L'Actualité économique

Modèles de comptage semi-paramétriques

Christian Gouriéroux et Alain Monfort

Volume 73, numéro 1-2-3, mars–juin–septembre 1997

L'économétrie appliquée

URI : https://id.erudit.org/iderudit/602238ar
DOI : https://doi.org/10.7202/602238ar

Aller au sommaire du numéro

Éditeur(s)

HEC Montréal

ISSN

0001-771X (imprimé)
1710-3991 (numérique)

Citer cet article

Gouriéroux, C. & Monfort, A. (1997). Modèles de comptage semi-paramétriques. *L'Actualité économique*, *73*(1-2-3), 525–550. https://doi.org/10.7202/602238ar

Résumé de l'article

Dans cet article nous définissons une nouvelle classe de modèles pour les variables endogènes entières : les modèles additifs log-différenciés en probabilité (ALDP). Cette classe a des analogies avec les modèles semi-paramétriques de hasard proportionnel pour les modèles de durées et a des interprétations intéressantes en terme de coûts (ou de bénéfices). Les propriétés asymptotiques des estimateurs du maximum de vraisemblance pour ces modèles sont étudiées et comparées à celles des estimateurs de l'analyse discriminante. On propose également des estimateurs adaptés au cas d'échantillons stratifiés de façon exogène ou endogène. Enfin, le cas d'observations hétérogènes est discuté.

L'Actualité économique, Revue d'analyse économique, vol. 73, nos 1-2-3, mars-juin-septembre 1997

MODÈLES DE COMPTAGE SEMI-PARAMÉTRIQUES

Christian GOURIÉROUX
CREST-CEPREMAP
Alain MONFORT
CREST

RÉSUMÉ – Dans cet article nous définissons une nouvelle classe de modèles pour les variables endogènes entières : les modèles additifs log-différenciés en probabilité (ALDP). Cette classe a des analogies avec les modèles semi-paramétriques de hasard proportionnel pour les modèles de durées et a des interprétations intéressantes en terme de coûts (ou de bénéfices). Les propriétés asymptotiques des estimateurs du maximum de vraisemblance pour ces modèles sont étudiées et comparées à celles des estimateurs de l'analyse discriminante. On propose également des estimateurs adaptés au cas d'échantillons stratifiés de façon exogène ou endogène. Enfin, le cas d'observations hétérogènes est discuté.

ABSTRACT – In this paper, we introduce a new class of models for count endogenous variables, i.e. the additive log-differentiated probability models (ALDP). This class is similar to the semi-parametric proportional hazard models used for duration data, and has some interesting implications in terms of costs or benefits. The asymptotic properties of the maximum likelihood estimators are studied and compared with the properties of the discriminant analysis estimators. We also explain why these models are suitable in the framework of endogenous sampling. Finally we discuss the introduction of heterogeneity.

INTRODUCTION

Les modèles économétriques sur données entières ont été introduits pour décrire des données de durée en temps discret et analyser des données de comptage.

Les modèles de durée ont été étudiés dans de nombreux articles, par exemple Kalbfleisch-Prentice (1980), Cox-Oakes (1984), Heckman-Singer (1984), Gouriéroux (1989), Lancaster (1990). Ils sont habituellement spécifiés en temps continu, ce qui implique une distribution continue des durées, mais admettent une contrepartie en temps discret. Ces modèles sont généralement fondés sur une spécification de la fonction de hasard :

$$\lambda(k) = P[\tau = k / \tau \geq k],\ k = 0, 1, 2, ...$$

où τ est la durée.

Il y a une relation biunivoque entre cette fonction de hasard et la loi de la durée τ, puisque :

$$p(k) = P[\tau = k] = \lambda(k)\prod_{j=0}^{k-1}[1-\lambda(j)], k = 0,1,2,...,$$

$$\lambda(k) = p(k) / \sum_{\ell \geq k} p(\ell).$$

Quand on veut étudier l'effet de variables exogènes x sur la durée τ, il est utile de séparer cet effet de la forme générale de la fonction de hasard. Par exemple, on peut introduire un modèle semi-paramétrique satisfaisant l'hypothèse de hasard proportionnel (voir Cox, 1975) : la fonction de hasard λ est le produit d'une fonction de hasard de base λ_0 par une fonction paramétrique des variables exogènes. Quand cette fonction paramétrique a une forme exponentielle, on obtient :

$$\lambda(k) = \lambda_0(k)\ exp(x\theta),$$

ou

$$Log\ \lambda(k) = Log\ \lambda_0(k) + x\theta.$$

Ce modèle a un aspect paramétrique, par la présence du paramètre θ dans la fonction des variables exogènes, et non paramétrique puisque la fonction de hasard de base est en général non contrainte. Dans la modélisation des données de durée, ce modèle à hasard proportionnel est souvent utilisé, parce qu'il fournit généralement un bon ajustement (en raison de son caractère semi-paramétrique), parce qu'il est facile à interpréter et à estimer (en particulier la détermination d'estimateurs du maximum de vraisemblance partiels est fondée sur des modèles logit multivariés), et parce qu'il se prête bien à un classement des individus par l'intermédiaire du score $x\theta$.

Les données de comptage apparaissent dans de nombreux problèmes pratiques. En recherche et développement, on s'intéresse au nombre de brevets déposés pendant une année (voir Hausman-Hall-Griliches, 1984 ; Hall-Griliches-Hausman, 1986); dans les études sur les crédits permanents, on analyse le nombre d'incidents de paiement, et dans les études actuarielles le nombre de sinistres pendant une période donnée (voir Lemaire, 1985 ; Boyer-Dionne, 1989 ; Dionne-Vanasse, 1992 ; Dionne-Gouriéroux-Vanasse, 1996)...

Les modèles de données de comptage sont souvent fondés sur le modèle de Poisson dans lesquels les probabilités élémentaires sont :

$$p(k) = exp(-exp\ (x\theta))\frac{exp(kx\theta)}{k!}, k \in \mathbb{N},$$

ou sur des extensions de ce modèle, comme les modèles binomiaux négatifs (Gouriéroux-Monfort-Trognon, 1984b ; Hausman-Hall-Griliches, 1984 ; Lemaire, 1985 ; Gouriéroux, 1989), introduits pour prendre en compte les problèmes d'hétérogénéité. Ces modèles sont paramétriques. Dans cet article, nous nous intéressons à l'introduction de modèles pour données de comptage, analogues des modèles semi-paramétriques à hasard proportionnel pour les données de durée.

Dans le modèle de Poisson, les probabilités log-différenciées.

$$Log\ p(k) - Log\ p(k-1) = Log\ \frac{1}{k} + x\theta,$$

sont somme d'une fonction de k et d'une fonction des variables exogènes ; il est donc naturel de fonder un modèle plus général sur la fonction log-différenciée en probabilité (LDP) :

$$\gamma(k) = Log\ p(k) - Log\ p(k-1),$$

qui aura le même rôle que la fonction de hasard dans les modèles de durée.

Dans la première section, on introduit le modèle additif log-différencié en probabilité (ALDP), défini par :

$$\gamma(k) = Log\ p(k) - Log\ p(k-1) = \gamma_0(k) + x\theta,$$

où γ_0 est une fonction LDP de base inconnue. On décrit quelques modèles compatibles avec cette forme ALDP et on fournit une interprétation du paramètre θ en terme de coût (ou de bénéfice). Ensuite, on examine les propriétés asymptotiques des estimateurs du maximum de vraisemblance de la fonction $[\gamma_0(k)]$ et du paramètre θ, et en particulier on fournit une expression simple de leur matrice de variance-covariance asymptotique. Ces résultats asymptotiques sont utilisés pour construire des procédures de test de certains sous-modèles paramétriques : modèle de Poisson, modèle géométrique.

Dans la deuxième partie, on relie le modèle ALDP à l'analyse discriminante. Après avoir rappelé le modèle sous-jacent à l'analyse discriminante, on remarque qu'il s'agit d'un sous-modèle du modèle ALDP. On donne alors les expressions des estimateurs de l'analyse discriminante de $[\gamma_0(k)]$ et θ, on étudie leur propriétés asymptotiques et on compare ces estimateurs aux estimateurs du maximum de vraisemblance analysés dans la deuxième partie.

