

ALAIN MONFORT

Variables latentes et modélisation statistique en assurance

Journal de la société française de statistique, tome 144, n° 3 (2003),
p. 69-95

http://www.numdam.org/item?id=JSFS_2003__144_3_69_0

© Société française de statistique, 2003, tous droits réservés.

L'accès aux archives de la revue « Journal de la société française de statistique » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

VARIABLES LATENTES ET MODELISATION STATISTIQUE EN ASSURANCE

Alain MONFORT *

RÉSUMÉ

Dans cet article nous présentons des utilisations possibles des variables latentes (ou inobservables) dans les modèles statistiques appliqués à l'assurance. On distingue les modèles statiques et dynamiques. Dans le cas statique, on examine les modèles de choix de contrats, l'hétérogénéité inobservable (en particulier le modèle de Poisson hétérogène et le modèle à hasard proportionnel hétérogène), les modèles polytomiques et l'estimation robuste ou semiparamétrique des modèles de comptage. Dans le cas dynamique on distingue deux types de modèles selon que les variables latentes sont exogènes ou non. On étudie en particulier les modèles à changements de régimes, les modèles à facteurs, les processus à valeurs entières ou positives et les modèles de structure par terme des taux d'intérêt.

ABSTRACT

In this paper we present some possible applications of latent variables in insurance statistical models. We distinguish the static case and the dynamic case. In the static case, we focus on the models of contract choice, the unobservable heterogeneity (in particular the heterogeneous Poisson model and the proportional hazard heterogeneous model), some polytomous models and robust or semiparametric estimation of count data models. In the dynamic case, we distinguish two types of models according to the exogeneity status of the latent variables. We consider in particular, the switching regimes models, the factor models, the integer valued or positive processes and the term structure of interest rate models.

1. Introduction

Les variables latentes, ou facteurs, 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

* CNAM et CREST INSEE.

Département de Mathématiques, CNAM, 292 rue Saint Martin, F75141 PARIS Cedex 03.
e-mail : monfort@cnam.fr

observabilité (la limite étant l'identifiabilité du modèle, c'est-à-dire la possibilité de remonter aux paramètres inconnus à partir des seuls éléments observables).

Les variables latentes ont déjà prouvé leur efficacité dans de nombreux domaines d'application, par exemple en économie pour l'analyse conjoncturelle ou la détermination des composantes saisonnières, en finance où la valorisation des produits dérivés repose principalement sur des modèles à volatilité stochastique de même que les modèles de structure par terme des taux supposent des facteurs correspondant aux caractéristiques niveau, pente, courbure des formes de courbe. Citons également les applications en biostatistique où les chaînes de Markov cachées sont utilisées pour le séquençage du génôme et en restauration d'images.

Dans le domaine de l'assurance, les approches par variables latentes sont encore peu utilisées, bien que les champs d'applications possibles soient très nombreux. L'objectif de cet article est de décrire rapidement les modèles et les méthodes statistiques qui pourraient être utilisés en distinguant les modèles statiques (section 2) et les modèles dynamiques (section 3).

Dans toute la suite U désigne la variable (éventuellement vectorielle) latente.

2. Modèles statiques

Les modèles statiques s'appliquent par exemple à l'analyse de risques individuels relatifs à une période donnée¹. L'indice i désignera selon les situations soit l'assuré, soit le contrat. Ces modèles ont pour but de décrire comment les sinistres repérés soit par leur survenance, soit par leur gravité dépendent de caractéristiques de l'assuré ou du contrat. Les variables latentes sont des facteurs de risque intermédiaires, qui seront considérés comme des variables endogènes non observables.

On va d'abord décrire les objectifs généraux de l'introduction de variables latentes dans le cas statique. On insistera particulièrement sur la prise en compte d'une hétérogénéité inobservable et ses conséquences en terme d'évaluation évolutive du risque. Finalement on discutera le nécessaire équilibre entre maniabilité et généralité des modèles.

2.1. Objectifs

On peut identifier deux raisons pour introduire des variables latentes dans le cas statique : i) la description des choix effectués par les individus, qui permet de prendre en compte les phénomènes d'antisélection, et ii) la prise en compte de l'hétérogénéité.

1. Mais aussi à des historiques des sinistres sous des hypothèses d'indépendance temporelle [voir section 2.2].

2.1.1. *Modéliser les préférences et les choix*

i) Risque et choix de contrats

Divers types de contrats sont souvent proposés aux futurs assurés, comme par exemple des contrats d'assurance automobile avec ou sans franchise. Les assurés ne vont pas choisir les contrats au hasard, mais en fonction de l'intérêt qu'ils y trouvent, fonction du coût du contrat et de leur risque anticipé. Cette sélection endogène de contrats doit être prise en compte pour éviter des biais dans la sélection, mais aussi pour savoir quelle information supplémentaire apporte le type de contrat choisi sur la prévision du risque objectif de l'individu, [voir Dionne-Gouriéroux-Vanasse (1998), (2001), Chiappori-Salanié (2000)].

Considérons n assurés (contrats) indicés par $i = 1, \dots, n$. La propension à provoquer un sinistre est définie par :

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

Dans les modèles les plus simples, la variable observable décrit l'occurrence ($Y_{1i} = 1$) ou la non occurrence ($Y_{1i} = 0$) d'un sinistre et est déduite de la propension sous-jacente par :

$$Y_{1i} = \mathbb{1}_{\mathbb{R}^+}(U_{1i}), \text{ (soit } Y_{1i} = 1 \text{ si } U_{1i} > 0, = 0 \text{ sinon).}$$

De même, s'il y a deux types de contrats « tous risques » ou avec franchise, on peut introduire la propension à choisir un contrat « tous risques » :

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

Le choix de contrat observé est alors décrit par :

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

ou $Y_{2i} = 1$ si le contrat tout risque est retenu, $Y_{2i} = 0$, sinon.

Ainsi la spécification fait intervenir divers types de variables : les caractéristiques exogènes x_i de l'assuré, les propensions latentes U_1, U_2 , qui soustendent risque et choix de contrat. Les variables endogènes observables sont ici qualitatives Y_1, Y_2 .

Une telle spécification doit être complétée en se donnant la loi des termes d'erreurs de façon à pouvoir déduire la loi des variables observables sachant les seules caractéristiques exogènes. Ainsi supposons la normalité des erreurs :

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

Les variances de termes d'erreurs peuvent être fixées à 1 pour des raisons d'identifiabilité, seul l'aspect qualitatif du risque étant observable. Concentrons nous maintenant sur l'importance du paramètre de corrélation.

Lorsque la corrélation ρ est non nulle, les propensions U_1, U_2 sont liées et la connaissance du contrat retenu apporte une information supplémentaire pour prévoir le risque. Ceci est souvent interprété en terme d'information asymétrique et d'antisélection. La compagnie d'assurance dispose d'information *a priori* sur l'assuré, résumée par les variables x_i . Cependant le choix de contrat reste informatif, signifiant que l'assuré connaît mieux son risque que la compagnie qui ne dispose que de la seule reconstitution fondée sur les caractéristiques *a priori*.

