Assurances et gestion des risques
Insurance and Risk Management

Crédibilité linéaire bivariée utilisant le nombre de périodes avec réclamations : modèles de Poisson, modèles à barrière et modèles gonflés à zéro

Jean-Philippe Boucher et Michel Denuit

Volume 75, numéro 4, 2008

URI : https://id.erudit.org/iderudit/1106754ar
DOI : https://doi.org/10.7202/1106754ar

Aller au sommaire du numéro

Éditeur(s)

Faculté des sciences de l'administration, Université Laval

ISSN

1705-7299 (imprimé)
2371-4913 (numérique)

Découvrir la revue

Citer cet article

Boucher, J.-P. & Denuit, M. (2008). Crédibilité linéaire bivariée utilisant le nombre de périodes avec réclamations : modèles de Poisson, modèles à barrière et modèles gonflés à zéro. *Assurances et gestion des risques / Insurance and Risk Management*, *75*(4), 487–520. https://doi.org/10.7202/1106754ar

Résumé de l'article

Dans certaines situations, il est bien connu que la crédibilité linéaire univariée est trop restrictive et ne modélise pas bien les primes prédictives. Récemment, Boucher, Denuit et Guillén (2006a, 2006b) ont proposé diverses généralisations du modèle de Poisson gonflé à zéro et du modèle à barrière. Pour ces modèles, les auteurs ont montré que la distribution prédictive du nombre de réclamations annuelles ne dépend pas seulement du nombre de réclamations passées, mais aussi du nombre de périodes au cours desquelles l'assuré a fait une réclamation. Nous utilisons la théorie de la crédibilité bivariée afin d'estimer la prime future en fonction de plusieurs prédicteurs afin de comparer l'ajustement des primes de crédibilité linéaire bivariées aux primes exactes des nouveaux modèles. Nous montrons que l'ajustement pour les modèles gonflés à zéro est satisfaisant alors que celui des modèles à barrière génère des erreurs d'approximation majeures, probablement dues au fait que le modèle de crédibilité linéaire classique néglige la dépendance des effets aléatoires.

Assurances et gestion des risques, vol. 75(4), janvier 2008, 487-520
Insurance and Risk Management, vol. 75(4), January 2008, 487-520

Crédibilité linéaire bivariée utilisant le nombre de périodes avec réclamations : modèles de Poisson, modèles à barrière et modèles gonflés à zéro

par Jean-Philippe Boucher et Michel Denuit

RÉSUMÉ

Dans certaines situations, il est bien connu que la crédibilité linéaire univariée est trop restrictive et ne modélise pas bien les primes prédictives. Récemment, Boucher, Denuit et Guillén (2006a, 2006b) ont proposé diverses généralisations du modèle de Poisson gonflé à zéro et du modèle à barrière. Pour ces modèles, les auteurs ont montré que la distribution prédictive du nombre de réclamations annuelles ne dépend pas seulement du nombre de réclamations passées, mais aussi du nombre de périodes au cours desquelles l'assuré a fait une réclamation. Nous utilisons la théorie de la crédibilité bivariée afin d'estimer la prime future en fonction de plusieurs prédicteurs afin de comparer l'ajustement des primes de crédibilité linéaire bivariées aux primes exactes des nouveaux modèles. Nous montrons que l'ajustement pour les modèles gonflés à zéro est satisfaisant alors que celui des modèles à barrière génère des erreurs d'approximation majeures, probablement dues au fait que le modèle de crédibilité linéaire classique néglige la dépendance des effets aléatoires.

Mots clés : Nombre de réclamations, données de panel, modèle gonflé à zéro, modèle à barrière, crédibilité linéaire bivariée.

Les auteurs :

Jean-Philippe Boucher est professeur au Département de Mathématiques, Université du Québec à Montréal, 201, avenue du Président-Kennedy, Montréal, Québec, Canada, H2X 3Y7. Courriel : boucher.jean-philippe@uqam.ca

Michel Denuit est professeur à l'Institut des Sciences Actuarielles, Université Catholique de Louvain, 6 rue des Wallons, B-1348, Louvain-la-Neuve, Belgique. Courriel : michel.denuit@uclouvain.be

ABSTRACT

In some situations, it is well-known that the linear credibility approximation is too restrictive and does not model correctly predictive premiums. Recently, Boucher, Denuit and Guillén (2006a, 2006b) proposed various generalisations of the zero-inflated Poisson and the hurdle distributions. For these models, the authors showed that the predictive premium do not depend exclusively on the number of past claims, but also on the number of insured periods with at least one claim. We use the bivariate linear credibility theory to estimate the future premiums based on these two statistics, and compare the approximation with the exact result using Bayesian theory. We show that the approximations for the zero-inflated models are satisfying. However, we also show that the approximations of the hurdle distributions cause major errors. We explain these results by showing that the linear credibility premiums do not suppose dependence between random effects of the models.

Keywords: Claims number, panel data, zero-inflated models, hurdle models, bivariate linear credibility.