La troisième section est consacrée à l'économétrie des modèles ALDP quand les données sont issues d'un échantillon stratifié de manière exogène ou endogène. En particulier on établit que, si l'échantillonnage est endogène, l'estimateur de θ fondé sur un échantillonnage exogène reste convergent et asymptotiquement normal. On propose également des estimateurs plus efficaces et facilement calculables.

Dans la quatrième partie, on considère le problème de l'hétérogénéité et on montre que les paramètres peuvent être estimés par la méthode du pseudo-maximum de vraisemblance et que le problème de la prévision des effets individuels peut être résolu d'une manière similaire. Les démonstrations des résultats sont regroupées dans des annexes.

1. LE MODÈLE ALDP

1.1 *Description du modèle*

Considérons une variable endogène Y, dont les valeurs possibles sont entières, notées $k = 0,1,...,K$ (où K est fini ou infini). À cette variable discrète, on associe un ensemble de variables binaires :

$$Z_k = \begin{cases} 1, \text{ si } Y = k, \\ 0, \text{ sinon.} \end{cases} \tag{1.1}$$

On s'intéresse à un modèle décrivant la loi conditionnelle de Y (et donc des Z_k, $k = 1,...,K$) sachant les valeurs de variables exogènes X.

Définition 1.2

Le modèle ALDP est tel que les probabilités conditionnelles :

$$P[Y = k/X = x] = p(k/x),$$

satisfont les relations :

$$Log\, p(k/x) - Log\, p(k-1/x) = \gamma_0(k) + x\theta,\ k = 1,...,K,$$

où les $\gamma_0(k)$ $k = 1,...,K$ sont des paramétres scalaires inconnus et θ un paramètre de taille p, p étant le nombre de variables exogènes. La fonction γ_0 est appelée la fonction LDP (log-différenciée en probabilité) de base[1].

Dans un modèle ALDP, les expressions des probabilités $p(k/x)$, peuvent être explicitées par substitutions répétées :

$$Log\ p(k/x) - Log\ p(0/x) = \sum_{\ell=1}^{k} \gamma_0(\ell) + k\ x\theta,\ k = 1,..., K,$$

$$\Leftrightarrow p(k/x) = p(0/x) exp\left(\sum_{\ell=1}^{k} \gamma_0(\ell) + k\ x\theta \right), k = 1,..., K.$$

1. Pour une variable continue, l'analogue des modèles ALDP est un modèle tel que :

$$\frac{d Log f(y/x)}{dy} = \gamma_0(y) + x\theta,$$

où $f(./x)$ est la densité conditionnelle de Y sachant $X = x$.

En tenant compte de la contrainte : $\sum_{k=0}^{K} p(k/x)$, on obtient :

$$p(k/x) = \frac{exp\left[\sum_{\ell=1}^{k} \gamma_0(\ell) + k\, x\theta\right]}{\sum_{k=0}^{K} exp\left[\sum_{\ell=1}^{k} \gamma_0(\ell) + k\, x\theta\right]}, \tag{1.3}$$

avec la convention : $\sum_{k=1}^{0} \gamma_0(\ell) = 0.$

L'expression de $p(k/x)$ n'a de signification que si la fonction LDP de base et les variables exogènes sont telles que :

$$\sum_{k=0}^{K} exp\left[\sum_{\ell=1}^{k} \gamma_0(\ell) + k\, x\theta\right] < +\infty,$$

(une inégalité qui peut ne pas être satisfaite si $K = +\infty$).

Dans le reste de l'article, on utilise également la fonction LDP de base cumulée :

$$\beta(k) = \sum_{\ell=1}^{k} \gamma_0(\ell), k = 0, 1, \ldots K, \text{ avec } \beta(0) = 0. \tag{1.4}$$

Avec cette notation les probabilités deviennent

$$p(k/x) = \frac{exp(\beta(k) + kx\theta)}{\sum_{\ell=0}^{K} exp(\beta(\ell) + \ell x\theta)}, k = 0, 1, \ldots K. \tag{1.5}$$

Le modèle ALDP contient comme cas particulier les modèles paramétriques classiques pour données de comptage.

i) Le modèle de Poisson correspond aux probabilités :

$$p(k/x) = exp(-exp\ (x\theta))\frac{1}{k!}exp(kx\theta), k = 0, 1, \ldots, \tag{1.6}$$

et la fonction LDP de base est l'opposée de la fonction logarithme :

$$\gamma_0(k) = Log\frac{p(k/x)}{p(k-1/x)} - x\theta = -Log\ k.$$

ii) Le modèle géométrique utilisé pour des données de durée en temps discret est obtenu avec :

$$p(k/x)=\frac{1}{1-\rho\ exp\ (x\theta)}\rho^{k}\ exp(kx\theta), k=0,1,... \qquad (1.7)$$

où $|\rho\ exp\ x\theta|<1$.

La fonction LDP de base associée est constante[2] :

$\gamma_0(k)=Log\ \rho$.

1.2 *Une interprétation en terme de coût (ou bénéfice)*

Dans la prévision des sinistres en assurance, on s'intéresse non seulement à leur nombre, mais aussi à leurs coûts. Dans la formulation standard utilisée par exemple en assurance automobile (voir Lemaire, 1985), le coût d'un accident, pour un individu donné, est indépendant du nombre d'accidents qu'il a dans la période. Le coût anticipé conditionnellement au nombre d'accident est supposé linéaire en Y : $C(Y) = C_0Y$, où C_0 est le coût moyen unitaire. Cette hypothèse sur la fonction de coût est restrictive et non compatible avec les données qui révèlent des corrélations entre occurrence et coût positives ou négatives selon les classes de risque.

Ici on adopte une hypothèse plus faible selon laquelle le coût anticipé conditionnel à Y est une fonction non décroissante de Y. On peut alors étudier le coût d'un assuré ayant plus de K_0 accidents, pour tout K_0. Une étude symétrique est possible si $C(Y)$ représente le bénéfice engendré par l'occurrence de Y, par exemple si Y est un nombre de brevets. On obtient le résultat suivant[3] :

Propriété 1.8

Si $C(Y)$ est une fonction de coût conditionnel (resp. bénéfice) non décroissante et si Y est une variable satisfaisant un modèle ALDP :

$$\frac{\partial}{\partial(x\theta)}E[C(Y)/Y\geq K_0]=Cov[(C(Y),Y)/Y\geq K_0],$$

et $E[C(Y)/Y\geq K_0]$ est une fonction non décroissante de l'indice $x\theta$.

Preuve : voir annexe 1.

2. Le modèle géométrique est caractérisé par la constance de la fonction de hasard de base, ou par la constance de la fonction LDP de base.

3. Deux lois ayant la même fonction γ_0 et différentes valeurs de θ peuvent être comparées par l'intermédiaire de l'ordre défini par :

$$P\geq P_0 \Leftrightarrow E_P(C(Y)/Y\geq K_0)\geq E_{P_0}(C(Y)/Y\geq K_0),$$

pour toute fonction C non décroissante et tout K_0.

Cet ordre a été étudié dans Fourgeaud-Gouriéroux-Pradel (1990), où il est comparé avec l'ordre fondé sur la comparaison des fonctions de hasard et avec la dominance stochastique d'ordre un.

Dans le modèle ALDP, l'indice $x\theta$ a l'interprétation d'un score, et permet de séparer les individus les plus risqués (resp. performants) des individus les moins risqués (resp. performants), quelque soit la fonction de coût non décroissante (bénéfice).

En particulier, si $C(Y)=C_0Y$ on obtient :

$$C_0 \frac{\partial EY}{\partial(x\theta)} = C_0 VY, \text{ ou : } \frac{\partial EY}{\partial(x\theta)} = VY.$$

L'effet marginal du score sur le nombre moyen de sinistres coïncide avec la variance.

1.3 *Estimation par la méthode du maximum de vraisemblance*

Afin d'éviter une complexité mathématique inutile, on suppose dans ce paragraphe que K est fini. Cette hypothèse implique que les deux sous-modèles paramétriques donnés en exemple, c'est-à-dire les modèles de Poisson et géométrique doivent être remplacés par les versions tronquées :

$$p(k/x) = \frac{(1/k!)\,exp(kx\theta)}{\sum_{\ell=0}^{K}(1/\ell!)\,exp(\ell x\theta)},$$

et $p(k/x) = \rho^k\, exp(kx\theta)\dfrac{1-\rho^{K+1}\, exp[(K+1)x\theta]}{1-\rho\, exp(x\theta)}$, respectivement. De plus, pour des raisons d'identification, on suppose que le score $x\theta$ ne contient pas de terme constant.

