_i, X_i\beta)[\text{Var } X_i\beta]^{-1}(z - E(X_i))$$

C'est-à-dire :

$$r_\beta(z) = \frac{< \beta, \alpha >}{\| \beta \|^2} \cdot z$$

avec ici $E(X_i) = E(Y_i) = 0$, puisque le couple (Y_i, X_i) est centré.

3. Estimation Semi Paramétrique

3.1 Estimateur du maximum de Pseudo Vraisemblance du MLG

Dans les modèles semi paramétriques, pour estimer le paramètre d'intérêt α de dimension finie, on construit préliminairement un estimateur non paramétrique du paramètre fonctionnel inconnu. C'est le cadre général des estimateurs $MINPIN$ (Andrews (1989)), qui MINimisent un critère qui dépend d'un estimateur Préliminaire du paramètre de dimension Infinie et de Nuisance.

La difficulté dans notre cas est que le paramètre fonctionnel dépend du paramètre d'intérêt.

Pour β fixé, on estime alors naturellement $r_\beta(x_i\beta) = E(Y_i \mid X_i\beta = x_i\beta)$ par l'estimateur à noyau de la régression de y_i par $x_i\beta$ de Nadaraya-Watson :

$$\hat{r}_\beta(x_i\beta) = \frac{\sum_{j \neq i}^{n} y_j \cdot K_{a_n}(x_i\beta - x_j\beta)}{\sum_{j \neq i}^{n} K_{a_n}(x_i\beta - x_j\beta)} \tag{6}$$

où $K_{a_n}(\cdot) = \frac{1}{a_n}K(\cdot/a_n)$ pour un noyau K symétrique et a_n une fenêtre. Nous considérons en fait pour K, le noyau gaussien $K(u) = (2\pi)^{-\frac{1}{2}} \cdot \exp(-u^2/2)$, ce choix étant suffisant pour les résultats théoriques de convergence et étant performant dans les simulations. Notons que la définition de $\hat{r}_\beta(x_i\beta)$ qui omet $x_i\beta$ dans l'estimation a l'avantage technique des estimateurs du jacknife (Bickel (1982)).

L'estimateur du paramètre d'intérêt n'est plus l'estimateur MV $\hat{\beta}_{MV}$ défini en (2), car la fonction $L_n(\beta)$ à maximiser dépend de $r_\beta(\cdot)$ qui est inconnue :

$$L_n(\beta) = \sum_{i=1}^{n} \{y_i \, A(r_\beta(x_i\beta)) - B(r_\beta(x_i\beta))\} \tag{7}$$

L'estimateur est du type maximum de pseudo vraisemblance (MPV) (Gong, Samaniengo (1981), Bonneu, Delecroix (1993)). On maximise alors la pseudo vraisemblance obtenue en remplaçant dans la fonction $L_n(\beta)$ définie en (7), $r_\beta(x_i\beta)$ par $\hat{r}_\beta(x_i\beta)$ qui est défini en (6).

Définition 3

On définit donc l'estimateur MPV par :

$$\hat{\beta}_{MPV} = \arg\max_{\beta \in \mathcal{B}} Q_n(\beta) \tag{8}$$

où $$Q_n(\beta) = \sum_{i=1}^{n} \{y_i \cdot A(\hat{r}_\beta(x_i\beta)) - B(\hat{r}_\beta(x_i\beta))\} \qquad \blacksquare$$

L'estimateur de la fonction de lien se déduit alors des estimateurs définis en (6) et (8) pour l'estimation de $m(x) = E(Y_i \mid X_i = x)$:

$$\hat{m}(x) = \hat{r}_{\hat{\beta}_{MPV}}(x\hat{\beta}_{MPV}) \qquad \text{pour tout } x \in S. \tag{9}$$

• **Convergence de l'estimateur $\hat{\beta}_{MPV}$**

a) Les propriétés de convergence en probabilité et p.s. de $\hat{\beta}_{MPV}$ vers α se déduisent d'un théorème démontré dans un cadre plus général que les modèles MLG (*cf.* Bonneu, Delecroix (1993)). Le principe de la démonstration de la convergence p.s. se déroule en 3 étapes :

(i) $Q(\beta) = E\{Y_i A(r_\beta(X_i\beta) - B(r_\beta(X_i\beta))\}$ a un maximum unique en α.