Il est important de tester si ce phénomène existe ou pas, sachant qu'il dépend de renseignements dont dispose *a priori* la firme. Plus ces renseignements sont nombreux (et bien utilisés), moins le phénomène d'antisélection devrait apparaître [voir Dionne, Gouriéroux, Vanasse (2001)]. Le test de l'hypothèse d'absence d'autosélection $\rho = 0$ peut être fondé sur les résidus généralisés (Gouriéroux-Monfort-Renault-Trognon (1987)). Les résidus généralisés calculent les prévisions de ε_{1i} sachant Y_{1i} et de ε_{2i} sachant Y_{2i} :

$$\hat{\varepsilon}_{j_i} = \frac{\varphi(x'_i \hat{b}_j)}{\Phi(x'_i \hat{b}_j)} Y_{j_i} - \frac{\varphi(x'_i \hat{b}_j)}{\Phi(-x'_i \hat{b}_j)} (1 - Y_{j_i}), \quad j = 1, 2.$$

Ils constituent des approximations des erreurs $\varepsilon_1, \varepsilon_2$ fondées sur les variables endogènes associées. L'idée du test est alors d'examiner si ces approximations sont non corrélées. Plus précisément le test du score de l'hypothèse ($\rho = 0$) consiste à accepter l'hypothèse d'antisélection si :

$$\frac{\left(\sum_i \hat{\varepsilon}_{1i} \hat{\varepsilon}_{2i} \right)^2}{\sum_i \hat{\varepsilon}_{1i}^2 \hat{\varepsilon}_{2i}^2} \geq \chi_{95\%}^2(1),$$

où $\chi_{95\%}^2(1)$ désigne le quantile à 91 % de la loi du khi-deux à un degré de liberté.

ii) Choix multiples

L'approche peut être étendue à des choix concernant plus de deux contrats. Le modèle repose sur une spécification des niveaux d'utilité attribués par les individus aux divers type de contrat. Plus précisément notons :

$j = 1, \dots, J$, les types de contrats,

$i = 1, \dots, n$, les individus,

$U_{j_i} = v_{j_i}(b) + \varepsilon_{j_i}$, l'utilité attribuée au type j par l'individu i .

L'assuré est supposé choisir le contrat maximisant son utilité. Désignons par Y_i la variable qualitative désignant le type de contrat choisit : l'individu i choisit le contrat j : $Y_i = j$ si :

$$U_{j_i} = \max(U_{1i}, \dots, U_{Ji}),$$

c'est-à-dire si : $U_{j_i} > U_{ki}, \forall k \neq j$.

Lorsque les erreurs associées aux niveaux d'utilité ε_{j_i} suivent indépendamment des lois de Gompertz, de fonction de répartition $F(\varepsilon) = \exp[-\exp(-\varepsilon)]$, nous avons :

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

Lorsque les niveaux moyens d'utilité correspondant à un score linéaire $v_{i,j}(b) = x_{i,j}b_j$, on obtient donc le modèle logit polytomique.

iii) Décision d'attribution de contrats

Dans les exemples précédents nous avons considéré les choix de l'assuré. Cependant des sélections endogènes peuvent aussi survenir du fait de l'agent d'assurance. Celui-ci peut préférer refuser de signer un contrat, s'il trouve trop risqué l'assuré potentiel.

Comme dans l'exemple i), nous devons décrire les risques potentiels.

Les propensions à provoquer un sinistre sont :

$$U_{1i} = x'_i b_1 + \varepsilon_{1i},$$

où $i = 1, \dots, n$ désigne l'indice de l'assuré potentiel. Le modèle est complété en introduisant la propension de l'agent d'assurance à accorder un contrat :

$$U_{2i} = x'_i b_2 + \varepsilon_{2i}.$$

Dans le cas le plus favorable, on observe la décision d'accord et le risque s'il y a eu accord, c'est-à-dire :

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

Comme dans l'exemple i), il faut spécifier la loi des termes d'erreurs :

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

et discuter l'existence d'une sélection endogène par l'agent d'assurance. Un tel effet dépend ici encore des caractéristiques *a priori* introduites et peut traduire des informations supplémentaires dont disposerait l'agent et qui ne figureraient pas dans la base de données. Des corrélations significatives peuvent aussi apparaître dans des cas de fraudes, si un agent favorise certains types de clients.

Dans un tel modèle la log-vraisemblance s'écrit $L = L_1 + L_2$, avec :

$$L_1 = \sum_{i=1}^n y_{2i} \log \Phi(x'_i b_2) + (1 - y_{2i}) \log \Phi(-x'_i b_2),$$

$$L_2 = \sum_{i=1}^n y_{2i} \left\{ y_{1i} \log \frac{\Phi_2(x'_i b_1, x'_i b_2, \rho)}{\Phi(x'_i b_2)} \right. \\ \left. + (1 - y_{1i}) \log \left[1 - \frac{\Phi_2(x'_i b_1, x'_i b_2, \rho)}{\Phi(x'_i b_2)} \right] \right\},$$

ou $\Phi_2(\cdot, \cdot, \rho)$ désigne la fonction de répartition jointe d'une loi normale bivariée de corrélation ρ . Par maximisation de cette vraisemblance, on peut déterminer un estimateur de la corrélation ρ et tester la présence de sélection endogène. Évidemment, s'il y a une telle sélection endogène ($\rho \neq 0$), estimer le paramètre b_1 par un modèle Probit sur les contrats acceptés conduit à des biais importants non seulement sur le paramètre, mais aussi sur les valorisations (primes) associées.

2.1.2. Hétérogénéité inobservable

Une autre utilisation des variables latentes est la prise en compte d'hétérogénéité non observable. Les modèles usuels introduisent l'hétérogénéité individuelle par l'intermédiaire des caractéristiques *a priori* x_i . Il s'agit en fait de la partie observable de l'hétérogénéité. D'autres caractéristiques non observables peuvent aussi influencer sur le risque, les choix. On appelle hétérogénéité non observable un résumé de ces variables. Plus précisément on introduit trois types de variables :

Y_i : désigne la variable endogène observable, mesurant par exemple le nombre de sinistres pour le contrat i ,

X_i : les caractéristiques exogènes observables,

U_i : une variable inobservable représentant l'hétérogénéité non prise en compte par X_i .

On peut alors distinguer dans la prévision du risque celle effectuée à partir des seules caractéristiques x_i , de celle effectuée sachant l'ensemble des caractéristiques x_i, u_i . Nous nous proposons de discuter ci-dessous les différences entre de tels conditionnements.