La fonction log-vraisemblance correspondant à n observations $i = 1,...,n$ est donnée par :

$$L(\beta;\theta) = \sum_{i=1}^{n}\sum_{k=0}^{K} Z_{ik} Log\left[\frac{exp[\beta(k)+kx_i\theta]}{\sum_{\ell=0}^{K} exp[\beta(\ell)+\ell x_i\theta]}\right]$$

$$= \sum_{i=1}^{n}\left\{\sum_{k=1}^{K} Z_{ik}[\beta(k)+kx_i\theta] - Log\left[\sum_{\ell=0}^{K} exp(\beta(\ell)+\ell x_i\theta)\right]\right\},$$

et coïncide avec la fonction log-vraisemblance d'un modèle logit multinomial, dans lequel les indices associés à chaque alternative sont contraints linéairement. Ces contraintes expliquent la forme simple de la matrice de variance-covariance asymptotique des estimateurs du maximum de vraisemblance.

En utilisant les notations : $Z_i = (Z_{i1},...,Z_{iK}), p_i = [p(1/x_i),...p(K/x_i)]'$, $\alpha = (1,...,K)'$, $\beta = [\beta(1),...,\beta(K)]'$, les dérivées premières de la log-vraisemblance s'écrivent :

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{n} (Z_i - p_i),$$

$$\frac{\partial L}{\partial \theta} = \sum_{i=1}^{n} x_i' \left\{ \sum_{k=1}^{K} k Z_{ik} - \frac{\sum_{\ell=0}^{K} \ell \, exp[\beta(\ell) + \ell x_i \theta]}{\sum_{\ell=0}^{K} exp[\beta(\ell) + \ell x_i \theta]} \right\}$$

$$= \sum_{i=1}^{n} x_i' \left(\sum_{k=1}^{K} k Z_{ik} - \sum_{\ell=0}^{K} \ell p(\ell / x_i) \right)$$

$$= \sum_{i=1}^{n} x_i' (Y_i - m_i),$$

où m_i est l'espérance conditionnelle de la variable discrète Y_i sachant $X_i = x_i$. Les deux fonctions p_i et m_i dépendent des paramètres β et θ et les estimateurs du maximum de vraisemblance sont obtenus en égalant les moyennes empiriques $\frac{1}{n}\sum_{i=1}^{n} Z_i$ et $\frac{1}{n}\sum_{i=1}^{n} x_i' Y_i$ à leurs contreparties théoriques :

$$\begin{cases} \dfrac{\partial L}{\partial \beta} = \sum_{i=1}^{n} [Z_i - p_i(\hat{\beta}, \hat{\theta})] = 0, \\ \dfrac{\partial L}{\partial \theta} = \sum_{i=1}^{n} x_i' [Y_i - m_i(\hat{\beta}, \hat{\theta})] = 0). \end{cases} \tag{1.9}$$

Puisque $Y_i = \alpha' Z_i$ et $m_i = \alpha' p_i$, ce système est un ensemble de conditions d'orthogonalité sur les résidus :

$$\hat{u}_i = Z_i - p_i(\hat{\beta}, \hat{\theta}),$$

et s'écrit :

$$\begin{cases} \sum_{i=1}^{n} \ {}_{\mathrm{i}} = 0, \\ \sum_{i=1}^{n} x_i' \alpha' \ {}_{i} = 0. \end{cases}$$

La matrice de variance-covariance asymptotique des estimateurs du maximum de vraisemblance est déduite des dérivées du deuxième ordre de la log-vraisemblance. Ces dérivées sont :

$$
\left[
\begin{aligned}
\frac{\partial^2 L}{\partial\beta\partial\beta'} &= -\sum_{i=1}^{n}\{diag(p_i)-p_i p_i'\} = -\sum_{i=1}^{n} V(Z_i),\\
\frac{\partial^2 L}{\partial\theta\partial\beta'} &= -\sum_{i=1}^{n} x_i' \left[\begin{bmatrix} 1\,p(1/x_i)\\ \vdots \\ \vdots \\ \vdots \\ K\,p(K/x_i)\end{bmatrix} - p_i m_i\right]'\\
&= -\sum_{i=1}^{n} x_i' Cov(Y_i, Z_i),\\
\frac{\partial^2 L}{\partial\theta\partial\theta'} &= -\sum_{i=1}^{n} x_i' x_i \left\{ \frac{\sum_{\ell=0}^{K} \ell^2 \exp(\beta(\ell)+\ell x_i\theta)}{\sum_{\ell=0}^{K} \exp(\beta(\ell)+\ell x_i\theta)} - \frac{\left[\sum_{\ell=0}^{K} \ell \exp(\beta(\ell)+\ell x_i\theta)\right]^2}{\left[\sum_{\ell=0}^{K} \exp(\beta(\ell)+\ell x_i\theta)\right]^2} \right\}\\
&= -\sum_{i=1}^{n} x_i' x_i V(Y_i).
\end{aligned}
\right.
$$

Ces dérivées sont non stochastiques (conditionnellement aux x_i) et la matrice d'information est :

$$
\begin{aligned}
I &= \sum_{i=1}^{n} \begin{bmatrix} V(Z_i) & Cov(Z_i, Y_i)x_i \\ x_i' Cov(Y_i, Z_i) & x_i' V(Y_i) x_i \end{bmatrix}\\
&= \sum_{i=1}^{n} \begin{bmatrix} V(Z_i) & V(Z_i)\alpha x_i \\ x_i'\alpha' V(Z_i) & x_i'\alpha' V(Z_i)\alpha x_i \end{bmatrix}.
\end{aligned}
$$

En particulier, la matrice de variance-covariance asymptotique de $\hat{\theta}$ peut être estimée par :

$$
\hat{V}\hat{\theta} = \left\{ \sum_{i=1}^{n} x_i'\alpha'\hat{\Omega}_i \alpha x_i - \left(\sum_{i=1}^{n} x_i'\alpha'\hat{\Omega}_i\right)\left(\sum_{i=1}^{n}\hat{\Omega}_i\right)^{-1}\left(\sum_{i=1}^{n}\hat{\Omega}_i\alpha x_i\right)\right\}^{-1}, \tag{1.10}
$$

où $\hat{\Omega}_i$ est déduite de $V(Z_i)$ en remplaçant les probabilités p_i par $\hat{p}_i = p_i(\hat{\beta},\hat{\theta})$.

1.4 *Tests sur la fonction LDP de base*

À partir de l'estimateur du maximum de vraisemblance $\hat{\beta}$ de β et de sa matrice de variance-covariance estimée $\hat{V}(\hat{\beta})$, on peut tester des formes particulières de la fonction LDP de base ou, de façon équivalente, de la fonction β. Par exemple on peut examiner l'hypothèse de Poisson :

$$H_0^P = \{\beta_k = -Log(k!),\ k = 1,...,K\}, \tag{1.11}$$

ou l'hypothèse géométrique :

$$H_0^G = \{\exists\rho : \beta_k = k\ Log\ \rho,\ k = 1,...,K\}. \tag{1.12}$$

L'hypothèse de Poisson peut être testée directement par un test de Wald. La statistique de Wald est :

$$\xi^P = [\hat{\beta}' + (Log1,...,LogK!)](\hat{V}\hat{\beta})^{-1}[\hat{\beta} + (Log1,...LogK!)'], \tag{1.13}$$

et l'hypothèse nulle H_0^P est acceptée si :

$$\xi^P < \chi^2_{95\%}(K),$$

rejetée sinon.