(ii) $\lim_{p.s.}(\sup_{\beta \in \mathcal{B}} \frac{1}{n} \mid L_n(\beta) - Q_n(\beta) \mid) = 0$

(iii) $\lim_{p.s.}(\sup_{\beta \in \mathcal{B}} \mid \frac{1}{n} L_n(\beta) - Q(\beta) \mid) = 0.$

Ces résultats de convergence uniforme p.s. nécessitent le résultat :

$$\lim_{p.s.}(\sup_{1 \leqslant i \leqslant n} \sup_{\beta \in \mathcal{B}} \mid \hat{r}_\beta(X_i\beta) - r(X_i\beta) \mid) = 0$$

Ils sont assurés sous certaines conditions, notamment sur la densité de $X\beta$, sur le support S de la densité de X_i et sur la suite a_n, mais aussi nécessitent une condition sur la densité de $Y_i \mid X_i$ qui est vérifiée dans le cas des MLG. Cette condition s'écrit, avec les notations du paragraphe 1.3 :

Etant donné un intervalle fermé F de $M \subset \boldsymbol{R}$, $\exists c \geqslant 0$, $\forall t_1$, $\forall t_2$ éléments de $F, \forall y \in I \subset \boldsymbol{R}$.

$$\mid f^*(y,t_1) - f^*(y,t_2) \mid = 0(\mid t_1 - t_2 \mid (\mid y \mid + c)) \tag{10}$$

En effet, le terme que l'on souhaite majorer et qui est écrit avec les notations du paragraphe 1, s'explicite encore :

$$f^*(y,t_1) - f^*(y,t_2) = y(A(t_1) - A(t_2)) - (B(t_1) - B(t_2))$$

En appliquant le théorème des accroissements finis, la condition (10) est vérifiée pour la constante $c = \sup\limits_{t \in F} \mid t \mid$ si l'on remarque que $A'(t) = tB'(t)$ et $\sup\limits_{t \in F} \mid A'(t) \mid < +\infty$.

b) D'autre part, les propriétés de convergence en loi et d'efficacité asymptotique semi paramétrique se déduisent des théorèmes de convergence dans le cas des M-estimateurs.

Sous certaines hypothèses (*cf.* Bonneu, Delecroix, Hristache 1994), on obtient le résultat :

$$\sqrt{n}(\hat{\beta}_{MPV} - \alpha) \xrightarrow{\mathcal{L}} \mathcal{N}_p(0, \Sigma_\alpha^{-1}) \quad \text{en loi}$$

$$\text{où} \quad \Sigma_\alpha = E_{X_i\alpha}\left[(r'_\alpha(X_i\alpha))^2 \cdot \mathrm{Var}\,(X_i \mid X_i\alpha)/A'(r_\alpha(X_i\alpha))\right]$$

• **Résolution numérique de** $\hat{\beta}_{MPV}$

L'algorithme de Newton Raphson pour résoudre la maximisation de la pseudo vraisemblance définie en (8) est identique à l'algorithme itératif défini en (3), si l'on considère l'estimateur $\hat{r}_\beta(x_i\beta)$ de la fonction de lien au point $x_i\beta$ et un estimateur non paramétrique $\hat{r}'_\beta(x_i\beta)$ de la dérivée de la fonction de lien au point $x_i\beta$. C'est la solution proposée par Weisberg et Welsh (1992) qui considèrent pour valeur initiale de β, l'estimateur $\tilde{\beta}_{MV}$ du maximum de vraisemblance quand la fonction de lien est canonique. En fait l'estimateur proposé est plus complexe et sous certaines conditions, des propriétés de convergence et de $\sqrt{n}$-normalité sont obtenus (Weisberg, Welsh (1992)).