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 traitements de l'hétérogénéité non observable sont alors possibles; l'un est fondé sur la marginalisation par rapport à U et l'autre sur un conditionnement.

a) Marginalisation

Dans cette démarche, on modélise aussi la densité conditionnelle de U sachant x . Par exemple on peut supposer que les variables U_i et X_i sont indépendantes et que la loi de l'hétérogénéité appartient à une famille paramétrée :

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

La loi du couple (u_i, y_i) sachant X_i est alors :

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

et appartient à la famille :

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

La marginalisation consiste à déterminer la loi de la variable endogène sachant les seules variables exogènes observables. Cette loi conditionnelle $f_o(y_i | x_i)$ s'obtient par intégration (marginalisation) et appartient à la famille :

$$\left\{ \int f(y_i | x_i, u_i; \theta) g(u_i; \theta) du, \theta \in \Theta \right\}.$$

Cette approche présente des avantages et des inconvénients. Ainsi la théorie du maximum de vraisemblance s'applique à l'estimation de θ ; de plus on peut aussi estimer $f_o(u_i | y_i, x_i)$ et s'en servir pour prévoir l'hétérogénéité et en déduire un coefficient de bonus-malus (voir section 2.2). En revanche la spécification repose sur des hypothèses fortes, comme l'indépendance entre U_i et X_i , et une forme particulière de loi pour l'hétérogénéité.

b) Conditionnement

L'idée est ici de trouver un conditionnement fondé sur les variables observables permettant d'éliminer l'effet de l'hétérogénéité.

Supposons qu'il existe un résumé $h(Y_i)$ tel que :

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

indépendant de u_i . On en déduit sans hypothèse supplémentaire une vraisemblance conditionnelle fondée sur $f_o(y_i | h(y_i), x_i; \theta)$.

L'avantage est d'éviter l'hypothèse d'indépendance entre U_i et X_i , et de ne pas faire d'hypothèse sur la forme de $f_o(u_i | x_i)$. Cependant cette approche n'est pas toujours possible et ne permet pas de prévoir l'hétérogénéité.

2.2. Deux modèles d'hétérogénéité classiques

L'introduction d'une hétérogénéité non observable est apparue très utile pour expliquer les mises à jour de primes (bonus-malus) dans des modèles fondamentalement statiques. Contrairement aux exemples précédents, on s'intéresse ici à des historiques individuels de sinistres.

2.2.1. Le modèle de Poisson hétérogène

Ce modèle est la base des formules de bonus-malus développées en assurance automobile [voir Lemaire (1995), Greene (1997)]. Dans cet exemple Y_{it} représente le nombre de sinistres du contrat $i = 1, \dots, n$ pour l'année $t = 1, \dots, T$.

Le modèle suppose que les variables Y_{it} sont indépendantes de loi de Poisson $\mathcal{P}[\lambda_{it}]$, avec $\lambda_{it} = u_i \exp(x'_{it}b)$, sachant les caractéristiques exogènes observables x_{it} et non observables u_i . L'aspect statique du modèle provient de l'hypothèse d'indépendance conditionnelle, qui élimine toute liaison dynamique entre les sinistres. Cependant cette propriété d'indépendance temporelle n'est pas stable lorsque l'ensemble d'information est modifié, par exemple lorsque la composante d'hétérogénéité individuelle u_i n'est pas observable.

La distribution conditionnelle aux exogènes, observables ou non, est :

$$f(y_i | x_i, u_i; b) = \prod_{t=1}^T \exp(-\lambda_{it}) \lambda_{it}^{y_{it}} \frac{1}{y_{it}!},$$

$$\text{avec : } \lambda_{it} = u_i \exp(x'_{it}b).$$

en notant

$$y_i = (y_{i1}, \dots, y_{iT})'$$

$$x_i = (x'_{i1}, \dots, x'_{iT})'$$

Il reste à expliquer comment mener l'inférence sur les paramètres d'intérêt à partir des seules variables observables y_{it} , x_{it} . Ceci peut être fait par marginalisation, après avoir spécifié la loi de l'hétérogénéité non observable, ou par conditionnement en essayant d'éliminer son effet.

a) Marginalisation

Le modèle classique, dit binomial négatif, repose sur une spécification gamma de la distribution d'hétérogénéité. Nous supposons les variables U_i , $i = 1, \dots, n$ indépendantes, de même loi gamma de densité :

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

On peut alors déduire la loi conditionnelle aux seules exogènes observables :

$$f(y_i|x_i; a, b) =$$

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

Il s'agit d'une loi binomiale négative multivariée, d'où le nom donné au modèle. Pour une telle loi les observations y_{it} , $t = 1, \dots, T$ correspondant à un même individu ne sont plus indépendantes. Cette dépendance temporelle conduit à une mise à jour des primes en fonction de l'historique des sinistres. Ainsi la prévision à la fin de l'année T du nombre de sinistres pour la période $T + 1$ pour le contrat i est :

$$E(Y_{i,T+1} | \underline{y}_{i,T}, \underline{x}_{i,T}, x_{i,T+1}) = \lambda_{i,T+1} E[u_i | \underline{y}_{i,T}, \underline{x}_{i,T}, x_{i,T+1}] = \lambda_{i,T+1} \frac{a + \sum_{t=1}^T y_{it}}{a + \sum_{t=1}^T \lambda_{it}}.$$

La mise à jour des primes entre deux dates successives est due i) au changement du risque *a priori* $\lambda_{i,T+1}$, ii) à la modification de l'hétérogénéité anticipée. Ce dernier effet est fonction de la séquence des sinistres passés et sert de base au bonus-malus (voir Dionne-Vanasse (1992), Lemaire (1995)). La formule de mise à jour peut par ailleurs s'interpréter en terme de crédibilité [Gerber, Jones (1975)].

b) Conditionnement

Une seconde approche ne suppose pas de loi particulière pour l'hétérogénéité non observable et cherche à éliminer son influence par un conditionnement bien choisi. Considérons le nombre total de sinistres de l'individu :

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

On peut examiner comment se répartit ce nombre entre les années. On a :

$$f(y_i | y_i, x_i, u_i; b) = \frac{y_i!}{\prod_t y_{it}!} \prod_{t=1}^T \left(\frac{\lambda_{it}}{\sum_t \lambda_{it}} \right)^{y_{it}}.$$

Il s'agit d'une loi multinomiale indépendante de l'hétérogénéité u_i . On peut ainsi estimer b par la méthode du maximum de vraisemblance fondée sur cette loi conditionnelle. Cependant cette approche ne permet pas de calculer la prime. En effet cette dernière suppose une prévision de $Y_{i,T+1}$, donc une spécification de la dépendance temporelle, qui dans ce modèle suppose une spécification de la distribution d'hétérogénéité.

Signalons que ces démarches peuvent être étendues de façon à prendre en compte les coûts des sinistres [voir Gouriéroux (1999) chapitre 8, Pinquet (2001), Frangos-Vrontos (2001)].