L'hypothèse géométrique est sous forme mixte. La statistique de test de Wald généralisé (voir Szroeter, 1983 ; Gouriéroux-Monfort, 1989b) :

$$\xi^G = Min_{Log\rho}(\hat{\beta} - \alpha Log\rho)'(\hat{V}\hat{\beta})^{-1}(\hat{\beta} - \alpha Log\rho), \tag{1.14}$$

est asymptotiquement distribuée selon une loi du khi-deux $\chi^2(K-1)$. Le test consiste :

$$\begin{cases} \text{à accepter l'hypothèse géométrique, si } \xi^G < \chi^2_{95\%}(K-1), \\ \text{à la rejeter, sinon.} \end{cases}$$

Il est aussi possible de proposer des procédures descriptives. Par exemple, on peut reporter graphiquement les valeurs $\hat{\beta}_k + Log(k!), k = 1,...,K$ avec leur bande de confiance, et examiner si elles sont proches de l'axe des abscisses, afin de voir si l'hypothèse de Poisson est approximativement satisfaite.

De façon analogue, on pourrait reporter les points $(k, \hat{\beta}_k)$ $k = 1,...,K$ avec leur bande de confiance et étudier s'ils sont approximativement sur une droite passant par l'origine, de façon à étudier l'hypothèse géométrique.

2. LIENS AVEC L'ANALYSE DISCRIMINANTE

2.1 *Deux approches*

Le modèle ALDP impose la forme de la loi conditionnelle des variables endogènes Y_i sachant les variables exogènes X_i. Il peut être complété en supposant que les variables exogènes sont indépendantes identiquement distribuées de densité f. On obtient un modèle pour (X_i, Y_i) de densité :

$$\ell(x_i, y_i) = f(x_i)p(y_i/x_i). \tag{2.1}$$

On a vu que, dans le modèle ALDP, la comparaison entre les individus peut être effectuée par la fonction de score $x_i\theta$, qui permet ainsi de discriminer entre les individus à partir de leurs caractéristiques.

Cette sélection des individus peut aussi être effectuée en utilisant d'autres méthodes comme l'analyse discriminante (AD). L'approche AD est symétrique de la précédente. Dans cette approche un modèle paramétrique est imposé pour la loi conditionnelle des variables exogènes X_i sachant les endogènes, la loi marginale des endogènes n'étant pas spécifiée. En général les variables exogènes sont supposées conditionnellement normales de moyenne conditionnelle linéaire en Y_i :

$$E(X_i/Y_i) = b + aY_i, \tag{2.2}$$

(où a et b sont des vecteurs de la taille p de X_i), et de variance conditionnelle constante :

$$V(X_i'/Y_i) = \Sigma. \tag{2.3}$$

Sous cette hypothèse le modèle AD est caractérisé par la famille de densités :

$$\begin{aligned} \ell_0(x_i;y_i) = {} & \frac{1}{(2\pi)^{p/2}}\frac{1}{(det\,\Sigma)^{1/2}} \\ & exp\left[-\frac{1}{2}(X_i'-b-aY_i)'\Sigma^{-1}(X_i'-b-aY_i)\right].p_0(y_i), \end{aligned} \tag{2.4}$$

où $(p_0(y_i))$ est la densité marginale de Y_i.

Comme pour les modèles logit et AD (voir Efron, 1975 ; McFadden, 1976 ; Amemiya-Powell, 1983 ; Lo, 1984, 1986), le modèle AD est emboîté dans le modèle ALDP.

Propriété 2.5

Le modèle d'analyse discriminante est un modèle ALDP.

Preuve : De la densité jointe (2.4), on déduit les probabilités conditionnelles élémentaires de Y_i sachant X_i

$$\begin{aligned} p_0(k/x_i) &= \frac{\ell_0(x_i;k)}{\sum_{\ell=0}^{K}\ell_0(x_i;\ell)} \\ &= \frac{exp\left[-\frac{1}{2}(x_i'-b-ak)'\Sigma^{-1}(x_i'-b-ak)\right]p_0(k)}{\sum_{\ell=0}^{K} exp\left[-\frac{1}{2}(x_i'-b-a\ell)'\Sigma^{-1}(x_i'-b-a\ell)\right]p_0(\ell)} \\ &= \frac{exp(x_i\Sigma^{-1}ak)\,exp\left[-b'\Sigma^{-1}ak-\frac{k^2}{2}a'\Sigma^{-1}a+Log\ p_0(k)\right]}{\sum_{\ell=0}^{K} exp(x_i\Sigma^{-1}a\ell)\,exp\left[-b'\Sigma^{-1}a\ell-\frac{\ell^2}{2}a'\Sigma^{-1}a+Log\ p_0(\ell)\right]} \end{aligned}$$

C'est un modèle ALDP, où les paramètres β et θ sont :

$$\begin{cases} \beta(k) = -b'\Sigma^{-1}ak - \dfrac{k^2}{2}a'\Sigma^{-1}a + Log\dfrac{p_0(k)}{p_0(0)}, \\ \theta = \Sigma^{-1}a. \end{cases}$$

Q.E.D.

En fait, il est facile de déterminer tous les modèles conditionnels $f(x/y)$ pour lesquels les lois conditionnelles de Y sachant X satisfont un modèle ALDP. Ces densités conditionnelles doivent être telles que :

$$\frac{f(x/k)p_0(k)}{f(x/0)p_0(0)} = exp[\beta(k) + kx\theta]$$

$$\Leftrightarrow f(x/k) = \left[\frac{p_0(0)}{p_0(k)} exp\,\beta(k)\right] f(x/0) \quad exp(kx\theta).$$

La densité conditionnelle $f(x/y)$ est donc de la forme :

$$f(x/y) = a_0(y)a_1(x) \quad exp(yx\theta).$$

Naturellement les densités normales satisfont cette condition.

2.2 *L'estimateur de l'analyse discriminante*

Dans le modèle d'analyse discriminante les estimateurs du maximum de vraisemblance de $a,b,\Sigma,p_0(k),k = 0,...,K$ sont facilement calculés. Les estimateurs de a,b,Σ sont les estimateurs des moindres carrés dans le modèle de régressions empilées :

$$X_i' = b + aY_i + u, \; i = 1,...n, \; Eu = 0, \; Vu = \Sigma \otimes Id_n. \tag{2.6}$$

Comme les variables explicatives, 1 et Y_i, sont les mêmes dans les différentes équations, l'estimateur des moindres carrés généralisés de (a,b) coïncide avec celui des moindres carrés ordinaires.

En outre les estimateurs du maximum de vraisemblance des probabilités marginales $p_0(k)$ $k = 0,...,K$ sont simplement les fréquences empiriques $\hat{p}_0(k) = \frac{1}{n}\sum_{i=1}^{n} 1\!\mathrm{I}_{Y_i=k}$. On en déduit que l'estimateur du maximum de vraisemblance du paramètre d'intérêt θ dans le modèle AD est :

$$\tilde{\theta} = \tilde{\Sigma}^{-1}\tilde{a}, \tag{2.7}$$

où : $\tilde{a}_j = Cov_e(X'_j, Y)/V_e(Y)$,

$$\tilde{\Sigma} = [\tilde{\sigma}_{j\ell}], \text{ avec } \tilde{\sigma}_{j\ell} = Cov_e(X'_j, X'_\ell) - \frac{Cov_e(X'_j, Y)Cov_e(X'_\ell, Y)}{V_e(Y)},$$

et V_e et Cov_e désignent les variances et covariances empiriques.

Des propriétés des estimateurs de régressions empilées, on déduit (voir annexe 2), la matrice de variance-covariance asymptotique de $\sqrt{n}(\tilde{\theta} - \theta)$ dans le modèle AD :

$$\begin{aligned} V_{as}\left[\sqrt{n}(\tilde{\theta} - \theta)\right] &= (VY)^{-1}\Sigma^{-1} + 2(a' \otimes Id)(\Sigma^{-1} \otimes \Sigma^{-1}) \\ &\times R'Q(\Sigma \otimes \Sigma)Q'R(\Sigma^{-1} \otimes \Sigma^{-1})(a \otimes Id), \end{aligned} \tag{2.8}$$

où R est une matrice de sélection bien choisie et Q est l'inverse de Penrose-Moore de R.

2.3 *Comparaison des estimateurs AD et ALDP de* θ

Quand le modèle d'analyse discriminante est satisfait, les deux estimateurs $\hat{\theta}$ et $\tilde{\theta}$ sont convergents, mais l'estimateur AD, à savoir $\tilde{\theta}$, est asymptotiquement efficace. Quand le modèle AD n'est pas satisfait, et que le modèle ALDP l'est, $\hat{\theta}$ reste convergent, mais $\tilde{\theta}$ en général ne l'est plus.