Par analogie à l'algorithme mis en œuvre dans GLIM (*cf.* paragraphe 1.3), la résolution numérique, de $\hat{\beta}_{MPV}$ peut être réalisée par l'algorithme des moindres carrés pondérés itératif relatif aux équations normales :

$${}^tX\hat{V}_\beta^{-1}(\hat{z}_\beta - X\beta) = 0$$

où $\hat{V}^{-1}(\beta)$ est la matrice diagonale de terme diagonal : $A'(\hat{r}_\beta(x_i\beta)) \times (\hat{r}'_\beta(x_i\beta))^2$ et le vecteur $\hat{z}_\beta$ de composante : $\hat{z}_i = x_i\beta + (y_i - \hat{r}_\beta(x_i\beta))/\hat{r}'_\beta(x_i\beta)$. Dans ce cas l'estimateur $\hat{r}'_\beta(x_i\beta)$ est obtenu par dérivation de la fonction $\hat{r}_\beta(x_i\beta)$ par rapport à β.

3.2 Estimation du lien dans les modèles semi paramétriques

Des approches non paramétriques, ont été proposées quand on ne tient plus compte d'information sur la densité conditionnelle de Y_i sachant X_i. C'est le cadre des modèles à direction révélatrice unique (single index models) définis par :

$$m(x) = E(Y_i \mid X_i = x) = r(x\beta) \tag{11}$$

où la fonction de lien r est supposée inconnue.

Une première approche purement non paramétrique est la technique ADE : «Average Derivative Estimation» (Powell, Stock, Stoker (1989), Härdle, Stoker (1989), Härdle (1992)), qui consiste à estimer $E(m'(X))$ à partir d'estimateur non paramétrique $\hat{g}$ (resp. $\hat{g}'$) de la densité g de X_i (resp. de la dérivée de g). Sous certaines conditions sur g, cet estimateur s'explicite :

$$\hat{\beta}_{ADE} = -\frac{1}{n}\sum_{i=1}^{n} Y_i \frac{\hat{g}'(X_i)}{\hat{g}(X_i)} \tag{12}$$

Les propriétés de convergence et de $\sqrt{n}$-normalité sont établis pour le vecteur aléatoire $\sqrt{n}[\hat{\beta}_{ADE} - E(m'(x))]$. Si l'on suppose que l'espace paramétrique $\mathcal{B}$ est contraint (*cf.* hypothèse H_2'), on identifie alors $E(m'(X))$ à la vraie valeur α du paramètre d'intérêt qui vérifie (11).

Une comparaison empirique des estimateurs $\hat{\beta}_{ADE}$ et $\hat{\beta}_{MPV}$ réalisée à partir de données simulées a mis en évidence les performances de $\hat{\beta}_{MPV}$ par rapport à $\hat{\beta}_{ADE}$ quand on connait la loi de Y_i/X_i (Bonneu, Delecroix, Malin (1993)).

Une deuxième approche consiste à déterminer l'estimateur des «moindres carrés»qui minimise.

$$\sum_{i=1}^{n}(Y_i - \hat{r}_\beta(X_i\beta))^2$$

C'est la démarche adoptée par Ichimura (1988), Härdle, Hall et Ichimura (1993). Cet estimateur coïncide avec l'estimateur du maximum de pseudo vraisemblance $\hat{\beta}_{MPV}$ quand on considère que la loi de Y_i/X_i est gaussienne de variance ξ.

D'autres approches consistent à estimer $m(x) = E(Y_i/X_i = x)$ de façon non paramétrique quand $m(x)$ ne vérifie plus (11).

Par exemple, dans le cas des modèles additifs généralisés (Hastié et Tibshirani (1986), (1987)), le paramètre β de dimension finie n'intervient plus dans $m(x)$ qui s'explicite :

$$m(x) = r_0\left(\sum_{j=1}^{p} g_j(x^j)\right)$$

où r_0 est une fonction donnée et les fonctions g_j sont inconnus. Le cas plus général des modèles de régression en «projection pursuit» (Friedman, Stuetzel (1981)), font intervenir un nombre inconnu K de direction révélatrice :

$$m(x) = \sum_{k=1}^{K} g_k(x\beta^{(k)}) \qquad (K < p)$$

Les K fonctions g_k réelles sont inconnues et le paramètre de dimension finie est dans ce cas un vecteur constitué par les p-vecteurs $\beta^{(k)}$. Les modèles additifs étudiés par Hardle, Tsybakov (1994) ne font plus intervenir de paramètre $\beta^{(k)}$ et supposent qu'il existe $\{j_1, \cdots, j_K\} \subset \{1, \cdots, p\}$ tel que :

$$m(x) = \sum_{k=1}^{K} g_{j_k}(x^{j_k}).$$

Dans toutes ces approches, l'objectif est de déterminer le sous ensemble de régresseurs minimal qui prédise le mieux $m(x)$ et non plus de quantifier les effets des variables explicatives X^j.