2.2.2. Modèle à hasard proportionnel hétérogène

L'approche par hétérogénéité non observable peut aussi être utilisée dans des modèles où les survénances des sinistres sont décrites de façon plus détaillée, par exemple par leur date d'arrivée. À titre d'illustration nous considérons le cas de l'assurance vie et notons Y_i la durée de vie de l'individu ² i . Y_i est une variable aléatoire positive dont la loi peut être caractérisée par la fonction de hasard, qui ici coïncide avec le taux de mortalité instantané. Celle-ci est définie par :

$$\lambda(y) = \lim_{dy \rightarrow 0} \frac{1}{dy} P[Y < y + dy | Y > y].$$

Le modèle est défini en faisant dépendre ce taux des variables exogènes observables ou pas. La fonction de hasard conditionnelle à $X_i = x_i$ et $U_i = u_i \geq 0$ est par exemple supposée à hasard proportionnel :

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

Cette spécification permet de distinguer les effets sur le taux de mortalité des exogènes, de l'hétérogénéité non observable, et de l'âge de l'individu. b et c sont des paramètres inconnus.

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 $H(y_i; c) = \int_0^{y_i} h(y; c) dy$ désigne le hasard de base intégré.

De façon à appliquer l'approche par marginalisation, il reste à spécifier la loi de u_i . Si on suppose que cette loi est gamma :

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

la fonction de hasard conditionnelle à x_i seulement est :

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

Le profil de la fonction de hasard en fonction de l'âge diffère sensiblement de celui conditionné par x_i et u_i . Ceci résulte du phénomène dit mobile-stable.

2. Les modèles de durée peuvent aussi être utilisés pour l'assurance automobile. Les variables sont alors doublement indexées par l'individu et l'indice du sinistre.

Les individus les plus risqués, c'est-à-dire avec une grande valeur du facteur d'hétérogénéité u_i , ont une plus forte probabilité de disparaître jeunes. La population survivante correspond donc à des individus moins risqués, ce qui fait décroître le taux de mortalité. Cet effet est évidemment plus marqué lorsque l'hétérogénéité est plus variable. Ainsi on remarque que la fonction de hasard décroît avec $\sigma^2 = 1/a$.

De même la densité de Y_i conditionnelle à $X_i = x_i$:

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

peut être très différente de f .

Les formules ci-dessus se simplifient dans le cas d'un hasard de base constant : $h(y_i; c) = 1$. Lorsque x_i et u_i sont toutes deux connues, la loi est exponentielle avec :

$$\begin{aligned} \lambda &= u_i \exp(x'_i b), \text{ constant en } y_i, \\ f &= u_i \exp(x'_i b) \exp[-u_i \exp(x'_i b)y_i]. \end{aligned}$$

Après marginalisation la loi est une loi de Pareto de fonction de hasard :

$$\tilde{\lambda} = \frac{\exp(x'_i b)}{1 + \frac{1}{a} y_i \exp(x'_i b)}, \text{ décroissante en } y_i, \text{ de densité :}$$

$$\tilde{f} = \frac{\exp(x'_i b)}{\left[1 + \frac{1}{a} \exp(x'_i b)y_i\right]^{a+1}}.$$

2.3. Maniabilité versus généralité

Dans la recherche d'un équilibre entre maniabilité et généralité d'un modèle, on peut avoir tendance à trop simplifier ce modèle. Cela est ainsi le cas pour le modèle logit polytomique. On souligne dans ce paragraphe l'importance de méthodes robustes et de modèles généraux de type semi-paramétrique (en prenant l'exemple des modèles de comptage).

2.3.1. Un inconvénient du modèle logit polytomique

Considérons un modèle de choix multiples logit polytomique (voir 2.1.1 ii) :

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

où les erreurs ε_{ji} sont indépendantes, de fonction de répartition $F(\varepsilon) = \exp[-\exp(-\varepsilon)]$. Les choix sont repérés par la variable qualitative Y_i telle

que : $Y_i = j$, si $U_{j_i} = \max(U_{1_i}, \dots, U_{J_i})$. D'après les calculs du paragraphe 2.1.1. pour tout J , et tout couple (j, k) , le rapport :

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

ne dépend pas des autres alternatives. Ceci peut ne pas être conforme à l'intuition, comme le montre l'exemple classique bus bleu-bus rouge.

Cet exemple concerne le choix d'un mode de transport : métro, bus-bleu, bus-rouge. Comme nous allons le voir, la difficulté vient du fait que les différences entre alternatives ne sont pas de même type, en particulier que la couleur a peu d'impact sur le choix.

Notons, 1 : métro, 2 : bus bleu, 3 : bus rouge, les alternatives de mode de transport, de probabilités p_1, p_2, p_3 .

Considérons aussi le cas où l'individu ne devrait choisir qu'entre $J = 2$ alternatives : 1 : métro, 2 : bus bleu, de probabilités \tilde{p}_1, \tilde{p}_2 .

Il est naturel de supposer que les probabilités de choix dans l'ensemble plus restreint sont liées aux probabilités de sélection initiale par :

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

On voit alors que les rapports des probabilités changent selon que la troisième alternative est ou non proposée :

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

Ceci contredit le modèle logit polytomique.

Il existe diverses façons d'introduire un effet des autres alternatives, comme le modèle logit séquentiel ou le modèle probit. Nous présentons ci-dessous ce dernier modèle. Il repose sur une hypothèse de normalité des termes d'erreurs intervenant dans la définition des niveaux d'utilité :

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

où la matrice de variance-covariance Σ admet des éléments diagonaux égaux à 1. Intuitivement c'est l'existence de corrélations non nulles entre termes d'erreurs, qui va permettre de gérer l'influence des autres alternatives. De façon plus précise, ceci peut être constaté en analysant la loi de la variable de choix Y_i . Nous avons :

$$P(Y_i = j) = p_j = P(U_{j_i} - U_{k_i} \leq 0, k \neq j).$$

Nous ne donnerons pas l'expression explicite de la probabilité élémentaire qui demande le calcul d'intégrales multiples et nécessite l'utilisation de méthodes

simulées d'estimation : méthode des moments simulés, méthode du maximum de vraisemblance simulé (Gouriéroux-Monfort (1996)a).

Il est cependant utile de décrire un simulateur sans biais de p_j souvent employé, appelé simulateur GHK (Geweke-Hajivassiliou-Keane). En effet la forme récursive du simulateur fait clairement apparaître l'influence des autres alternatives.

Nous commençons par écrire sous une forme plus appropriée la probabilité élémentaire. On a : $p_j = P(\varepsilon_{j_i} - \varepsilon_{k_i} \geq v_{k_i} - v_{j_i}, k \neq j)$, ou en notant $W_j = (\varepsilon_{j_i} - \varepsilon_{k_i})_{k \neq j}$, $w_j = (v_{k_i} - v_{j_i})_{k \neq j}$,

$$p_j = P(W_j \geq w_j),$$

où l'inégalité doit être comprise composante par composante.