Comme dans (Lo, 1984, 1986), on peut construire un test de spécification d'Hausman de l'hypothèse d'analyse discriminante en utilisant la statistique :

$$\xi = (\hat{\theta} - \tilde{\theta})'(\widehat{V_{as}\hat{\theta} - V_{as}\tilde{\theta}})^{-}(\hat{\theta} - \tilde{\theta}), \tag{2.9}$$

où $(\widehat{V_{as}\hat{\theta} - V_{as}\tilde{\theta}})^{-}$ est un estimateur convergent de l'inverse généralisée de $(V_{as}\hat{\theta} - V_{as}\tilde{\theta})$ Sous l'hypothèse AD, cette statistique est asymptotiquement distribuée selon une loi du khi-deux, dont le nombre de degrés de liberté est : $r_0 = \text{rang } (V_{as}\hat{\theta} - V_{as}\tilde{\theta})$. Donc

$$\begin{cases} \text{on rejette l'hypothèse d'analyse discriminante, si } \xi > \chi^2_{95\%}(r_0), \\ \text{on ne la rejette pas, sinon.} \end{cases}$$

3. ÉCHANTILLONS STRATIFIÉS

3.1 *Échantillonnage exogène*

Considérons un modèle ALDP dans lequel l'ensemble des valeurs possibles des variables exogènes est fini, $\{x_j, j = 1,...,J\}$. On suppose que, pour toute strate $X = x_j$, on tire un échantillon aléatoire de taille $n_j = \alpha_j n$, où les $\alpha_j, j = 1,...,J$ sont des nombres positifs fixés tels que : $\sum_{j=1}^{J} \alpha_j = 1$.

Les paramètres $\beta_k, k = 1,...,K$ et θ peuvent être estimés efficacement par la méthode du maximum de vraisemblance, en maximisant :

$$L(\beta, \theta) = \sum_{j=1}^{J} \sum_{k=0}^{K} n_{jk} \quad Log \frac{exp[\beta(k) + kx_j\theta]}{\sum_{\ell=0}^{K} exp[\beta(\ell) + \ell x_j \theta]} \quad [\text{avec } \beta(0) = 0]$$

où n_{jk} est le nombre d'observations pour lesquelle $Y = k$ et $X = x_j$. Elles s'appliquent lorsque les tailles $n_j, j = 1,...,J$ des sous-échantillons sont grandes. Ces conditions sont réalisées pour les applications au domaine de l'assurance automobile par exemple, où le nombre d'observations est couramment de l'ordre de 200-300 000 permettant la construction de quelques centaines de classes de risques *a priori* $j = 1,...,J$. Tous les résultats du paragraphe 1.3 s'appliquent.

Cependant, il existe d'autres procédures asymptotiquement efficaces, plus simples numériquement, fondées sur la méthode des moindres carrés asymptotiques, semblables aux méthodes de Berkson développées pour les modèles qualitatifs dichotomiques.

La méthode repose sur les estimateurs du maximum de vraisemblance non contraints des probabilités : $p(k/j) = p(Y = k/X = x_j)$:

$$\hat{p}(k/j) = \frac{n_{jk}}{n_j} .$$

Il existe une transformation biunivoque entre les paramètres

$$\{p(k/j), k = 0,..., K; j = 1,..., J\} \text{ et } \left\{Log \frac{p(k/j)}{p(k-1/j)}, k = 1,..., K; j = 1,..., J\right\},$$

et une procédure asymptotiquement efficace pour estimer $\gamma_0(k), k = 1,...,K$ et θ peut être fondée sur le modèle linéaire :

$$Log \frac{\hat{p}(k/j)}{\hat{p}(k-1/j)} = \gamma_0(k) + x_j\theta + u_{jk}, \quad j = 1,..., J \quad k = 1,..., K.$$

Cette procédure se réduit à une procédure des moindres carrés quasi généralisés associée à la matrice $\frac{\hat{\Omega}}{n}$, où $\hat{\Omega}$ est un estimateur convergent de $V_{as}\left\{\sqrt{n}[u_{jk}]\right\}$, $[u_{jk}]$ étant le vecteur de taille JK dont les composantes sont u_{jk}, k variant le premier. Une version de $\hat{\Omega}$ est la matrice bloc-diagonale dont les blocs sont :

$$n\begin{bmatrix} \left(\frac{1}{n_{j1}}+\frac{1}{n_{j0}}\right) & -\frac{1}{n_{j1}} & 0 & 0 & 0 & 0 \\ -\frac{1}{n_{j1}} & \left(\frac{1}{n_{j2}}+\frac{1}{n_{j1}}\right) & -\frac{1}{n_{j2}} & 0 & 0 & 0 \\ 0 & -\frac{1}{n_{j2}} & \left(\frac{1}{n_{j3}}+\frac{1}{n_{j2}}\right) & -\frac{1}{n_{j3}} & & -\frac{1}{n_{j,K-1}} \\ \vdots & & & & & \\ \vdots & & & & & \\ 0 & & & & -\frac{1}{n_{j,K-1}} & \left(\frac{1}{n_{jK}}+\frac{1}{n_{j,K-1}}\right) \end{bmatrix}$$

Une autre méthode asymptotiquement efficace, plus simple à mettre en oeuvre, est fondée sur le modèle :

$$Log\frac{\hat{p}(k/j)}{\hat{p}(0/j)} = Log\frac{n_{jk}}{n_{0k}} = \beta(k)+kx_j\theta+v_{jk}, \quad j=1,\dots,J, \quad k=1,\dots,K,$$

avec une matrice de variance-covariance des termes d'erreurs égale à $\frac{\Omega^*}{n}$, où Ω^* est la matrice bloc-diagonale de blocs :

$$\Omega_j^* = n\left[diag\left(\frac{1}{n_{j.}}\right)+\frac{1}{n_{j0}}U\right],$$

$diag\left(\frac{1}{n_{j.}}\right)$ est la matrice $(K\times K)$ diagonale de termes diagonaux $1/n_{jk}$ et U est la matrice $(K\times K)$ dont les éléments sont égaux à 1.

L'inverse Ω_j^{*-1} est égale à : $\frac{1}{n}\left[diag(n_{j.})-\frac{1}{n_j}[n_{j.}][n_{j.}]'\right]$, où $[n_{j.}]$ est le vecteur de taille K dont les composantes sont n_{jk}, $k = 1,\dots,K$. Cette procédure est particulièrement simple, puisque la fonction objectif à minimiser est :

$$\sum_{j=1}^{J} v_{j.}'\left[diag(n_{.j})-\frac{1}{n_j}[n_{j.}][n_{j.}]'\right]v_{j.}$$

où $v_{j.}$ est le vecteur dont les composantes sont $v_{jk}, k = 1,...,K$. Toutes les procédures de test classiques peuvent être utilisées, y compris celle fondée sur la différence des valeurs optimales de la fonction objectif (voir Gouriéroux-Monfort, 1989a : chapitre 18).

3.2 *Échantillonnage endogène*

On suppose maintenant que, pour chaque valeur k de Y, on tire un échantillon de taille $n_k = \alpha_k n$, où $\alpha_k, k = 0,...,K$, sont des nombres positifs tels que :

$$\sum_{k=0}^{K} \alpha_k = 1.$$

De tels tirages endogènes sont classiques dans les applications à l'assurance. Pour éviter le traitement d'un trop grand nombre de contrats, on trie souvent dans le fichier initial un échantillon de taille raisonnable de l'ordre de 4-5 000 contrats. Mais si le tirage est effectué au hasard, on risque de n'avoir dans cet échantillon que peu de contrats avec sinistres. On préfère alors tirer des nombres de contrats prédéfinis avec 0,1,2... sinistres, ce qui revient à surpondérer les contrats risqués. Pour estimer les paramètres du modèle ALDP dans le cas de sélection endogènes, nous distinguons le cas où l'ensemble des valeurs possibles de X est infini et celui où cet ensemble est fini.

3.2.1 *Ensemble infini de valeurs*

Notons :

$$\begin{aligned} p(k/x;\beta,\theta) &= \frac{exp[\beta(k)+kx\theta]}{\sum_{\ell=0}^{K} exp[\beta(\ell)+\ell x\theta]} \\ &= P[Y=k/X=x], \\ p(k) &= P[Y=k], \end{aligned}$$

et $\pi(x)$ la densité marginale de X (par rapport à une mesure μ).

