VARIABLES LATENTES ET MODÉLISATION STATISTIQUE EN ASSURANCE

A. MONFORT*

*CNAM et CREST-INSEE

1. INTRODUCTION.

Les variables latentes, ou inobservables, jouent un rôle de plus en plus important dans la modélisation statistique des problèmes de toute nature. Les raisons de ce succès sont essentiellement de deux types. Sur un plan théorique, les variables latentes permettent de prendre en compte plus finement les différentes composantes d'un phénomène complexe sans se soucier nécessairement de leur observabilité (la limite étant l'identifiabilité du modèle, c'est-à-dire la possibilité de remonter aux paramètres inconnus à partir des éléments observables).

Les variables latentes ont déja prouvé leur efficacité dans de nombreux domaines, par exemple en économie (analyse conjoncturelle, composantes saisonnières), en finance (modèles à facteurs, modèles à volatilité stochastique), en biostatistique (chaînes de Markov cachées pour le séquençage du génôme), en traitement d'images (restauration).

En assurance, le rôle des variables latentes est encore peu important, pourtant les champs d'applications possibles sont très nombreux. L'objectif de cet article est de décrire rapidement les modèles et les méthodes statistiques utilisées en distinguant deux grands domaines : les modèles statiques et les modèles dynamiques.

Dans toute la suite U designera une variable (éventuellement vectorielle) aléatoire latente.

2. CAS STATIQUE.

On va d'abord décrire les objectifs généraux de l'introduction de variables latentes dans le cas statique, on présentera plus en détail deux modèles classiques de traitement de l'hétérogénéité inobservable et on discutera le problème de l'équilibre entre la maniabilité et la généralité des modèles.

2.1 Objectifs.

Les deux objectifs principaux de l'introduction de variables latentes dans le cas statique sont la modélisation des propensions (ou des utilités) et de l'hétérogénéité.

2.1.1 Modéliser des propensions ou des utilités.

Information asymétrique.
[voir Dionne-Gouriéroux-Vanasse (1998), Chiappori-Salanié (2000)]

Dans le modèle le plus simple, on considère n contrats indicés par $i = 1, ..., n$. La propension à provoquer un sinistre est définie par :

$$U_{1i} = x_i' b_1 + \varepsilon_{1i} \quad i = 1, \dots, n$$

La variable observable décrivant l'occurence ($Y_{1i} = 1$) ou la non occurence ($Y_{1i} = 0$) d'un sinistre est déduite de U_{1i} par :

$$Y_{1i} = \mathbb{1}_{\mathbb{R}^+}(U_{1i})$$

De même la propension à choisir un contrat "tous risques" est définie par :

$$U_{2i} = x_i' b_2 + \varepsilon_{2i} \quad i = 1, \dots, n$$

et le choix observé est décrit par :

$$Y_{2i} = \mathbb{1}_{\mathbb{R}^+}(U_{2i})$$

On suppose la normalité des erreurs :

$$\varepsilon_i \;=\; \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \sim IIN\left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$$

Il y a information asymétrique (sélection adverse ou risque moral) si V_{1i} et V_{2i} ne sont pas conditionnellement indépendantes, c'est-à-dire si ρ est non nul. Le test de $\rho = 0$ peut être fondé sur les résidus généralisés (qui sont des estimations des espérances conditionnelles de ε_{1i} sachant Y_{1i} et de ε_{2i} sachant Y_{2i}) :

$$\hat{\varepsilon}_{ji} = \frac{\varphi(x_i'\hat{b}_j)}{\Phi(x_i'\hat{b}_j)} Y_{ji} - \frac{\varphi(x_i'\hat{b}_j)}{\Phi(-x_i'\hat{b}_j)}(1 - Y_{ji})$$

et le test du score a pour région critique :

$$\frac{(\sum_i \hat{\varepsilon}_{1i}\hat{\varepsilon}_{2i})^2}{\sum_i \hat{\varepsilon}_{1i}^2 \hat{\varepsilon}_{2i}^2} \geq \chi^2_{1-\alpha}(1)$$

(Gouriéroux-Monfort-Renault-Trognon (1987))

Choix multiples.

Dans les modèles de choix multiples, par exemple le choix entre divers types de contrats, les variables latentes représentent les diverses utilités attribuées par les individus observés aux diverses modalités du choix. Plus précisément on a :

J modalités $j = 1, \ldots, J$
n individus $i = 1, \ldots, n$
$U_{ji} = v_{ji}(b) + \varepsilon_{ji}$ est l'utilité de la modalité j pour l'individu i.

L'individu i choisi $Y_i = j$ si :

$$U_{ji} = \max(U_{1i}, \ldots, U_{Ji})$$

Si les U_{ji} suivent indépendamment une loi de Gompertz, de fonction de répartition $F(\varepsilon) = \exp[-\exp(-\varepsilon)]$, on a :

$$P(Y_i = j) = \frac{\exp[v_{ji}(b)]}{\sum_{j=1} \exp[v_{ji}(b)]}$$

On obtient le modèle logit polytomique.

Décision d'attribution de contrats.