À ce stade, on voit que ce sont les différences de niveaux d'utilité, composantes de W_j , plus que les niveaux eux-mêmes, qui importent, donc aussi la matrice de variance-covariance Ω_j de W_j , plus que celle initiale Σ . En fait tout vecteur gaussien peut être écrit sous une forme récursive à partir de variables gaussiennes indépendantes. Plus précisément, on a :

$$p_j = P(A_j Z \geq w_j), \text{ avec } Z \sim N(0, I_{J-1}),$$

et

$$A_j Z = \begin{pmatrix} Z_1 \\ Z_2 + a_{2,1} Z_1 \\ \vdots \\ Z_{J-1} + a_{J,J-1} Z_{J-2} + \dots + a_{J,1} Z_1 \end{pmatrix}.$$

La matrice A_j correspond à celle intervenant dans une décomposition de Cholevsky de la matrice de variance-covariance $\Omega_j = A_j A_j'$.

L'algorithme de simulation est le suivant.

On tire Z_1^* dans la loi normale $N(0, 1)$ restreinte à $(w_{1_j}, +\infty)$, ce qui fournit une valeur simulée z_1^* ;

Z_2^* dans la loi normale $N(0, 1)$ restreinte à $(w_{1_j} - a_{2,1} z_1^*, +\infty)$, ce qui fournit une valeur simulée z_2^* ;

⋮

Z_{J-1}^* dans la loi normale $N(0, 1)$ restreinte à $w_{J-1,j} - a_{J,J-1} z_{J-2}^* - \dots - a_{J,1} z_1^*$, ce qui fournit une valeur simulée z_{J-1}^* .

Un simulateur sans biais de p_j est :

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

La difficulté rencontrée avec le modèle logit polytomique est due à un modèle paramétrique trop restrictif, qui peut facilement conduire à des erreurs de spécification et à de mauvais ajustements. Cette difficulté peut être contournée en utilisant des modèles semi-paramétriques plus robustes, avec en contrepartie une perte de précision éventuelle sur le calcul des primes. Des exemples de telles approches semi-paramétriques sont données ci-dessous pour étendre le modèle de Poisson.

2.3.2. Estimation robuste d'un modèle de comptage

Reprenons le modèle de Poisson considéré auparavant. La loi conditionnelle du nombre de sinistres Y_i sachant X_i, U_i est la loi de Poisson $\mathcal{P}[u_i \exp(x'_i b)]$. Le paramètre b peut être estimé de façon convergente sans faire d'hypothèse forte sur la distribution d'hétérogénéité. Ainsi supposons que ses deux premiers moments conditionnels soient du type :

$$E(U_i|x_i) = 1, \quad V(U_i|x_i) = \sigma^2.$$

On en déduit alors des formes fonctionnelles pour les deux premiers moments conditionnels de la variable endogène observable :

$$E(Y_i|x_i) = \exp(x'_i b),$$

$$V(Y_i|x_i) = \exp(x'_i b) + \sigma^2 \exp(x'_i b).$$

Bien que la méthode du maximum de vraisemblance ne soit plus disponible les expressions des moments servent de base à des méthodes d'inférence semi-paramétriques : méthode du pseudo-maximum de vraisemblance et du pseudo-maximum de vraisemblance quasi-généralisé (Gouriéroux-Monfort-Trognon 1984a, 1984b) ou méthode de moments (Pinquet 2001).

2.3.3. Modèle de comptage semi-paramétrique

Nous nous proposons dans ce paragraphe de discuter une formulation semi-paramétrique, qui permet de modifier simplement la forme des probabilités élémentaires d'une loi de Poisson. Reprenons le modèle de régression 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 la 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 (la fonction $-\log k$) et une fonction inconnue linéaire en x_i . Il est naturel

de généraliser le modèle de Poisson en remplaçant la fonction $-\log k$ par une fonction $\gamma_o(k)$ quelconque inconnue. On a alors :

$$\log [p(k|x_i)] - \log [p(k-1)|x_i] = \gamma_o(k) + x'_i b,$$

et par remplacement récursif :

$$p(k|x_i) = \exp \left[\sum_{l=1}^k \gamma_o(l) + \lambda x'_i b \right] / \sum_{k=0}^K \exp \left[\sum_{l=1}^k \gamma_o(l) + k x'_i b \right],$$

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

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

Examinons alors comment cette spécification peut inclure une hétérogénéité non observable. Vue l'interprétation de celle-ci en terme de variables exogènes omises, il est naturel d'écrire :

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

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

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

Ce type de modèle, c'est-à-dire les paramètres $b, a, \gamma_o(\cdot)$, peut être estimé par la méthode du maximum de vraisemblance simulé ou du pseudo-maximum de vraisemblance simulé (voir Gouriéroux-Monfort (1991), (1993), (1996a)). De plus il est aussi adapté au calcul des primes de fréquences :

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

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 $y^T = (y_1, \dots, y_T)$. Par conditionnement successif, on a :

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

où ψ est une fonction connue. De plus

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

On peut alors évaluer chacun des termes de ce rapport par une méthode de Monte-Carlo fondée sur des simulations dans la distribution d'hétérogénéité.

3. Cas dynamique

La discussion de la section précédente a montré l'importance de la spécification des liaisons temporelles entre sinistres successifs, notamment pour obtenir des formules de mise à jour des primes. Cette dépendance temporelle a été introduite par l'intermédiaire d'un facteur d'hétérogénéité non observable indépendant du temps, donc de façon indirecte. Il existe cependant d'autres façons d'introduire cet aspect dynamique i) en incluant des résumés des sinistres passés parmi les variables explicatives observables, ii) en permettant au facteur non observable³ de varier dans le temps. Dans ce dernier cas il est souvent interprété comme une variable mesurant les efforts faits par l'individu pour influencer sur son risque, donc en terme de hasard moral.

Dans le premier paragraphe, nous distinguons deux types de dynamiques pour le couple (Y_t, U_t) . Nous donnons des exemples de telles spécifications dans les paragraphes suivants et terminons par une discussion de l'inférence statistique. Pour simplifier la présentation nous ne faisons plus figurer l'indice individuel i , ni les caractéristiques individuelles observables x_i .

3.1. Deux types de dynamiques

Considérons le processus joint de la variable de risque et du facteur : $(Y_t, U_t), t = 1, \dots, T$, où en général le facteur est inobservable. Pour simplifier, on suppose que ce processus est markovien d'ordre 1. On note :

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

3. Nous verrons plus loin que ce facteur ne s'interprète plus nécessairement en terme d'hétérogénéité.

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

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

Spécifier la loi jointe revient alors à spécifier les composantes de cette décomposition, notamment la façon dont elles dépendent des variables conditionnantes.