Pour un tel tirage endogène la méthode du maximum de vraisemblance est impraticable. Cependant si la loi marginale $p(k)$ est connue, on peut appliquer une méthode du maximum de vraisemblance pondérée (Manski-Lerman, 1977). La vraisemblance pondérée est :

$$L_n^* = \sum_{k=0}^{K} \frac{p(k)}{\alpha_k} \sum_{i=1}^{n_k} Log\ p(k/x_i^k;\beta,\theta),$$

où $x_i^k (i = 1,...,n_k)$ est la valeur de la variable exogène pour le ième individu du kième échantillon. Cette procédure est convergente ; en effet :

$$\begin{aligned}\lim_{n\to\infty} \frac{L_n^*}{n} &= \lim_{n\to\infty} \sum_{k=0}^{K} \frac{p(k)}{n_k} \sum_{i=1}^{n_k} Log\ p(k/x_i^k;\beta,\theta) \\ &= \sum_{k=0}^{K} p(k) E_0 [Log\ p(k/X;\beta,\theta)/Y=k] \\ &= E_0\ Log\ p(Y/X;\beta,\theta),\end{aligned}$$

où l'indice 0 indique que l'espérance est par rapport à la vraie loi du couple (Y,X).

Cette limite est égale à :

$$\int \pi(x) \left[\sum_{k=0}^{K} Log\ p(k/x;\beta,\theta) p(k/x;\beta_0,\theta_0) \right] \mu(dx).$$

Sous les conditions usuelles d'identification, le maximum de cette fonction est atteint pour la valeur unique $\beta = \beta_0, \theta = \theta_0$, d'où la convergence. La normalité asymptotique est une conséquence de la théorie des M-estimateurs (voir Gouriéroux-Monfort, 1989a : chapitre 8), et la matrice de variance-covariance asymptotique de $\sqrt{n}(\hat{\varphi}_n - \varphi_0)$, où $\hat{\varphi}_n$ est l'estimateur de $\varphi = (\beta',\theta')'$, est estimée de façon convergente par :

$$n \left[\frac{\partial^2 L_n^*(\hat{\varphi}_n)}{\partial\varphi\partial\varphi'} \right]' \sum_{k=0}^{K} \sum_{i=1}^{n_k} \frac{p^2(k)}{\alpha^2(k)} \frac{\partial Log\ p(k/x_i^k;\hat{\varphi}_n)}{\partial\varphi} \frac{\partial Log\ p(k/x_i^k;\hat{\varphi}_n)}{\partial\varphi'} \left[\frac{\partial^2 L_n^*(\hat{\varphi}_n)}{\partial\varphi\partial\varphi'} \right]^{-1}.$$

3.2.2 *Ensemble fini de valeurs*

Supposons (comme dans le cas d'échantillonnage exogène) que l'ensemble des valeurs possibles de X est fini et noté par $\{x_j, j = 1,...,J\}$.

La procédure du maximum de vraisemblance pondérée est encore applicable lorsque la distribution marginale $p(k)$ est connue. Cependant, il existe une procédure alternative ayant de meilleures propriétés :

- elle est plus simple numériquement ;
- quand les probabilités $p(k)$ sont connues, elle est asymptotiquement plus efficace que la procédure du maximum de vraisemblance pondérée ;
- quand la distribution $p(k)$ est inconnue, elle fournit un estimateur convergent et asymptotiquement normal de θ.

L'idée de base est de partir d'un modèle linéaire analogue à celui de la section 3.1 :

$$Log \frac{n_{jk}}{n_{j0}} = \tilde{\beta}(k) + kx_j\theta + \tilde{v}_{jk}, \quad j = 1, \ldots, J, \quad k = 1, \ldots, K.$$

La limite de $Log \frac{n_{jk}}{n_{j0}}$ est :

$$\lim_{n \to \infty} \quad Log \frac{n_{jk}}{n} . \frac{n}{n_{j0}} = Log \frac{\alpha_k \pi(j/k)}{\alpha_0 \pi(j/0)},$$

où : $\pi(j/k) = p(X = j / Y = k)$.

En utilisant la notation $\pi(j) = p(X = j)$, cette limite est aussi égale à :

$$\begin{aligned} & Log \frac{\alpha_k}{\alpha_0} + Log \left\{ \frac{p(k/j)\pi(j)}{p(k)} \frac{p(0)}{p(0/j)\pi(j)} \right\} \\ = \; & Log \frac{\alpha_k}{\alpha_0} + Log \frac{p(0)}{p(k)} + \beta(k) + kx_j\theta \\ = \; & \tilde{\beta}(k) + kx_j\theta, \end{aligned}$$

avec : $\tilde{\beta}(k) = \beta(k) + Log \frac{\alpha_k}{\alpha_0} + Log \frac{p(0)}{p(k)}.$

Donc toute procédure de type moindres carrés généralisés fournit des estimateurs convergents et asymptotiquement normaux des paramètres $\tilde{\beta}$ (k) et θ. Ainsi tous les biais provenant du tirage endogène sont capturés dans les paramètres de type β, mais ne concernent pas l'estimation du score servant au classement des individus. Il s'agit d'une propriété importante de la formulation semi-paramétrique ALDP introduite. De plus si la fonction $p(.)$ est connue, on en déduit des estimateurs convergents des paramètres $\beta(k)$. En particulier, si la matrice de variance-covariance est à tort fixée au niveau de celle Ω^* correspondant à l'échantillonnage exogène (voir section 3.1), des estimateurs convergents sont néanmoins obtenus. Les matrices de variance-covariance, correctement et incorrectement calculées, sont étudiées et comparées dans Gouriéroux-Monfort (1989b) .

La procédure optimale, qui domine la procédure du maximum de vraisemblance pondérée, correspond à la matrice de variance-covariance des erreurs estimées :

$$\begin{bmatrix} diag\left[\frac{1}{n_{1.}}-\frac{1}{n_{.}}\right]+\left[\frac{1}{n_{10}}-\frac{1}{n_0}\right]U\ldots & diag\left[-\frac{1}{n_{.}}\right]-\frac{1}{n_0}U\ldots & diag\left[-\frac{1}{n_{.}}\right]-\frac{1}{n_0}U \\ & diag\left[\frac{1}{n_{2.}}-\frac{1}{n_{.}}\right]+\left[\frac{1}{n_{20}}-\frac{1}{n_0}\right]U & \cdots \\ diag\left[-\frac{1}{n_{.}}\right]-\frac{1}{n_0}U & \cdots\cdots & diag\left[\frac{1}{n_{J.}}-\frac{1}{n_{.}}\right]+\left[\frac{1}{n_{J0}}-\frac{1}{n_0}\right]U \end{bmatrix}$$

où U est la matrice $(K \times K)$ dont tous les éléments sont égaux à 1. Cette matrice est facilement inversée (voir Gouriéroux-Monfort, 1989b), et la procédure peut être mise en oeuvre simplement.

4. HÉTÉROGÉNÉITÉ

4.1 *Estimation*

Jusqu'à présent, la loi de la variable endogène Y_i dépend de l'individu i seulement à travers la valeur x_i. Autrement dit, l'hétérogénéité de la population a été complètement capturée par les variables exogènes. Cette hypothèse peut être trop forte et il peut exister des caractéristiques non connues de l'économètre comme les variables d'effort à la base des théories du hasard moral. Nous notons ε_i l'agrégat de ces effets, supposons qu'il affecte additivement l'indice $x_i\theta$ et que les ε_i sont indépendants de même loi et indépendants des x_i. La loi conditionnelle de $Y_i = k$ sachant x_i et ε_i est :

$$p(k/x_i,\varepsilon_i) = \frac{exp[\beta(k)+k(x\theta+\varepsilon_i)]}{\sum_{\ell=0}^{K} exp[\beta(\ell)+\ell(x_i\theta+\varepsilon_i)]}.$$

ε_i est inobservable et on peut marginaliser en supposant que la densité $g(\varepsilon,a)$ de ε_i, appartient à une famille indicée par un paramètre a, la probabilité conditionnelle de $Y_i = k$ sachant x_i :

$$p(k/x_i) = \int \frac{exp[\beta(k)+kx_i\theta](exp\,\varepsilon_i)^k}{\sum_{\ell=0}^{K} exp[\beta(\ell)+\ell x_i\theta](exp\,\varepsilon_i)^\ell} g(\varepsilon_i,a)d\varepsilon_i,$$

Cette expression n'est pas, en général, facilement calculable et la méthode du maximum de vraisemblance n'est pas simple à utiliser. Cependant il est possible d'appliquer une méthode du pseudo-maximum de vraisemblance ou du maximum de vraisemblance simulé (voir Gouriéroux-Monfort, 1996) ou de retenir des approximations non paramétriques de la loi de g permettant des intégrations explicites (McFadden-Train, 1996).