Les éléments d'un modèle d'attribution de contrat avec sélection endogène sont :

$$i = 1, \ldots, n \text{ candidats}$$

la propension à provoquer un sinistre :

$$U_{1i} = x_i' b_1 + \varepsilon_{1i}$$

la propension à accorder un contrat :

$$U_{2i} = x_i' b_2 + \varepsilon_{2i}$$

On observe (par convention) :

$$\begin{aligned}
Y_{1i} = 1, Y_{2i} &= 1 \text{ si } U_{2i} > 0 \; U_{1i} > 0 (\text{contrat accordé, sinistre}) \\
Y_{1i} = 0, Y_{2i} &= 1 \text{ si } U_{2i} > 0 \; U_{1i} < 0 (\text{contrat accordé, pas de sinistre}) \\
Y_{2i} &= 0 \text{ si } U_{2i} < 0 (\text{contrat refusé})
\end{aligned}$$

$$\begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \sim IIN \left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right)$$

Il y a sélection endogène si $\rho \neq 0$
La log vraisemblance est $L = L_1 + L_2$ avec :

$$\begin{aligned}
L_1 &= \sum_{i=1}^{n} y_{2i} \log \Phi(x_i' b_2) + (1 - y_{2i}) \log \Phi(-x_i' b_2) \\
L_2 &= \sum_{i=1}^{n} y_{2i} \left\{ y_{1i} \log \frac{\Phi_2(x_i' b_1, x_i' b_2, \rho)}{\Phi(x_i' b_2)} \right. \\
&+ \left. (1 - y_{1i}) \log \left[1 - \frac{\Phi_2(x_i' b_1, x_i' b_2, \rho)}{\Phi(x_i' b_2)} \right] \right\}
\end{aligned}$$

Il y a un biais (même asymptotique) si on estime b_1 par un Probit sur les contrats acceptés.

2.1.2. Hétérogénéité inobservable : marginalisation ou conditionnement ?

On suppose qu'on a trois types de variables :
(par exemple le nombre de sinistres pour le contrat i)
Y_i : variable endogène observable
X_i : variable exogène observable
U_i : variable exogène inobservable représentant l'hétérogénéité (non prise en compte par X_i).

On s'intéresse à la vraie densité conditionnelle (inconnue) :

$$f_o(y_i/x_i, u_i)$$

On postule :

$$f_o(y_i/x_i, u_i) \in \{f(y_i/x_i, u_i; \theta), \theta \in \Theta\}$$

où f est une fonction donnée et θ un paramètre inconnu.

Deux approches statistiques sont possibles, l'une est fondée sur la marginalisation et l'autre sur conditionnement

a) Marginalisation.

On modèlise aussi la densité conditionnelle $g_o(u_i/x_i)$:
U_i et X_i sont supposées indépendantes et

$$g_o(u_i) \in \{g(u_i; \theta), \theta, \in \Theta\}$$

Alors :

$$f_o(y_i, u_i/x_i) = f_o(y_i/x_i, u_i) g_o(u_i)$$

appartient à :

$$\{f(y_i/x_i, u_i; \theta) g(u_i; \theta), \theta \in \Theta\}$$

et on en déduit par marginalisation que :

$f_o(y_i/x_i)$ appartient à

$$\{\int f(y_i/x_i, u; \theta) g(u; \theta) du, \theta \in \Theta\}$$

θ étant supposé identifiable.

Avantages

• la théorie du maximum de vraisemblance pour l'estimation de θ fonctionne

• on peut estimer $f_o(y_i, u_i/x_i)$ donc $f_o(u_i/y_i, x_i)$ et prévoir U_i (et donc obtenir un coefficient de bonus-malus)

Inconvénients

•U_i et X_i sont supposées indépendantes

• il faut postuler un type de loi pour U_i

• il y a des calcul d'intégrales

b) Conditionnement.

Si on peut trouver $h(Y_i)$ telle que :

$$f_o(y_i/h(y_i), x_i, u_i) = f_o(y_i/h(y_i), x_i)$$

on déduit sans hypothèse supplémentaire une vraisemblance conditionnelle fondée sur $f(y_i/h(y_i), x_i; \theta)$ (θ étant supposé identifiable).

Avantages

• pas d'hypothèse d'indépendance entre U_i et X_i

• pas d'hypothèse sur $f_o(u_i/x_i)$

Inconvénients

• pas toujours possible

• on ne peut pas prévoir U_i

2.2 Deux modèles d'hétérogénéité classiques.

2.2.1) Le modèle de Poisson hétérogène.

Y_{it} représente le nombre de sinistres du contrat i pour l'année t.

On observe :

$$Y_i = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{iT} \end{pmatrix}, X_i = \begin{pmatrix} X_{i1} \\ \vdots \\ X_{iT} \end{pmatrix},$$

U_i représente l'hétérogénéité inobservable et on suppose que ;

$$\begin{aligned} f(y_i/x_i, u_i; b) &= \prod_{t=1}^{T} \exp(-\lambda_{it}) \lambda_{it}^{y_{it}} \frac{1}{y_{it}!} \\ \lambda_{it} &= u_i \exp(x'_{it} b) \end{aligned}$$

a) Marginalisation : Poisson-gamma.

Si on suppose en outre que :

$$\begin{aligned} g(u_i; a) &= \frac{1}{\Gamma(a)} \exp(-a u_i) u_i^{a-1} a^a \\ E(U_i) &= 1, V(U_i) = \frac{1}{a} \end{aligned}$$

on a :
$f(y_i/x_i; a, b) =$

$$\Pi_t \frac{\lambda_{it}^{y_{it}}}{y_{it}!} \left(\frac{a}{\sum_t \lambda_{it} + a} \right)^a \left(\sum_t \lambda_{it} + a \right)^{-\sum_t y_{it}} . \frac{\Gamma(\sum_t y_{it} + a)}{\Gamma(a)}$$

On obtient la loi binomiale négative multivariée et on peut estimer les paramètres a et b par la méthode du maximum de vraisemblance.

La prévision du nombre de sinistre en $T+1$ pour le contrat i est :

$$E(Y_{i,T+1}/y_i, x_i, x_{i,T+1}) = \lambda_{i,T+1} \frac{a + \sum_{t=1}^{T} y_{it}}{a + \sum_{t=1}^{T} \lambda_{it}}$$

Cete prévision fait intervenir un coefficient de bonus-malus fonction de la séquence des sinistres passés (voir Dionne-Vanasse (1992)).

b) Conditionnement : effets fixes.