Premier type de dynamique

Supposons :

$$f_o(y_t | y_{t-1}, u_t, u_{t-1}) = f_o(y_t | u_t), \text{ indépendant de } y_{t-1}, u_{t-1};$$

$$f_o(u_t | y_{t-1}, u_{t-1}) = f_o(u_t | u_{t-1}), \text{ indépendant de } y_{t-1}.$$

Le facteur U_t admet une évolution autonome et il n'y a pas de «feedback» de Y vers U . En particulier le processus d'hétérogénéité est (marginale) markovien.

Second type de dynamique

Nous pouvons aussi supposer :

$$f_o(y_t | y_{t-1}, u_t, u_{t-1}) = f_o(y_t | u_t), \text{ indépendant de } y_{t-1}, u_{t-1},$$

$$f_o(u_t | y_{t-1}, u_{t-1}) = f_o(u_t | y_{t-1}), \text{ indépendant de } u_{t-1}.$$

Le processus d'hétérogénéité définit alors une valeur intermédiaire expliquant comment s'effectue la transition de y_{t-1} vers y_t . Dans ce cas, on remarque que $f_o(y_t, u_t | y^{t-1}, u^{t-1}) = f_o(y_t | u_t) f_o(u_t | y_{t-1})$, et, en intégrant par rapport à u_t on voit que le processus Y est (marginale) markovien.

3.2. Objectifs des modèles du premier type

Les modèles du premier type peuvent être utilisés pour introduire une dynamique dans l'hétérogénéité inobservable, prendre en compte des changements de régimes ou construire des modèles à facteurs.

3.2.1. Hétérogénéité dynamique

L'exemple le plus simple de ce type d'utilisation est un modèle de Poisson dans lequel le facteur d'hétérogénéité présente une dépendance réelle. Par exemple, il peut être supposé tel que $\log U_{it} = a + \rho \log U_{i,t-1} + \varepsilon_{it}$, où les erreurs sont gaussiennes indépendantes identiquement distribuées. Une telle spécification autorégressive est par exemple considérée dans Pinquet (2001), Pinquet-Guillen-Bolancé (2001).

3.2.2. Modèles à changements de régimes

La variable U_t est ici un indicateur de divers régimes possibles pour le risque. Elle est qualitative avec des modalités correspondant à chacun des régimes. Des modèles de ce type sont les suivants :

- a) Un modèles autorégressif-moyenne mobile (ARMA) à changement de régimes, ont été initialement introduits pour prendre en compte les cycles. Il y a alors deux alternatives : récession, expansion (voir *e.g.* Hamilton (1989)). Ils sont aussi utilisés pour les calculs de Valeur à Risque (VaR) en finance (voir *e.g.* Billio-Pelizzon (2000)).
- b) Le risque financier étant généralement mesuré par la volatilité, cette approche a été aussi introduite dans les modèles à volatilité stochastique. Il y a alors divers régimes de volatilité, par exemple petite ou grande volatilité [voir Hamilton (1988), pour une application aux taux d'intérêt].
- c) Finalement cette approche a aussi été utilisée pour des risques mesurés par l'intermédiaire du nombre de sinistres. Ainsi un modèle de Poisson à changement de régimes a été introduit pour 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.

i) Simplifier la dynamique

C'est le cas d'un modèle GARCH à facteur défini par :

$$Y_t = aU_t + \varepsilon_t,$$

où Y_t est une mesure multidimensionnelle du risque de dimension K , U_t est un facteur scalaire satisfaisant un modèle GARCH : $U_t = (c + dU_{t-1}^2)^{1/2}\eta_t$, et ε_t , η_t sont deux bruits blancs gaussiens indépendants.

Dans un tel modèle la dynamique des composantes du risque passe par l'intermédiaire d'un seul facteur sous-jacent (inobservable). Ceci conduit habituellement à des évolutions erratiques, mais assez parallèles des composantes de risque Y_{1t}, \dots, Y_{Kt} (voir Diebold-Nerlove (1989), pour une application aux taux de change).

ii) Enrichir la dynamique

Un exemple classique est celui des modèles à volatilité stochastique du type :

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

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

ε_t , η_t sont indépendants gaussiens.

Une approche similaire a été considérée pour des séries temporelles de durées, comme des données entre transaction ou des durées entre sinistres. Un exemple est le modèle SVD (Stochastic Volatility Duration) introduit par Ghysels, Gouriéroux et Jasiak (2003) pour analyser le risque de liquidité. Il est défini par :

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

où le facteur bivarié (F_{1t}, F_{2t}) satisfait un processus Vectoriel Autorégressif Gaussien, de loi marginale $N(0, I_2)$, $G(b, \cdot)$ est la fonction quantile de la loi $\gamma(b, b)$.

3.3. Objectifs des modèles du deuxième type

L'objectif principal de ces modèles est de construire des modèles dynamiques pour des variables à support contraint : variables entières, variables positives.

3.3.1. Processus autorégressif à valeurs entières

L'introduction d'effets autorégressifs pour des séries à valeurs entières n'est pas immédiate. En effet deux idées naturelles sont les suivantes :

- i) On peut écrire une formule autorégressive linéaire assure $Y_{it} = \rho Y_{i,t-1} + \varepsilon_{it}$, avec $-1 \leq \rho \leq 1$, et ε_{it} un ensemble de variables indépendantes de même loi. Cependant le fait que Y soit à valeurs entières implique que l'erreur l'est également et que de plus ρ est nécessairement égal à 0 ou 1. Ceci rend le modèle impraticable.
- ii) On peut aussi introduire directement dans le modèle de regression de Poisson la variable retardée parmi les variables explicatives. La loi de $Y_{i,t}$ conditionnelle à l'information disponible est alors $\mathcal{P}\{\exp(x_t b + y_{i,t-1} c)\}$, disons. Cependant les propriétés dynamiques d'un tel modèle, notamment les conditions de stationarité ou les formules de prévision à l'horizon h sont mal connues.

Pour contourner ces difficultés, il est possible de rendre stochastique le coefficient autorégressif usuel. Le modèle obtenu est appelé INAR (Integer Valued Autoregressive) model. Il est défini de la façon suivante :

$$Y_t = \sum_{j=1}^{Y_{t-1}} U_{jt} + \varepsilon_t$$

où la variable U_{jt} sont indépendantes de même loi de Bernoulli de paramètre p et indépendantes du terme d'erreur ε_t . De plus les variables ε_t sont supposées indépendantes de même loi de Poisson $P(\lambda)$. Cette décomposition peut encore s'écrire :

$$Y_t = U_t + \varepsilon_t,$$

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

Ainsi la composante autorégressive U_t est stochastique de moyenne pY_{t-1} similaire à celle d'une autorégression linéaire. Cependant sa loi assure l'aspect discret de l'espace des valeurs possibles.