Une dernière approche est la technique SIR (Sliced Inverse Regression) proposée par K.C. Li (1991). Le modèle est dans ce cas défini par :

$$Y_i = r(X_i\beta, \varepsilon)$$

La loi de X_i est supposée elliptique. La fonction inconnue r est dans ce cas une fonction qui lie Y_i au couple constitué du prédicteur linéaire $X_i\beta$ et de l'erreur ε du modèle. Ce n'est plus le cadre des modèles linéaires généralisés.

3.3 Mise en œuvre pratique

On considère le modèle dichotomique défini au paragraphe 1.2. Dans le cas d'expériences indépendantes de Bernouilli, la fonction $Q_n(\beta)$ de pseuso vraisemblance définie en (8) s'écrit :

$$Q_n(\beta) = \sum_{i=1}^{n} \{y_i \operatorname{logit} \hat{r}_\beta(x_i\beta) + \ln(1 - \hat{r}_\beta(x_i\beta))\}$$

Pour calculer l'estimateur $\hat{\beta}_{PMV}$ qui maximise $Q_n(\beta)$, on considère l'estimateur à noyau $\hat{r}_\beta(x_i\beta)$ défini en (6) pour un noyau gaussien et une largeur de fenêtre proportionnelle à $n^{1/5}$. Ce choix est suffisant pour obtenir les convergences énoncés au paragraphe 3.

Les performances sont mesurées par l'estimation de l'erreur quadratique moyenne : $E \parallel \hat{\beta} - \alpha \parallel^2$. Pour étudier les performances de cet estimateur, des simulations du modèle dichotomique ont été réalisées pour différentes fonctions de lien r_α. L'estimateur $\hat{\beta}_{MPV}$ peut être alors comparé aux différents estimateurs :

estimateur MV dans le cas où on considère r_α; estimateur MV dans le cas où on spécifie incorrectement la fonction de lien; estimateur ADE défini en (12) pour lequel le choix de la fenêtre est crucial. (Härdle, Hall, Ichimura (1993), Bonneu, Delecroix, Malin (1993)). Dans le cas où la fonction de lien est correctement spécifié à r_α, l'estimateur $\hat{\beta}_{MPV}$ reste proche de l'estimateur MV. Dans le cas contraire, les performances de $\hat{\beta}_{MPV}$ sont meilleures, l'estimateur MV étant de plus biaisé (*cf.* figure 1, où les graphiques donnent un exemple de résultats, pour une composante du paramètre α, et pour 100 échantillons simulés de taille 500).

D'autres estimateurs semi ou non paramétriques sont définis dans le modèle dichotomique (Klein, Spady (1989), Manski (1985), Horowitz (1992). Une étude de comparaison par simulation (Malin (1992)) permet de mettre en évidence des performances pour $\hat{\beta}_{MPV}$ meilleures que celles de l'estimateur du score de Manski.

Il faut noter que l'on doit imputer les bonnes performances de l'estimateur $\hat{\beta}_{MPV}$ par rapport aux autres estimateurs semi ou non paramétriques, au fait que l'on dispose d'une information justifiée sur la distribution de $Y_i \mid X_i$.

On peut donc conclure que si la loi de $Y_i \mid X_i$ est correctement spécifiée, il est préférable de considérer un modèle linéaire généralisé semi paramétrique et d'estimer le paramètre d'intérêt α par l'estimateur $\hat{\beta}_{MPV}$. On en déduit alors une estimation définie en (9) de la fonction de lien qui n'était pas *a priori* spécifiée.