Si on pose :

$$Y_{i.} = \sum_{t=1}^{T} Y_{it}$$

on a :

$$f(y_i/y_{i.}, x_i, u_i; b)$$

$$= \frac{y_{i.}!}{\Pi_t y_{it}!} \Pi_{t=1}^{T} \left(\frac{\lambda_{it}}{\sum_t \lambda_{it}} \right)^{y_{it}}$$

On obtient donc une loi multinomiale indépendante de u_i et on peut estimer b par la méthode du maximum de vraisemblance.

Il existe des généralisations, faisant éventuellement intervenir les coûts [voir Gouriéroux (1999) chapitre 8, Pinquet (2001), Frangos-Vrontos (2001)]

2.2.2 Modèle à hasard proportionnel hétérogène.

Y_i est une variable aléatoire positive, représentant une durée entre deux sinistres, ou une durée de vie.

La fonction de hasard conditionnel à $X_i = x_i$ et $U_i = u_i \geq 0$ est supposée du type :

$\lambda(y_i/x_i, u_i; b, c) = u_i \exp(x_i' b) h(y_i; c)$

La densité conditionnelle de Y_i sachant $X_i = x_i$ et $U_i = u_i$ est donc :

$f(y_i/x_i, u_i; b, c) = u_i \exp(x_i' b) h(y_i; c) \exp[-u_i \exp(x_i' b) H(y_i, c)]$

où $H(y_i; c) = \int_o^{y_i} h(y; c) dy$ est le hasard de base intégré.

Hétérogénéité

Si on suppose que :

$$g(u_i; a) = \frac{1}{\Gamma(a)} \exp(-a u_i) u_i^{a-1} a^a$$

la fonction de hasard conditionnel à x_i est :

$$\tilde{\lambda}(y_i/x_i; a, b, c) = \frac{\exp(x_i' b) h(y_i; c)}{1 + \frac{1}{a} \exp(x_i' b) H(y_i; c)}$$

et cette fonction décroît lorsque $\sigma^2 = \frac{1}{a}$ croît (c'est le phénomène mobile - stable)
Le profil de hasard conditionnel à x_i peut être très différent de celui conditionnel à x_i et u_i. De même la densité de Y_i conditionnelle à x_i

$$\tilde{f}(y_i/x_i; a, b, c) = \frac{\exp(x_i' b) h(y_i; c)}{[1 + \frac{1}{a} \exp(x_i' b) H(y_i; c)]^{a+1}}$$

peut être très différente de f.

Cas $h(y_i; c) = 1$

$$\lambda = u_i \exp(x_i' b), \text{ constant en } y_i$$

$$f = u_i \exp(x_i' b) \exp[-u_i \exp(x_i' b) y_i], \text{ loi exponentielle}$$

$$\tilde{\lambda} = \frac{\exp(x_i' b)}{1 + \frac{1}{a} y_i \exp(x_i' b)}, \text{ décroissant en } y_i$$

$$\tilde{f} = \frac{\exp(x_i' b)}{[1 + \frac{1}{a} \exp(x_i' b) y_i]^{a+1}}, \text{ loi de Paréto}$$

2.3 Maniabilité contre généralité.

On va d'abord montrer que dans la recherche d'un équilibre entre la maniabilité et la généralité d'un modèle on peut avoir tendance à trop simplifier ce modèle ; cela peut être le cas pour le modèle logit polytomique. On souligne donc l'importance de méthodes robustes et de modèles généraux de type semi-paramétriques (en prenant l'exemple des modèles de comptage).

2.3.1 Bus bleu, bus rouge.

Considérons un modèle logit polytomique

$$U_{ji} = v_{ji}(b) + \varepsilon_{ji} \quad j = 1, \ldots, J, i = 1, \ldots, n$$

où les ε_{ji} sont indépendants de fonction de répartition $F(\varepsilon) = \exp[-\exp(\varepsilon)]$
$Y_i = j$ si $U_{ji} = \max(U_{1i}, \ldots, U_{Ji})$

Alors, pout tout J, et pout tout couple (j, k) le rapport

$$\frac{p_j}{p_k} = \frac{P(Y_i = j)}{P(Y_i = k)} = \exp[v_{ji}(b) - v_{ki}(b)]$$

ne dépend pas des autres alternatives, ce qui peut ne pas être conforme à l'intention.

Considérons le cas $J = 3$ et les alternatives de mode de transport :

1 : métro, 2 : bus bleu, 3 : bus rouge

de probabilités p_1, p_2, p_3
et le cas $J = 2$ avec les alternatives :

1 : métro, 2 : bus bleu

de probabilités $\tilde{p}_1, \tilde{p}_2$
On peut supposer :

$$p_2 = p_3, \tilde{p}_1 = p_1, \tilde{p}_2 = p_2 + p_3$$

alors :

$$\frac{\tilde{p}_1}{\tilde{p}_2} = \frac{p_1}{p_2 + p_3} = \frac{1}{2}\frac{p_1}{p_2}$$

Ce qui est en contradiction avec le modèle logit polytonique.

Une alternative au modèle logit polytomique est le modèle probit polytomique :

$$\begin{pmatrix} \varepsilon_{1i} \\ \vdots \\ \varepsilon_{Ji} \end{pmatrix} \sim N(0, \Sigma)$$

avec les éléments diagonaux de Σ à 1.

$P(Y_i = j)$ est alors une intégrale de dimension $J - 1$:

$$p_j = P(U_{ji} - U_{ki} \leq 0, k \neq j)$$

La présence d'intégrales de grande dimension nécessite l'utilisation de méthodes simulées : méthode des moments simulés, méthode du maximum de vraisemblance simulé (Gouriéroux-Monfort (1996)a).