Ce modèle est très maniable. Ainsi les moments d'ordre un et deux sont :

$$EY_t = \frac{\lambda}{1-p}, \quad \gamma(h) = \frac{p^h \lambda}{1-p}.$$

De plus le modèle est adapté à la prévision. Ainsi la prévision à l'horizon h est :

$$E(Y_{t+h}|Y_t) = p^h Y_t + \frac{1-p^h}{1-p} \lambda,$$

et la distribution prédictive est facilement caractérisée par sa transformée de Laplace :

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

avec $a(v) = \log[p \exp(v) + 1 - p]$ et $b(v) = \lambda[\exp(v) - 1]$.

Finalement le processus est stationnaire si le paramètre p est strictement inférieur à 1, et sa loi marginale est la loi de Poisson $P\{\lambda/(1-p)\}$. Il s'agit clairement d'une extension du modèle de Poisson usuel, qui correspond au cas $p = 0$ d'absence de liaison temporelle.

3.3.2. Processus à valeurs positives

De façon analogue des modèles autorégressifs non linéaires particuliers doivent être utilisés lorsque les variables endogènes sont à valeurs continues positives. De tels modèles sont particulièrement utiles pour analyser les séries de durées entre sinistres. La spécification la plus classique est le processus autorégressif gamma (ARG), qui constitue la version en temps discret du processus de Cox, Ingersoll, Ross (voir Gouriéroux-Jasiak (2000), Gouriéroux-Monfort (2002)).

Un processus ARG(1) est défini en deux étapes, avec un facteur intermédiaire U_t . On a :

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

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

Ainsi la variable Y_t suit une loi gamma avec un degré de liberté stochastique dont la valeur dépend de Y_{t-1} .

Ce modèle est également simple d'utilisation. Par exemple la fonction d'auto-corrélation $\text{Corr}(Y_{t+h}, Y_t)$ est similaire à celle d'un processus AR(1) de corrélation ρ . Les formules de prévisions sont :

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

La transformée de Laplace conditionnelle est :

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

avec :

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

La loi marginale de Y_t est telle que $\frac{1 - \rho}{c} Y_t$ suit la loi $\gamma(\nu)$.

3.4. Dépendances « réelle » et « apparente »

Nous avons vu que les facteurs sous-jacents peuvent être introduits de diverses façons avec des interprétations différentes. Ils peuvent donc modéliser divers types de dépendance temporelle. Dans ce paragraphe nous nous proposons de distinguer des notions de dépendance réelle et apparente à partir d'un exemple pour le nombre de sinistres. Supposons que ce nombre s'écrive :

$$Y_t = U_{1t} + \varepsilon_t,$$

où, conditionnellement à l'information disponible à la date $t - 1$, les composantes U_{1t} , ε_t sont indépendantes de loi :

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

$$\varepsilon_t \sim \mathcal{P}(\lambda u_2), \text{ Poisson,}$$

avec :

$$U_2 \sim \gamma(a, a).$$

Dans ce modèle la dépendance temporelle est de deux types :

- i) la composante U_{1t} dépend de façon explicite du nombre passé de sinistres (dépendance directe) ;
- ii) la seconde composante ε_t comporte un facteur d'hétérogénéité non observable U_2 qui, après marginalisation, créera une autodépendance temporelle (dépendance apparente).

Ce modèle contient comme cas particuliers :

- i) le modèle classique Poisson-Gamma sans dépendance « réelle », pour $p = 0$;
- ii) le modèle INAR(1), sans dépendance « apparente » induite par l'hétérogénéité, lorsque $a = +\infty$.

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

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

et peut être calculée par des méthodes de Monte Carlo.

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

$$\lambda_{it} = \exp(x'_{it}b_1), \quad p_{it} = \frac{1}{1 + \exp(-x'_{it}b_2)}.$$

3.5. Structure par terme des taux d'intérêt

La détermination de la valeur future d'un portefeuille de contrats d'assurance nécessite non seulement de prévoir les répartitions de sinistres, mais aussi de savoir comment actualiser les flux futurs. Il est donc intéressant d'expliquer comment des modèles à facteurs peuvent aussi être utilisés pour cette actualisation, c'est-à-dire pour analyser la forme et l'évolution de la structure par terme des taux. Rappelons à ce stade que le prix à la date t d'un zéro-coupon de maturité résiduelle h (valant un euro à la date $t+h$), noté $B(t, h)$, peut s'écrire : $B(t, h) = E_t(M_{t,t+1}, \dots, M_{t+h-1,t+h})$, où $M_{t,t+1}$ désigne le facteur d'escompte stochastique pour la période $t, t+1$ et E_t l'espérance conditionnelle à l'information disponible.

Ce facteur d'escompte stochastique doit être positif pour satisfaire les contraintes d'absence d'opportunité d'arbitrage.

Nous allons décrire ci-dessous le modèle à facteurs conduisant aux modèles affines de taux d'intérêt [Duffie, Kan (1996), Gouriéroux, Monfort, Poliménis (2003)].

Dans ce modèle on suppose que le facteur d'escompte stochastique s'écrit comme une fonction exponentielle affine :

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

de deux variables d'états, où r_{t+1} est le taux court et U_t est vecteur latent. De plus on suppose que $X_t = (r_{t+1}, U'_t)$ suit un modèle autorégressif composé d'ordre 1 (CAR(1)), dont la loi conditionnelle est définie par la transformation de Laplace conditionnelle :

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

supposée exponentielle affine dans les états retardés.

Les paramètres du modèle ci-dessus sont cependant contraints. En effet la formule de valorisation doit être valable en particulier pour l'actif sans risque

de maturité 1 et de prix $B(t, t+1) = \exp(-r_{t+1})$. Écrivant cette condition : $\exp(-r_{t+1}) = E_t \exp(\alpha r_{t+2} + \delta' U_{t+1} + \beta_t)$, on voit que le facteur d'escompte peut être réécrit :

$$M_{t,t+1} = \exp(\gamma_0 + \gamma'_1 X_t + \gamma'_2 X_{t+1}),$$

avec $\gamma_0 = -b(\alpha, \delta)$, $\gamma'_1 = -a(\alpha, \delta) - \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\gamma'_2 = \begin{pmatrix} \alpha \\ \delta \end{pmatrix}$.

La spécification est très adaptée pour calculer des formes explicites de structures par terme de taux. En effet on peut montrer que :

$$B(t, h) = \exp(c'_h X_t + d_h),$$

où les coefficients se calculent par récurrence

$$c_h = a \left[c_{h-1} + \begin{pmatrix} \alpha \\ \delta \end{pmatrix} \right] - a \begin{pmatrix} \alpha \\ \delta \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

$$d_h = d_{h-1} - b \begin{pmatrix} \alpha \\ \delta \end{pmatrix} + b \left[c_{h-1} + \begin{pmatrix} \alpha \\ \delta \end{pmatrix} \right],$$

avec la condition initiale $c_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $d_0 = 0$.