4.2 Prévision de l'effet individuel

Une fois les paramètres (β,θ,a) estimés, il peut être utile de prévoir les caractéristique non observables $v_i = exp\varepsilon_i$ de l'individu i. La prévision optimale de v_i est l'espérance conditionnelle :

$$E(v_i / Y_i = k) = \frac{\int v^{k+1} \frac{exp[\beta(k) + kx_i\theta]}{\sum_{\ell=0}^{K} exp[\beta(\ell) + \ell x_i\theta]v^\ell} f(v,a)dv}{\int v^{k} \frac{exp[\beta(k) + kx_j\theta]}{\sum_{\ell=0}^{K} exp[\beta(\ell) + \ell x_i\theta]v^\ell} f(v,a)dv},$$

où $f(v,a)$ est la densité de v. Cette expression peut être calculée par simulation puisqu'il s'agit d'un rapport d'espérances par rapport à la loi de $v = exp\varepsilon$, de densité $f(v,a)$. En notant $v_h = exp(\varepsilon_h)$ $h = 1,...,H$, des tirages indépendants dans cette loi, après remplacement de a par son estimation, on obtient :

$$\hat{v}(k,i) = \frac{\sum_{h=1}^{H} \frac{exp[\beta(k) + kx_i\theta]v_h^{k+1}}{\sum_{\ell=0}^{K} exp[\beta(\ell) + \ell x_i\theta]v_h^\ell}}{\sum_{h=1}^{H} \frac{exp[\beta(k) + kx_i\theta]v_h^{k}}{\sum_{\ell=0}^{K} exp[\beta(\ell) + \ell x_i\theta]v_h^\ell}} = \frac{\sum_{h=1}^{H} \frac{v_h^{k+1}}{d(v_h,i)}}{\sum_{h=1}^{H} \frac{v_h^{k}}{d(v_h,i)}},$$

avec : $d(v_h,i) = \sum_{\ell=0}^{K} exp[\beta(\ell) + \ell x_i\theta]v_h^\ell.$

Une démarche analogue peut être appliquée dans un cadre dynamique supposons que, pour un individu i donné, on ait observé (en omettant l'indice i pour simplifier les notations) :

$\underline{y_T} = (y_1,...,y_T),$

$\underline{x_{T+1}} = (x_1,...,x_{T+1}),$

et que l'on veuille prévoir y_{T+1} de façon optimale. La prévision est :

$$\begin{aligned}
&E[y_{T+1} / \underline{y_T}, \underline{x_{T+1}}] \\
&= E[E[y_{T+1} / \underline{y_T}, \underline{x_{T+1}}, v] / \underline{y_T}, \underline{x_{T+1}}] \\
&= E\left[\frac{\sum_{\ell=0}^{K} \ell v^\ell exp[\beta(\ell) + \ell x_{T+1}\theta]}{\sum_{\ell=0}^{K} v^\ell exp[\beta(\ell) + \ell x_{T+1}\theta]} / \underline{y_T}, \underline{x_{T+1}}\right] \\
&= E[\Psi_{T+1}(v) / \underline{y_T}, \underline{x_{T+1}}]
\end{aligned}$$

avec des notations évidentes. Cette espérance conditionnelle peut aussi être calculée par simulation. En effet la densité conditionnelle de v sachant $\underline{y_T}/\underline{x_{T+1}}$ est :

$$\frac{f(v)\prod_{t=1}^{T} p(y_t/x_t, v)}{\int f(\theta)\prod_{t=1}^{T} p(y_t/x_t, v)dv}$$

et l'espérance conditionnelle recherchée s'écrit :

$$\frac{\int \Psi_{T+1}(v)\prod_{t=1}^{T} p(y_t/x_t, v)f(v)dv}{\int \prod_{t=1}^{T} p(y_t/x_t, v)f(v)dv}$$

$$= \frac{E_v\left[\Psi_{T+1}(v)\prod_{t=1}^{T} p(y_t/x_t, v)\right]}{E_v\left[\prod_{t=1}^{T} p(y_t/x_t, v)\right]}$$

Les deux espérances apparaissant au numérateur et au dénominateur peuvent être approchées de la même façon que précédemment à partir de tirages dans la loi de v. Dans le contexte actuariel cette prévision et sa mise à jour seront la base des systèmes de bonus-malus (voir Lemaire, 1985 et Dionne-Vanasse, 1992).

ANNEXE 1

PREUVE DE LA PROPRIÉTÉ 1.8

On a :

$$E[C(Y)/Y \geq K_0] = \frac{\sum_\ell C(\ell) 1\mathrm{I}_{\ell \geq K_0}\, exp(\beta(\ell) + \ell x\theta)}{\sum_\ell 1\mathrm{I}_{\ell \geq K_0}\, exp(\beta(\ell) + \ell x\theta)}.$$

La dérivée de cette fonction par rapport à $x\theta$ est :

$$\frac{\partial E[C(Y)/Y \geq K_0]}{\partial(x\theta)} = \frac{\sum_\ell C(\ell) 1\mathrm{I}_{\ell \geq K_0}\, \ell\, exp(\beta(\ell) + \ell x\theta)}{\sum_\ell 1\mathrm{I}_{\ell \geq K_0}\, exp(\beta(\ell) + \ell x\theta)}$$

$$- \frac{\sum_\ell C(\ell) 1\mathrm{I}_{\ell \geq K_0}\, exp(\beta(\ell) + \ell x\theta)}{\sum_\ell 1\mathrm{I}_{\ell \geq K_0}\, exp(\beta(\ell) + \ell x\theta)} \frac{\sum_\ell 1\mathrm{I}_{\ell \geq K_0}\, \ell\, exp(\beta(\ell) + \ell x\theta)}{\sum_\ell 1\mathrm{I}_{\ell \geq K_0}\, exp(\beta(\ell) + \ell x\theta)}$$

$$= E[C(Y)Y/Y \geq K_0] - E(C(Y)/Y \geq K_0)E(Y/Y \geq K_0)$$

$$= Cov[(C(Y), Y)/Y \geq K_0].$$

Cette covariance est non négative, puisque $C(Y)$ est fonction non décroissante de Y.

ANNEXE 2

MATRICE DE VARIANCE-COVARIANCE ASYMPTOTIQUE DE L'ESTIMATEUR DE L'ANALYSE DISCRIMINANTE

La preuve est analogue à celle donnée dans Lo (1986) pour une variable qualitative Y, et on en décrit seulement les arguments différents.

Soit $\sigma = (\sigma_{11},...,\sigma_{p1},\sigma_{22},...,\sigma_{p2},...,\sigma_{pp})'$ le vecteur de taille $p(p+1)/2$ des éléments distincts de Σ et R la matrice de sélection de taille $[1/2\ p(p+1)] \times p^2$ telle que : $R'\sigma = vec(\Sigma)$. Soit Q' l'inverse de Penrose-Moore de R, telle que $RQ' = Id$.

On a : $\sqrt{n}\begin{bmatrix} \hat{a}-a \\ vec(\hat{\Sigma}) - vec(\Sigma) \end{bmatrix} \xrightarrow{L} N[0, V]$,

où : $V = \begin{bmatrix} \Sigma(VY)^{-1} & 0 \\ 0 & 2R'Q(\Sigma \otimes \Sigma)Q'R \end{bmatrix}$.