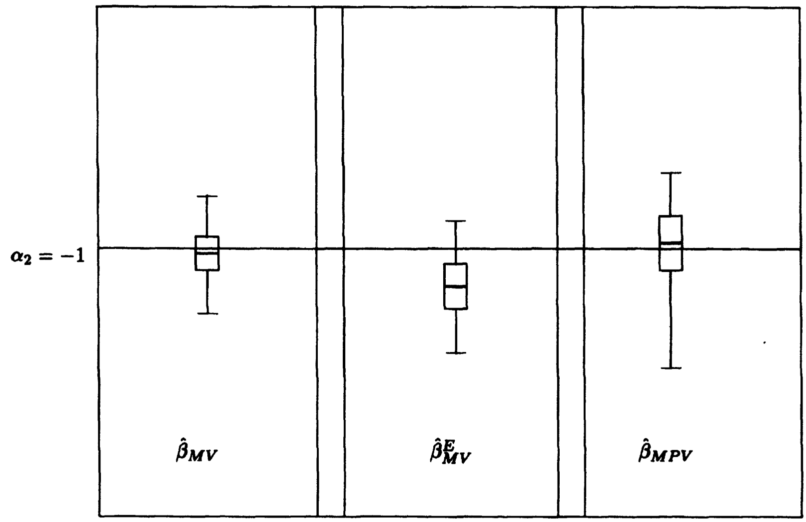

FIGURE 1 :

Boîtes à moustaches représentant les distributions empiriques des différents estimateurs de la composante $\alpha_2 = -1$ de la vraie valeur α_2 du paramètre d'intérêt.

$\hat{\beta}_{MV}$ = *estimateur du maximum de vraisemblance pour le lien correct* r_α
$\hat{\beta}^E_{MV}$ = *estimateur du MV pour un lien mal spécifié différent de* r_α
$\hat{\beta}_{MPV}$ = *estimateur du maximum de pseudo vraisemblance*

Bibliographie

ANDREWS, D.W.K. (1989), «Semi parametric Models : Estimation», *Manuscript*, Cowles Foundation, Yale University.

AITKIN, M., D. ANDERSON, B. FRANCIS AND J. HINDE (1990), «Statistical Modelling in *GLIM*», *Oxford Science Publications.*

ATKINSON, A.C. (1982), «Regression Diagnostics, Transformations and constructed Variables», *JRSS*, **B44**, p. 1-36.

BICKEL, P. (1982), «On Adaptive Estimation», *Ann. Stat.*, **10**, p. 647-671.

BONNEU, M. (1988), «Model Choice for Prediction in Generalized Linear Models», *Statistics*, **19,3**, p. 369-382.

BONNEU, M., M. DELECROIX (1993), «Convergence d'un estimateur semi paramétrique dans les modèles à direction révélatrice unique», GREMAQ n° 920 9256.

BONNEU, M., M. DELECROIX, M. HRISTACHE (1994), «Efficient Semiparametric Estimation in a Class of Single Index Models», CREST, n° 9447.

BONNEU, M., M. DELECROIX, E. MALIN (1993), «Semiparametric Versus Non Parametric Estimation in Single Index Regression Model : a Computational Approach», *Computational Statistics*,**8**, p. 207-222.

BONNEU, M., X. MILHAUD (1993), «A Modified Akaike Criterion for Model Choice in Generalized Linear Models», *Statistics*, (à paraître).

CZADO, C. (1992), «On Link Selection in Generalized Linear Models», *Lecture Notes in Statistics*, **78**, p. 60-65.

FAHRMEIR, L. AND H. KAUFMANN (1985), «Consistency and Asymptotic Normality of the Maximum Likelihood Estimator in Generalized Linear Models», *Ann. Stat.*, **13**, p. 342-368.

FAHRMEIR, L (1987), «Asymptotic Testing Theory for Generalized Linear Models», *Statistics*, **18,1**, p. 65-76.

FAHRMEIR, L (1990), «Maximum Likelihood Estimation in Misspecified Generalized Linear Models», *Statistics*, **21,4**, p. 487-502.

FRIEDMAN, J.H., W. STUETZEL (1981), «Projection Pursuit Regression», *JASA*, **76**, p. 817-823.

GALLANT, A.R., D.W. NYCHKA (1987), «Semi Parametric Maximum Likelihood Estimation», *Econometrica*, **55,2**, p. 363-390.

GONG, G., F.J. SAMANIENGO (1981), «Pseudo Maximum Likelihood Estimation : Theory and Applications», *Ann. Stat.*, **9,4**, p. 861-869.