En particulier les taux géométriques aux divers termes : $r(t, h) = -\frac{1}{h} \log B(t, h)$ s'écrivent : $r(t, h) = -\frac{c'_h}{h} X_t - d_h$. Ils sont en relation linéaire affine avec les facteurs, d'où le nom du modèle.

3.6. Structure par terme des taux d'intérêt

Nous terminerons en discutant l'inférence statistique pour les modèles de type 1 et 2.

3.6.1. Inférence pour les modèles de type 1

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

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

et donc la forme de la loi jointe :

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

En revanche si on souhaite calculer la vraisemblance :

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

il faut évaluer une intégrale de très grande dimension, impossible à calculer avec précision.

Diverses solutions sont envisageables :

- i) On peut retenir des modèles simples où le calcul peut s'effectuer explicitement ou récursivement. Si le modèle est markovien linéaire le filtre de Kalman permet de résoudre le problème. De même si le modèle est markovien et U_t prend un nombre fini de valeurs, le filtre de Kitagawa-Hamilton (Hamilton (1989)) fournit la solution.
- ii) On peut aussi utiliser des théories structurelles pour reconstruire U_t . Ainsi dans les modèles affines de structure par terme, il existe une relation linéaire entre taux aux divers termes et facteurs sous-jacents. Ces derniers peuvent donc être retrouvés à partir des taux observés, et alors c'est la vraisemblance non marginalisée qui est à utiliser.
- iii) Sinon, on peut soit :
 - remplacer le problème initial par un problème plus simple en changeant de critère d'estimation. Ceci est fait dans l'inférence indirecte (Gouriéroux-Monfort-Renault (1993)) où la vraisemblance est remplacée 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 apparemment plus compliqué, comme dans l'approche bayésienne (voir Billio-Monfort-Robert (1999)).

L'idée générale de cette dernière méthode 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 une méthode de simulation de type MCMC (Monte-Carlo Markov Chain) comme la méthode de Metropolis-Hastings, qui ne nécessite 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 :

$$E_t \{ \exp (w'_j Y_{t+1}) - \exp [a(w_j, \varphi)'Y_t + b(w_j, \varphi)] \} = 0, \quad j = 1, \dots, J.$$

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. Leur mise en application pratique est cependant encore limitée 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.

Références

- 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 Multivariate Switching Regime Approach », *Journal of Empirical Finance*, 135.
- CHIAPPORI P.A. et SALANITÉ 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.
- DIONNE G., GOURIÉROUX C. et VANASSE C. (1998), « Evidence of Adverse Selection in Automobile Insurance Markets », dans *Automobile Insurance*, eds. D. Dionne, C. Laberge-Nadeau, Kluwer, 13-4.
- DIONNE G., GOURIÉROUX C. et VANASSE C. (2001), « Evidence Adverse Selection in Automobile Insurance Market », *Journal of Political Economy*.
- DIONNE G. et VANASSE C. (1992), « Automobile Insurance Ratemaking in the Presence of Asymmetrical Information », *Journal of Applied Econometrics*, 7, 149-165.
- DUFFIE D. et KAN R. (1996), « A Yield Factor Model of Interest Rates », *Mathematical Finance*, 379-406.
- 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.
- GERBER H. et JONES D. (1975), « Credibility Formulas of the Updating Type », *Transactions of the Society of Actuaries*, 27, 21-52.

- GHYSELS E., GOURIÉROUX C. et JASIAK J. (2003), « Stochastic Volatility Duration Models », à paraître *Journal of Econometrics*.
- GOURIÉROUX C. (1999), « Statistique de l'assurance », *Economica*, Paris
- GOURIÉROUX C. et JASIAK J. (2000a), « Autoregressive Gamma Processes », Document CREST.
- GOURIÉROUX C. et JASIAK J. (2000b), « Nonlinear Panel Data Models with Dynamic Heterogeneity », dans Krishuahumar J. et Ronchetti E. eds, *Panel Data Econometrics, Future directions*, 127-147.
- GOURIÉROUX C. et JASIAK J. (200.), « Heterogenous INAR(1) Model with Application to Car Insurance », Document CREST.
- GOURIÉROUX C. et MONFORT A. (1991), « Simulation Based Inference in Models with Heterogeneity », *Annales d'Économie et de Statistique*, n° 20/21, 39-107.
- GOURIÉROUX C. et MONFORT A. (1993), « Simulation Based Inference », *Journal of Econometrics*, 59, 5-33.
- GOURIÉROUX C. et MONFORT A. (1996a), « Simulation Based Econometric Methods », Oxford University Press.
- GOURIÉROUX C. et MONFORT A. (1996b), « Statistique et modèles économétriques », (2 volumes) *Economica*.
- GOURIÉROUX C. et MONFORT A. (1997), « Modèles de comptage semi-paramétriques », dans *Econometrie Appliquée*, *Economica*, Presses des HEC de Montréal.
- GOURIÉROUX C. et MONFORT A. (2002), « Affine Term Structure Models », Document CREST.
- GOURIÉROUX C., MONFORT A. et 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., MONFORT A. et TROGNON A. (1984a), « Pseudo Maximum Likelihood Methods Theory », *Econometrica*, 52, 681-700.
- GOURIÉROUX C., MONFORT A. et TROGNON A. (1984b), « Pseudo Maximum Likelihood Methods Applications to Poisson Models », *Econometrica*, 52, 701-720.
- GREENE W. (1997), « Econometric Analysis », Rentice Hall.
- 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.
- LEMAIRE J. (1995), « Bonus-Malus System in Automobile Insurance », Kluwer.
- PINQUET J. (2000), « Experience Rating Through Heterogenous Models », *Handbook of Insurance*, ed. G. Dionne, Chap. 14, 459-500, Kluwer.
- PINQUET J., (2003), « Prise en compte de l'ancienneté des périodes dans les modèles de risque en fréquence : aspects théoriques et applications à la tarification des risques en assurance automobile », ce numéro.
- PINQUET J. (2001), « Experience Rating Through Heterogeneous Models », dans *Handbook of Insurance*, Chapitre 15, Kluwer.
- PINQUET J., GUILLEN M. et BOLANCÉ C. (2001), « Allowance for the Age of Claim in Bonus-Malus System », *ASTIN Bulletin*, 31, n°2, 337-348.

VARIABLES LATENTES ET MODELISATION STATISTIQUE EN ASSURANCE

- PUELZ R. et SNOW A. (1994), « Evidence en Adverse Selections : Equilibrium Signalling and Cross Solidization in the Insurance Market », *Journal of Political Economy*.
- WEDEL M., DESARBO W.S., BULT J.R et RAMASWAMY V. (1993), « A Latent Class Poisson Regression Model for Heterogenous Count Data », *Journal of Applied Econometrics*, 8, 397-411. 33