Par exemple un simulateur sans biais de p_j souvent employé est le simulateur GHK (Geweke-Hajivassiliou-Keane), défini de la façon suivante.

$$\text{On a : } p_j = P(\varepsilon_{ji} - \varepsilon_{ki} \geq v_{ki} - v_{ji}, k \neq j)$$

En notant $W_j = (\varepsilon_{ji} - \varepsilon_{ki})_{k\neq j}, w_j = (v_{ki} - v_{ji})_{k\neq j}$

$$\begin{aligned}
p_j &= P(W_j \geq w_j) \text{ (inégalité composante par composante)} \\
&= P(A_j Z \geq w_j) \;\; Z \sim N(0, I_{J-1}) \\
&\text{avec } A_j Z = \begin{pmatrix} Z_1 \\ Z_2 + a_{2,1} Z_1 \\ \vdots \\ Z_{J-1} + a_{J,J-1} Z_{J-2} + \ldots + a_{J,1} Z_1 \end{pmatrix}
\end{aligned}$$

On tire Z_1^* dans $N(0,1)$ restreinte à $(w_{1j}, +\infty)$: z_1^*

Z_2^* dans $N(0,1)$ restreinte à $(w_{2,j} - a_{2,1} z_1^*, +\infty)$: z_2^*

$\vdots$

Z_{J-1}^* dans $N(0,1)$ restreinte à $w_{J-1,j} - a_{J,J-1} z_{J-2}^* - \ldots - a_{J,1} z_1^*$: z_{J-1}^*

Le simulateur de p_j est :

$$\tilde{p}_j = \Phi(-w_{1j}) \prod_{k=2}^{J-1} \Phi(-w_{kj} + a_{k,k-1} z_{k-1}^* + \ldots + a_{k1} z_1^*)$$

On vérifie que ce simulateur est sans biais ; par exemple si $J = 3$ on a :

$$\begin{aligned}
\tilde{p}_j(z_1^*) &= \Phi(-w_{1j}) \Phi(-w_{2j} + a_{21} z_1^*) \\
E\tilde{p}_j(Z_1^*) &= \int_{[w_{1j}, +\infty]} \tilde{p}(z_1^*) \frac{\varphi(z_1^*)}{\Phi(-w_{1j})} dz_1^* \\
&= \int_{[w_{1j}, +\infty]} \Phi[-w_{2j} + a_{21} z_1^*] \varphi(z_1^*) dz_1^* \\
&= p_j
\end{aligned}$$

2.3.2 Estimation robuste d'un modèle de comptage.

La loi conditionnelle $Y_i/(x_i, u_i)$ est la loi de Poisson $\mathcal{P}[u_i \exp(x_i'b)]$.

On ne fait pas l'hypothèse sur la loi de U_i mais seulement sur ses deux premiers moments :

$E(U_i) = 1, V(U_i) = \sigma^2$

on a alors

$E(Y_i/x_i) = \exp(x_i'b)$

$V(Y_i/x_i) = \exp(x_i'b) + \sigma^2 \exp(2x_i'b)$

La méthode du maximum de vraisemblance n'est plus disponible mais on peut utiliser les méthodes du pseudo maximum de vraisemblance et du pseudo-maximum de vraisemblance quasi généralisé (Gouriéroux-Monfort-Trognon 1984a, 1984b)ou la méthode de moments (Pinquet 2001).

2.3.3 Modèle de comptage semi-paramétrique.

Dans le modèle de Poisson on a :

$$p(k/x_i) = \exp(-\exp(x_i'b))\frac{\exp(kx_i'b)}{k!}$$

et donc :

$$\log[p(k/x_i)] - \log[p(k-1)/x_i] = -\log k + x_i'b$$

Le logarithme de $p(k/x_i)$ (probabilité conditionnelle de $Y_i = k$ sachant $X_i = x_i$) se décompose donc en une somme de deux fonctions : une fonction connue de k uniquement (la fonction - log$k$) et une fonction inconnue linéaire en x_i. Il est donc naturel de généraliser ce modèle en remplaçant la fonction -log$k$ par une fonction $\gamma_o(k)$ inconnue quelconque. On a donc :

$$\log[p(k/x_i)] - \log[p(k-1)/x_i] = \gamma_o(k) + x_i'b$$

et :

$$p(k/x_i) = \exp[\sum_{l=1}^{k} \gamma_o(l) + kx_i'b] / \sum_{k=0}^{K} \exp[\sum_{l=1}^{k} \gamma_o(l) + kx_i'b]$$

avec par convention $\sum_{l=1}^{0} \gamma_o(l) = 0$.

Dans ce contexte on peut estimer $\gamma_o(.)$ et b (par la méthode du maximum de vraisemblance).

Hétérogénéité :

On peut généraliser encore le modèle précédent en introduisant un terme d'hétérogénéité inobservable U_i. On a alors :

$$\log[p(k/x_i, u_i)] - \log[p(k-1)/x_i, u_i] = \gamma_o(k) + x_i'b + u_i$$

où la loi de U_i appartient à une famille de loi $g(u_i; a)$ quelconque avec une convention d'identification du type :

$$EU_i = 0, \text{ ou } E\exp(U_i) = 1$$

Ce type de modèle peut être estimé par la méthode du maximum de vraisemblance simulé ou du pseudo-maximum de vraisemblance simulé (voir Gouriéroux-Monfort (1991), (1993), (1996a)).

Le calcul de

$$E(Y_{i,T+1}/y_{i1}, \ldots, y_{iT}, x_{i1}, \ldots, x_{i,T+1})$$

peut se faire par simulation (Gouriéroux-Monfort (1997)).