Si on considère l'estimateur par analyse discriminante : $\tilde{\theta}_0 = \tilde{\Sigma}^{-1}\tilde{a},$ et la fonction reliant (a,Σ) et $\theta = f(a, vec\ \Sigma)$, on a :

$$\sqrt{n}(\tilde{\theta} - \theta) \xrightarrow{d} N[0, JVJ'],$$

où : $J = \left[\frac{\partial f}{\partial a}, \frac{\partial f}{\partial vec\Sigma}\right] = [\Sigma^{-1}, -(a' \otimes Id)(\Sigma^{-1} \otimes \Sigma^{-1})]$.

On en déduit :

$$JV J' = (VY)^{-1}\Sigma^{-1} + 2(a' \otimes Id)(\Sigma^{-1} \otimes \Sigma^{-1})$$
$$\times R'Q(\Sigma \otimes \Sigma)Q'R(\Sigma^{-1} \otimes \Sigma^{-1})(a \otimes Id).$$

ANNEXE 3

PSEUDO-MAXIMUM DE VRAISEMBLANCE SIMULÉ

Par exemple, l'approche du pseudo-maximum de vraisemblance simulé consiste à calculer par simulation une approximation des deux premiers moments conditionnels de Y sachant x_j. En notant u_{ih}, $h = 1,...,H$ les valeurs obtenues à partir de tirages indépendants dans une loi fixe telle que, pour une fonction φ bien choisie, $\varphi(u_{ih}, a)$ ait pour loi la loi de densité $g(.,a)$, les deux premiers moments de Y_i sachant x_i sont approximés par :

$$m_{1H}(x_i; \beta, \theta, a) = \frac{1}{H} \sum_{h=1}^{H} \sum_{k=0}^{K} kp(k / x_i) \varphi(u_{ih}, a)$$

$$m_{2H}(x_i; \beta, \theta, a) = \frac{1}{H} \sum_{h=1}^{H} \sum_{k=0}^{K} k^2 p(k / x_i) \varphi(u_{ih}, a)$$

Ensuite une pseudo vraisemblance fondée sur m_1 et m_2 peut être calculée afin d'obtenir des estimateurs convergents et asymptotiquement normaux de β,θ et a. Il faut noter que l'approximation des moments par m_{1H} et m_{2H} ne change pas les propriétés des estimateurs pourvu que m et H tendent vers $+\infty$ de façon que $H / \sqrt{T}$ tende vers $+\infty$ (n est le nombre d'observations).

Il serait aussi possible d'utiliser une méthode de vraisemblance simulée puisque $p(k/x_i)$ peut être approximé par :

$$\frac{1}{H} exp[\beta(k) + kx_i\theta] \sum_{h=1}^{H} \frac{exp\ (k\varphi(u_{ih}, a))}{\sum_{\ell=0}^{K} exp[\beta(\ell) + \ell x_i \theta + \ell \varphi(u_{ih}, a)]}$$

Pour le choix de $g(.,a)$, il n'y a *a priori* pas de restriction pourvu que la simulation soit aisée et que le paramètre a soit identifiable. Ainsi, comme dans le cas d'un modèle de Poisson avec hétérogénéité conduisant au modèle binomial négatif, on peut supposer que $exp(\varepsilon_i)$ suit une loi $\gamma(a,a)$ dont la densité est

$\amalg_{\mathbb{R}^+}(z) \frac{1}{\Gamma(a)} z^{a-1} a^a e^{-az}$ la variance de cette loi est $\frac{1}{a}$ et sa moyenne 1.

Un autre choix naturel pour la densité de ε_i serait une loi normale ; afin que le modèle soit identifiable on pourrait imposer

$E\varepsilon_i = 0$ (et $a = V\varepsilon_i$) ou $E\varepsilon_i = -a/2, V\varepsilon_i = a$, ce qui implique $E\ exp(\varepsilon_i) = 1$.

BIBLIOGRAPHIE

AMEMIYA, T., et J. POWELL (1983), « A Comparison of the Logit Model and Normal Discriminant Analysis when the Independent Variables Are Binary », dans S. KARLIN, T. AMEMIYA et L. GOODMAN, Eds., *Studies in Econometrics, Time Series and Multivariate Statistics,* Academic Press, New York.

BOYER, M., et G. DIONNE (1989), « An Empirical Analysis of Moral Hazard and Experience Rating », *Review of Economics and Statistics,* LXXI : 128-134.

CAMERON, A., et R. TRIVEDI (1986), « Econometric Models Based on Count Data : Comparison and Application of Some Estimators and Tests », *Journal of Applied Econometrics,* 1 : 29-53.

COX, D.R. (1975), « Partial Likelyhood », *Biometrika,* 62 : 269-276.

COX, D.R., et D. OAKES (1984), Analysis of Survival Data, Chapman Hall.

DIONNE, G., et C. VANASSE (1992), « Automobile Insurance Rate Making in the Presence of Asymmetrical Information », *Journal of Applied Econometrics,* 7 : 149-165.

DIONNE, G., C. GOURIÉROUX, et C. VANASSE (1996), « Evidence of Adverse Selection in Automobile Insurance Markets », CREST DP.

EFRON, B. (1975), « A Comment on Discriminant Analysis Versus Logit Analysis », *Annals of Economic and Social Measurement,* 5 : 511-523.

FOURGEAUD, C., C. GOURIÉROUX, et J. PRADEL (1990), « Hétérogénéité et dominance des fonctions de hasard », *Annales d'Économie et de Statistique,* 18 : 1-24.

GOURIÉROUX, C. (1989), *Économétrie des variables qualitatives,* 2e Édition, Economica.

GOURIÉROUX, C., et A. MONFORT (1989a), *Statistique et modèles économétriques* (2 volumes), Economica, Paris.

GOURIÉROUX, C., et A. MONFORT (1989b), « Econometrics Based on Endogenous Samples », CREST 8903.

GOURIÉROUX, C., et A. MONFORT (1989c), « A General Framework for Testing a Null Hypothesis in a Mixed Form », *Econometric Theory,* 5 : 63-82.

GOURIÉROUX, C., et A. MONFORT (1996), *Simulation Based Econometric Methods,* Oxford University Press.

GOURIÉROUX, C., A. MONFORT, et A. TROGNON (1984), « Pseudo-Maximum Likelyhood Methods : Application to Poisson Models », *Econometrica,* 52 : 701-720.

HALL, B., A. GRILICHES, et J. HAUSMAN (1986), « Patents and R and D : Is there a Lag ? », *International Economic Review,* 27 : 265-283.

HAUSMAN, J., B. HALL, et Z. GRILICHES (1984), « Econometric Models for Count Data with an Application to the Patents-R and D Relationship », *Econometrica,* 52 : 910-938.

HECKMAN, J.J., et B. SINGER (1984), « Econometric Duration Analysis », *Journal of Econometrics,* 24 : 63-132.

HOY, M. (1982), « Categorizing Risks in the Insurance Industry », *Quarterly Journal of Economics*, 97 : 321-337.

KALBFLEISCH, J., et R. PRENTICE (1980), *The Statistical Analysis of Failure Time Data*, Wiley.

LANCASTER, T. (1990), *The Analysis of Transition Data*, Cambridge University Press.

LEMAIRE, J. (1985), *Automobile Insurance : Actuarial Models*, Kluwer, Boston.

LO, A. (1984), « Essays in Financial and Quantitative Economics », Ph.D., Harvard Univ.

LO, A.W. (1986), « Logit versus Discriminant Analysis », *Journal of Econometrics*, 31 : 151-178.

MANSKI, C., et F. LERMAN (1977), « The Estimation of Choice Probabilities from Choice Based Samples », *Econometrica*, 45 : 1977-1988.

MCFADDEN, D. (1979), « A Comment on Discriminant Analysis Versus Logit Analysis », *Annals of Economic and Social Measurement*, 5 : 511-524.

MCFADDEN, D., et K. TRAIN (1996), « Simulation Estimation of MNP and Non Parametric Discrete Choice Models Using Mixed MNL Kernels », paper presented at the EC^2 conference, Florence.

PUELZ, R., et A. SNOW (1994), « Evidence on Adverse Selection : Equilibrium Signalling and Cross Subsilization in the Insurance Market », *Journal of Political Economy*, 102 : 236-257.

SZROETER, J. (1983), « Generalized Wald Methods for Testing Nonlinear Implicit and Overidentifying Restrictions », *Econometrica*, 51 : 335-353.