GREEN, P.J., B.S. YANDELL (1985), «Semi Parametric Generalized Linear Models», *Lecture Notes in Statistics*,**32**, p. 44-55, Springer.

HÄRDLE, W. (1992), «Non Parametric Approaches to Generalized Linear Models» *Lectures Notes in Statistics*, **78**, p. 213-225.

HÄRDLE, W., T. STOKER (1989), «Investigating Smooth Multiple Regression by the Method of Average Derivatives», *JASA*, **84,408**, p. 986-995.

HÄRDLE, W., A.B. TSYBAKOV (1994), «How Many Terms Should Be Added into a Additive Model».

HÄRDLE, W., P. HALL AND H. ICHIMURA (1993), «Optimal Smoothing in Single Index Models», *Ann. Stat.*, **21**, 157-178.

HASTIE, T., R. TIBSHIRANI (1986), «Generalized Additive Models» *Stat. Science*, **1**, p. 297-318.

HASTIE, T., R. TIBSHIRANI (1987), «Generalized Additive Models : Somes Applications», *JASA*, **82**, p. 371-386.

HOROWITZ, J.L., (1992), «A Smoothed Maximum Score Estimator for the Binary Response Model», *Econometrica*, **60,3**, p. 505-531.

ICHIMURA, H. (1988), «Semi Parametric Least Squares Estimation of Single Index Models», Technical Report, University Minnesota.

ICHIMURA, H. AND L.F. LEE (1991), «Semi Parametric Least Squares Estimation of Multiple Index Models : Single Equation Estimation», Proceeding of the fifth Economic Theory and Econometrics, p. 3-49.

KLEIN, R.L. AND R.H. SPADY (1989), «An Efficient Semi Parametric Estimator for Discrete Choice Models», Manuscript, Bell Communication Research Corp.

LINHART H., ZUCCHINI W. (1986), «Model selection», J. Wiley and sons, New York.

LI, K.C. AND N. DUAN (1989), «Regression Analysis under Link Violation», *Ann. Stat.*, **17**, p. 1009-1052.

LI, K.C. (1991), «Sliced Inverse Regression : Dimension Reduction Via Inverse Regression», *JASA*, **86**, p. 316-342.

LINDSEY, J.K. (1989), «The Analysis of Categorical Data Using GLIM», *Lecture Notes in Statistics*, **56**, Berlin-Springer

MALIN, E. (1992), «Modèles à Revenu-Seuil de Demande Téléphonique. Comparaison par Simulation d'Estimateurs Semi et Non Parametriques», Cahier GREMAQ N°92.19.266.

MANSKI, C.F. (1985), «Semi Parametric Analysis of Discrete Response», *J. Econometrics*, **27**, p. 313-333.

MATHIEU, J.R. (1981), «Tests of χ^2 in the Generalized Linear Model», *Math. Op. Statist.*, **12**, p. 509-527.

MC CULLAGH, P. (1983), «Quasi-Likelihood Functions», *Ann. Stat.*, **11**, p. 59-67.

MC CULLAGH, P. (1984), «Generalized Linear Models», *European Journal of Operational Research*, **16**, p. 285-292, North-Holland.

MC CULLAGH, P. AND J.A. NELDER (1983), «Generalized Linear Models», London New York Chapman and Hall.

NELDER, J.A. AND R.W.M. WEDDERBURN (1972), «Generalized Linear Models», *JRSS*, **A147**, p. 389-425.

POWELL, J.L., J.M. STOCK AND T.M. STOCKER (1989), «Semi Parametric Estimation of Index Coefficients», *Econometrica*, **57**, p. 1403-1430.

PREGIBON, D. (1981), «Logistic Regression Diagnostics», *Ann. Statist.*, **9**, p. 705-724.

STUKEL, T. (1988), «Generalized Logistic Models», *JASA*, **83**, p. 426-431.

VUONG, Q. (1989), «Likelihood Ratio Tests for Model Selection and Non Nested Hypotheses», *Econometrica*, **57**, p. 257-306.

WEISBERG, S. AND A.M. WELSH (1992) «Adapting for the Missing Link», Preprint, Australian National Univ. Camberra.