Le principe de ce calcul est le suivant en supprimant l'indice i et les x_{it} pour simplifier et en notant $y^T = (y_1, ..., y_T)$:

$$\begin{aligned} E(Y_{T+1}/y^T) &= E[E(Y_{T+1}/y^T, U)/y^T] \\ &= E[E(Y_{T+1}/U)/y^T] \\ &= E[\psi(U)/y^T] \end{aligned}$$

où ψ est une fonction connue.

$$\begin{aligned} E[\psi(U)/y^T] &= \int \frac{\psi(u) f(y^T, u)}{f(y^T)} \; du \\ &= \frac{\displaystyle\int \psi(u) f(y^T/u) g(u) du}{\displaystyle\int f(y^T/u) g(u) du} \\ &= \frac{E_U[\psi(U) f(y^T/U)]}{E_U[f(y^T/U)]} \end{aligned}$$

On peut alors évaluer chacun des termes de ce rapport par une méthode de Monte Carlo fondée sur des simulations dans la loi de U.

3. CAS DYNAMIQUE.

On va d'abord distinguer deux types de dynamiques selon que la variable latente U est exogène ou endogène. On précisera ensuite les objectifs classiques des modèles des deux types. On présentera un modèle incorporant à la fois une dépendance temporelle "réelle" (ou directe) et "apparente" (ou indirecte). On traitera en détail un exemple de modèle dynamique avec variables latentes : celui de la structure par terme des taux d'interêt. Enfin on discutera le problème de l'inférence.

3.1. Deux types de dynamiques.

On considère le processus :

$$\begin{pmatrix} Y_t \\ U_t \end{pmatrix} \quad t = 1, \ldots, T, \; U_t \text{ inobservables}$$

Pour simplifier, on n'introduit pas de variables exogènes et on suppose le processus markovien d'ordre 1. On note :

$$y^T = (y_1, \dots, y_T), \ u^T = (u_1, \dots, u_T)$$

La vraie densité (inconnue) de (Y^T, U^T) peut toujours s'écrire :

$$\begin{aligned} f_o(y^T, u^T) &= \prod_{t=1}^T f_o(y_t, u_t / y_{t-1}, u_{t-1}) \\ &= \prod_{t=1}^T f_o(y_t / y_{t-1}, u_t, u_{t-1}) f_o(u_t / y_{t-1}, u_{t-1}) \end{aligned}$$

<u>1er type de dynamique</u>

$$f_o(y_t / y_{t-1}, u_t, u_{t-1}) = f_o(y_t / u_t)$$

$$f_o(u_t / y_{t-1}, u_{t-1}) = f_o(u_t / u_{t-1})$$

U est exogène, Y ne cause pas U, il n'y a pas de "feedback" de Y vers U

En outre le processus U est (marginalement) markovien

<u>2ème type de dynamique</u>

$$f_o(y_t / y_{t-1}, u_t, u_{t-1}) = f_o(y_t / u_t)$$

$$f_o(u_t / y_{t-1}, u_{t-1}) = f_o(u_t / y_{t-1})$$

U est endogène, Y cause U, il y a feedback de Y vers U

En outre, comme $f_o(y_t, u_t / y^{t-1}, u^{t-1}) = f_o(y_t / u_t) f_o(u_t / y_{t-1})$ on voit (en integrant par rapport à u_t) que le processus Y est (marginalement) markovien.

3.2 Objectifs des modèles du premier type.

On peut distinguer trois types d'objectifs : l'introduction d'une dynamique dans l'hétérogénéité inobservable, les modèles à changements de régimes et les modèles à facteurs.

3.2.1 Hétérogénéité dynamique

Dans ce cas U_t est un processus stationnaire, par exemple autorégressif

[voir Pinquet (2001), Pinquet-Guillen-Bolancé (2001)].

3.2.2 Modèles à changements de régimes

U_t est une chaîne de Markov (dite "chaîne de Markov cachée) ; les modèles de ce type les plus courants sont les suivants :

a) Modèles ARMA à changements de régimes (applications à l'analyse conjoncturelle : Hamilton (1989)), la VaR : (voir Billio-Pelizzon (2000))

b) Modèles à volatilité stochastique discrète, (application aux taux d'intérêt : voir Hamilton (1988))

c) Modèle de Poisson à changement de régimes, (application à la segmentation en marketing (voir Wedel-Desarbo-Bult-Ramaswamy (1993))

3.2.3 Modèles à facteurs

Les modèles à facteurs peuvent avoir des objectifs très différents.

Simplifier la dynamique

Par exemple, c'est le cas d'un modèle GARCH à facteur défini par :

$$Y_t = aU_t + \varepsilon_t$$

$$Y_t : \text{ taille } K$$

$$U_t : \text{ scalaire, modèle GARCH}$$

$$\varepsilon_t : \text{ bruit blanc gaussien}$$

Ce modèle fait passer la dynamique d'un vecteur de taille K par celle d'un scalaire (inobservable).

(Pour une application aux taux de change, voir Diebold-Nerlove (1989))

<u>Enrichir la dynamique</u>

C'est le cas pour les modèles à volatilité stochastique

$$Y_t = \exp(U_t)\varepsilon_t$$

$$U_t = a + bU_{t-1} + \eta_t$$

$$\varepsilon_t, \eta_t \text{ indépendants}$$

$$\varepsilon_t \sim IIN(0, 1)$$

$$\eta_t \sim IIN(0, \sigma^2)$$

ainsi que pour les modèles de durée stochastique :

$$Y_t = \frac{G(1, \Phi(F_{1t})}{aG(b, \Phi(F_{2t}))}$$

(F_{1t}, F_{2t}) processus VAR, de loi marginale $N(0, I_2)$

$G(b, .)$ fonction quantile de $\gamma(b, b)$

(voir Ghysels E, Gouriéroux, C, Jasiak J. (1999)

3.3 Objectifs des modèles du deuxième type.

L'objectif principal de ces modèles est de construire des modèles dynamiques pour les variables à support différent de $I\!R$: par exemple $I\!N$ ou $I\!R^+$.

3.3.1 Processus à valeurs entières : INAR (1)

Ils sont définis par :

$$\begin{cases} Y_t & = & \sum_{i=1}^{Y_{t-1}} U_{it} + \varepsilon_t \\ \\ U_{it} & \sim & II\mathcal{I}(p) \;\; \mathcal{I}(p) \;\; : \text{ loi indicatrice de paramètre } p \\ \\ \varepsilon_t & \sim & \mathcal{P}(\lambda) \text{ indépendamment} \end{cases}$$

ou : $Y_t = U_t + \varepsilon_t$

U_t et ε_t indépendants conditionnellement à $Y_{t-1}, \ldots, U_{t-1} \ldots$ et
$U_t \sim \mathcal{B}(y_{t-1}, p), \varepsilon_t \sim \mathcal{P}(\lambda)$

Alors :

$$\begin{aligned} & EY_t = \frac{\lambda}{1-p} \\ & \gamma(h) = \frac{p^h \lambda}{1-p} \\ & E(Y_{t+h}/Y_t) = p^h Y_t + \frac{1-p^h}{1-p}\lambda \\ & Y_t \sim \mathcal{P}[\lambda/(1-p)] \\ & E\exp(vY_{t+1}/Y_t) = \exp[a(v)Y_t + b(v)] \\ & \text{avec } a(v) = \log[p\exp(v) + 1 - p] \\ & b(v) = \lambda[\exp(v) - 1] \end{aligned}$$

(voir Darolles-Gouriéroux-Jasiak (2001), Gouriéroux-Monfort (2002)

3.3.2 Processus à valeurs positives

Par exemple un processus autorégressif gamma est défini par :

$$\frac{Y_t}{c}/U_t \sim \gamma(\nu + U_t) \ c > 0, \nu > 0$$

$$U_t/Y_{t-1} \sim \mathcal{P}(\rho Y_{t-1}/c) \ \rho > 0$$

et on a :

$$E(Y_{t+h}/Y_t) = \frac{c(1-\rho^h)}{1-\rho}\nu + \rho^h Y_t$$

$$\text{corr } (Y_{t+h}, Y_t) = \rho^h$$

$$\frac{(1-\rho)}{c} Y_t \sim \gamma(\nu)$$

$$E[\exp(vY_{t+1})/Y_t] = \exp[a(v)Y_t + b(v)]$$

avec :

$$a(v) = \frac{\rho v}{1 - vc}, b(v) = -\nu \log(1 - vc)$$

(voir Gouriéroux-Jasiak (2000), Gouriéroux-Monfort (2002)

3.4 Dependance "réelle" et "apparente".

Considérons le modèle défini par :

$U_t = (U_{1t}, U_2)$
$Y_t = U_{1t} + \varepsilon_t$, (nombre de sinistres de la période t)

On suppose que conditionnellement à $Y_{t-1}, U_{1,t-1}, ..., U_2$ les variables U_{1t} et ε_t sont indépendantes

$$U_{1t} \sim \mathcal{B}(y_{t-1}, p) \quad \text{(loi binomiale)}$$

$$\varepsilon_t \sim \mathcal{P}(\lambda u_2) \quad \text{(loi de Poisson)}$$

et, en outre : $U_2 \sim \gamma(a, a)$

Ce modèle contient comme cas particuliers un modèle où la dépendance temporelle est "apparente" (ou "directe") et un modèle où la dépendance est "réelle" (ou "indirecte") :

$p = 0$ cas classique Poisson-Gamma pas de dépendance "réelle"

$\frac{1}{a} = 0$ cas précédent INAR(1), pas de dépendance "apparente" induite par l'hétérogénéité.

Le modèle général contient les deux types de dépendance et peut donc servir à les tester. La prime pure est proportionnelle à :

$$\begin{aligned} E(Y_{T+1}/Y^T) &= E[E(Y_{T+1}/Y^T, U_2)/Y^T] \\ &= E[(pY_T + \lambda U_2)/Y^T] \\ &= pY_T + \lambda E(U_2/Y^T) \end{aligned}$$

qui peut être calculée par des méthodes de Monte Carlo.

On peut aussi introduire des variables exogènes, par exemple en posant :

$\lambda_{it} = \exp(x'_{it} b_1)$

$p_{it} = \dfrac{1}{1 + \exp(-x'_{it} b_2)}$

3.5 Structure par terme des taux d'intérêt.

La modélisation des taux d'intérêt est très importante dans de nombreux domaines, par exemple la gestion actif-passif et l'assurance-vie.

Dans la modèlisation proposée ici on introduit une variable d'état

$$X_t = \begin{pmatrix} r_{t+1} \\ \\ U_t \end{pmatrix} \text{ de taille } p$$

où r_{t+1} est le taux court et U_t un vecteur latent, et on suppose que X_t suit un modèle CAR (compound autoregressive), défini par la transformation de Laplace conditionnelle :

$$E[\exp(w'X_{t+1})/X^t] = \exp[a'(w)X_t + b(w)]$$

On introduit un facteur d'escompte stochastique $M_{t,t+1} \geq 0$, tel que, en absence d'opportunité d'arbitrage le prix en t d'un actif contingent fournissant un flux aléatoire g_{t+h} à la date $t+h$ est donné par :

$$C_t(g_{t+h}) = E_t\left(M_{t,t+1} \dots M_{t+h-1,t+h} g_{t+h}\right)$$

L'absence d'opportunité d'arbitrage pour r_{t+1} implique :

$$\exp(-r_{t+1}) = E_t M_{t+1}$$

et la probabilité risque neutre Q est de densité par rapport à la probabilité historique :

$$\frac{M_{t,t+1} \dots M_{t+h-1,t+h}}{E_t M_{t,t+1} \dots E_{t+h-1} M_{t+h-1,t+h}}$$

$$= \exp(r_{t+1} + \dots + r_{t+h}) M_{t,t+1}, \dots M_{t+h-1,t+h}$$

On peut donc écrire :

$$C_t(g_t + h) = E_t^Q[\exp(-r_{t+1} - \dots - r_{t+h}) g_{t+h}]$$

On suppose en outre que $M_{t,t+1}$ est de la forme :

$$M_{t,t+1} = \exp(\alpha r_{t+2} + \delta' U_{t+1} + \beta_t)$$

L'absence d'opportunité d'arbitrage pour r_{t+1} entraîne :

$$M_{t,t+1} = \exp(\gamma_o + \gamma_1' X_t + \gamma_2' X_{t+1})$$

$$\begin{array}{lrcl} \text{avec} & \gamma_o & = & -b(\alpha, \delta) \\ & \gamma_1 & = & -a(\alpha, \delta) - \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ & \gamma_2 & = & \begin{pmatrix} \alpha \\ \delta \end{pmatrix} \end{array}$$

Structure par terme.

Le prix en t d'un zéro coupon d'échéance $t+h$ est noté $B(t,h)$

et le taux à horizon h :

$$r_{t,t+h} = -\frac{1}{h}\log B(t,h)$$

Alors :

$$\begin{aligned} B(t,h) &= \exp(c'_h X_t + d_h) \\ r_{t,t+h} &= -\frac{c'_h}{h} X_t - d_h \end{aligned}$$

avec :

$$\begin{aligned} c_h &= a\left[c_{h-1} + \begin{pmatrix} \alpha \\ \delta \end{pmatrix}\right] - a\begin{bmatrix} \alpha \\ \delta \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ d_h &= d_{h-1} - b\begin{pmatrix} \alpha \\ \delta \end{pmatrix} + b\left[c_{h-1} + \begin{pmatrix} \alpha \\ \delta \end{pmatrix}\right] \end{aligned}$$

La structure par terme à la date t, constituée des taux $r_{t,t+h}, h$ variant, est donc fonction affine du vecteur des facteurs X_t.

En particulier si on pose :

$$X_t^* = \begin{pmatrix} r_{t+1} \\ r_{t,t+2} \\ \vdots \\ r_{t,t+p} \end{pmatrix}$$

ce vecteur est observable et on a $X_t^* = SX_t + s$

X_t^* a aussi une dynamique CAR

$$E_t[\exp(w' X_{t+1}^*)/X_t^*] = \exp[a^{*'}(w) X_t^* + b^*(w)]$$

et $M_{t,t+1}$ peut s'écrire :

$$M_{t,t+1} = \exp(\gamma_0^* + \gamma_1^{*'} X_t^* + \gamma_2^{*'} X_{t+1}^*)$$

où $a^*(.), b^*(.), \gamma_o^*, \gamma_1^*, \gamma_2^*$ sont déduits de $a(.), b(.), \alpha, \beta$. On peut alors proposer des méthodes d'estimation de α, β et des paramètres θ apparaissant dans $a(.), b(.)$.

La dynamique risque neutre : de $X_t^\star$ est aussi de type CAR, et sa transformée de Laplace conditionnelle est :

$$\exp[A^*(w)X_t^* + B^*(w)]$$

avec :

$$A^*(w) = a^*(w + \gamma_2^*) - a^*(\gamma_2^*)$$

$$B^*(w) = b^*(w + \gamma_2^*) - b^*(\gamma_2^*)$$

On peut donc proposer des méthodes générales de valorisation de produits dérivés de taux (voir Gouriéroux-Monfort (2002))

3.6 Inférence.

Considérons successivement les modèles de type 1 et 2.

3.6.1 Modèles de type 1

Dans ces modèles le problème essentiel est le risque d'explosion numérique. En effet on connaît :

$$f(y_t/y^{t-1}, u^t; \theta)$$

$$g(u_t/y^{t-1}, u^{t-1}; \theta)$$

et donc :

$$f(y^T, u^T; \theta) = \prod_{t=1}^{T} f(y_t/y^{t-1}, u^t; \theta) g(u_t/y^{t-1}, u^{t-1}; \theta)$$

mais on souhaite calculer la vraisemblance :

$$f(y^T;\theta) = \int f(y^T, u^T;\theta)dy^T$$

On est donc confronté à une intégrale de très grande taille impossible à calculer avec précision.

Solutions

1) Structures simples

• markoviennes linéaires : le filtre de Kalman permet de résoudre le problème

• markoviennes et U_t prend un nombre fini de valeur : le filtre de Kitagawa-Hamilton (Hamilton 1989) fournit la solution

2) Reconstruire les U_t grâce à la théorie : c'est le cas du modèle de structure par terme des taux d'intérêt vu plus haut.

3) Sinon :

• remplacer le problème par un problème plus simple en changement de fonction objectif, c'est le cas de la méthode d'inférence indirecte (Gouriéroux-Monfort-Renault (1993)) qui remplace la vraisemblance par une fonction plus simple.

• conserver le même problème, mais utiliser la méthode du rapport de vraisemblance simulé (voir Billio-Monfort-Robert (1998)

• remplacer le problème par un problème plus compliqué (apparemment), c'est le cas de l'approche bayésienne (voir Billio-Monfort-Robert (1999)). L'idée générale est la suivante :

on connaît la densité $f(y^T, u^T, \theta) = f(y^T, u^T/\theta)\pi(\theta)$ et donc on connaît à une constante multiplicative près la densité $f(u^T, \theta/y^T)$. On peut donc approximer cette densité [et donc la densité a posteriori $f(\theta/y^T)$] par des méthodes de simulation de type MCMC (Monte Carlo Markov Chain) comme la méthode de Metropolis Hastings, qui ne nécessitent que la connaissance à une constante multiplicative près de la densité.

3.6.2. Modèles de type 2

Si on connaît la transformée de Laplace conditionnelle

$$E(\exp w'Y_{t+1}/y_t) = \exp[a(w,\varphi)'y_t + b(w,\varphi)]$$

au paramètre φ près [dans le modèle de taux on a $\varphi = (\theta, \alpha, \delta)$], on a des contraintes sur les moments conditionnels du type :

$$E_t \left\{ \exp(w_j' Y_{t+1}) - \exp[a(w_j,\varphi)'y_t + b(w_j,\varphi)] \right\} = 0, \quad j = 1, ..., J$$

On peut donc utiliser la méthode des moments généralisés (voir Gouriéroux-Monfort 1996 b).

4. CONCLUSION.

Comme on vient de le voir rapidement, les potentialités des modèles statistiques utilisant des variables latentes sont très importantes dans le domaine de l'assurance, que ce soit dans une approche statique ou dynamique des problèmes. Leurs mises en application pratiques sont cependant encore limitées en raison essentiellement d'une plus grande complexité de calcul que dans les modèles classiques. On peut cependant espérer que cette situation va évoluer avec la diffusion de logiciels scientifiques performants.

Billio M., Monfort A. et Robert C.P (1998) : "The Simulated Likelihood Ratio Method", Document CREST.

Billio M., Monfort A. et Robert C.P (1999) :"Bayesian Estimation of Switching ARMA Models", Journal of Econometrics, 93, 229-255.

Billio M. et Pelizzon L. (2000) : "Value-at-Risk : a Multivarite Switching Regime Approach", Journal of Empirical Finance, 135.

Chiappori P.A. et Salanié B (2000) : "Testing for Asymmetric Information in Insurance Markets", Journal of Political Economy, 108, 56-78.

Darolles, S., Gouriéroux C. et Jasiak J. (2001) : "Compound Autoregressive Models", Document CREST.

Diebold F. et Nerlove M. (1989) : "The Dynamics of Exchange Rate Volatility : A Multivariate Latent Factor ARCH Model", Journal of Applied Econometrics, 4, 1-21.

Dione G., Gouriéroux C. et Vanasse C. (1998) : "Evidence of Adverse Selection in Automobile Insurance Markets", Document de travail CREST, n° 9824.

Dionne G. et Vanasse (1992) : "Automobile Insurance Ratemaking in the Presence of Asymmetrical Information", Journal of Applied Econometrics, 7, 149-165.

Frangos N.E et Vrontos S.D. (2001) : "Design of Optimal Bonus-malus Systems with a Frequency and a Severity Component on an Individual Basis in Automobile Insurance", ASTIN Bulletin, 31, n°1, 5-26.

Ghysels E., Gouriéroux C. et Jasiak J. (1999) : "Stochastic Volatility Duration Models", Document CREST.

Gouriéroux C. (1999) : "Statistique de l'assurance", Economica, Paris

Gouriéroux C. et J. Jasiak (2000) : "Autoregressive Gamma Processes", Document CREST.

Gouriéroux C. et J. Jasiak (2002) : "Heterogenous INAR(1) Model with Application to Car Insurance", Document CREST.

Gouriéroux C. et A. Monfort (1991) : "Simulation Based Inference in Models with Heterogeneity", Annales d'Économie et de Statistique, n° 20/21, 39-107.

Gouriéroux C. et A. Monfort (1993) : "Simulation Based Inference", Journal of Econometrics, 59, 5-33.

Gouriéroux C. et A. Monfort (1996a) : "Simulation Based Econometric Methods", Oxford University Press.

Gouriéroux C. et A. Monfort (1996b) : "Statistique et modèles économétriques, (2 volumes) Economica.

Gouriéroux C. et A. Monfort (1997) : "Modèles de comptage semi-paramétriques", in Econometrie Appliquée, Economica, Presses des HEC de Montréal.

Gouriéroux C. et A. Monfort (2002) : "Affine Term Structure Models", Document CREST.

Gouriéroux C., Monfort A., Renault E. (1993) : "Indirect Inference", Journal of Applied Econometrics, 8, 85-118.

Gouriéroux C., Monfort A., Renault E. et Trognon A. (1987) : "Generalized Residuals", journal of Econometrics, 34, 5-32.

Gouriéroux C., A. Monfort et A. Trognon (1984a) : "Pseudo Maximum Likelihood Methods Theory", Econometrica, 52, 681-700.

Gouriéroux C., A. Monfort et A. Trognon (1984b) : "Pseudo Maximum Likelihood Methods Applications to Poisson Models" Econometrica, 52, 701-720.

Hamilton J.D (1988) : "Rational Expectations Econometrics Analysis of Changes in Regime : an Investigation of the Term Structure of Interest Rates", Journal of Economic Dynamic and Control, 12, n° 213, 385-423.

Hamilton J.D (1989) : "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle", Econometrica, 57, 357-384.

Pinquet J. (2001) : "Experience Rating Through Heterogeneous Models", in. Handbook of Insurance, Chapitre 15, Kluwer.

Pinquet J., Guillen M., Bolancé C. (2001) : "Allowance for the Age of Claim in Bonus-Malus System", ASTIN Bulletin, 31, n°2, 337-348.

Wedel M., Desarbo W.S., Bult J.R et V. Ramaswamy (1993) : "A Latent Class Poisson Regression Model for Heterogenous Count Data", Journal of Applied Econometrics, 8, 397-411.