1. INTRODUCTION

Plusieurs caractéristiques d’un assuré ne peuvent pas être utilisées dans la tarification, soit parce qu’elles ne sont pas mesurables, soit parce qu’elles sont difficilement justifiables socialement. Néanmoins, certaines de ces variables ont un impact sur la fréquence de réclamations d’un assuré (nous n’avons qu’à penser à l’agressivité au volant, aux réflexes, au salaire annuel ou à la fréquence de conduite sous l’effet de l’alcool...). Afin de modéliser ces caractéristiques inconnues de chaque assuré, la modélisation bayésienne empirique est une alternative intéressante. En effet, un terme aléatoire ajouté à la distribution de fréquence de sinistres peut ainsi être mis à jour annuellement selon l’expérience de réclamation de l’assuré. L’analyse *a posteriori* de ce terme aléatoire permet d’ajuster la prime annuelle de l’assuré, révelant ainsi partiellement l’effet des caractéristiques non observées sur la fréquence de réclamations.

Habituellement, pour modéliser le nombre de réclamations rapportées à l’assureur, les actuaires ont tendance à utiliser la loi de Poisson, couplée à une distribution gamma pour modéliser l’effet aléatoire. Récemment, afin de travailler avec des données de panel, Boucher, Denuit et Guillén (2006b) ont proposé diverses généralisations du modèle de Poisson gonflé à zéro (*zero-inflated Poisson model*) alors que Boucher, Denuit et Guillén (2006a) ont généralisé de diverses façons un modèle à barrière (*hurdle model*). Pour le modèle gonflé à zéro, un seul facteur d’hétérogénéité a été ajouté à la distribution alors que pour le modèle à barrière, deux effets aléatoires ont été ajoutés pour chaque processus de la distribution.

Parce qu'il arrive fréquemment que le développement des distributions *a posteriori* et prédictives soient complexes, les actuaires utilisent la théorie de la crédibilité linéaire pour estimer les primes prédictives. Puisque pour ces nouveaux modèles, la distribution prédictive du nombre de réclamations ne dépend pas seulement du nombre de réclamations passées, mais aussi du nombre de périodes au cours desquelles l'assuré a fait une réclamation, nous utilisons aussi la théorie de la crédibilité bivariée afin d'estimer la prime future en fonction de plusieurs prédicateurs. Dans certaines situations, il est bien connu que la crédibilité linéaire univariée est trop restrictive et ne modélise pas bien les primes prédictives. En conséquence, le but de ce papier est de comparer l'ajustement des primes de crédibilité linéaire bivariées avec les primes exactes. Nous montrons que l'ajustement pour les modèles gonflés à zéro est satisfaisant alors que celui des modèles à barrière génère des erreurs d'approximation majeures, probablement dues au fait que le modèle de crédibilité linéaire classique néglige la dépendance des effets aléatoires.

Cet article est construit de la manière suivante. À la section 2, nous effectuons une revue des modèles gonflés à zéro et des modèles à barrière pour données de panel. Les propriétés des modèles sont explorées et différentes interprétations des modèle sont données dans le cas de diverses applications actuarielles. À la troisième section, nous explorons les primes prédictives des modèles ainsi que la théorie de la crédibilité linéaire. Nous y développons la théorie de la crédibilité bivariée utilisant le nombre de périodes d'assurance avec réclamation comme prédicteur. La section suivante illustre nos résultats numériquement en comparant les résultats obtenus par Boucher, Denuit et Guillén (2006a, 2006b) à ceux résultant de nos modèles de crédibilité linéaire. Alors que la dernière section conclut, une annexe en fin de papier développe plus en détail certains résultats mathématiques.

1.1. Hypothèses et notations

Tout comme Dionne et Vanasse (1989, 1992), nous exprimons les caractéristiques des assurés par une fonction de score $h(\beta_0 + x_i' \beta)$, où β_0 représente l'ordonnée à l'origine et $\beta' = (\beta_1, ..., \beta_p)$ est un vecteur de paramètres pour les variables explicatives $x_i = (x_{i,1}, ..., x_{i,p})$. En effet, puisque la majorité des compagnies d'assurance travaillent avec des variables catégorielles, les caractéristiques d'un assuré sont exprimées par des variables indicatrices prenant les valeurs 0 ou 1.

Parce que l'analyse des primes prédictives utilise un historique de perte, nous le noterons $\text{II}_{i,T}$ car de nombreuses informations peuvent y être emmagasinées. Cette notation représente toutes les informations recueillies au temps T, c'est-à-dire $n_{i,1}$, ..., $n_{i,T}$, le nombre de réclamations passées et $k_{i,1}$, ..., $k_{i,T}$, une variable indiquant si un assuré i a réclamé au moins une fois pendant la période d'assurance t. Notons aussi la variable $K_{i\bullet} = \sum_{t=1}^{T} k_{i,t} = T - T_0$, où T_0 peut être vu comme le nombre de périodes d'assurance où aucune réclamation n'a été effectuée. De même, $N_{i\bullet} = \sum_{t=1}^{T} n_{i,t}$ représente le nombre total de réclamations pour l'assuré i, du temps $t = 1$ à $t = T$. Les moyennes sur T observations de ces deux variables se notent $\bar{K}_i = K_{i\bullet} / T$ et $\bar{N}_i = N_{i\bullet} / T$. Conditionnellement à un effet alatoire θ, la moyenne de la variable aléatoire $N_{i,t}$ est exprimées comme $\mu_{i,t}(\theta) = E[N_{i,t} | \theta]$, alors que la valeur espérée du nombre de sinistres rapportés est notée par $\mu_{i,t} = E[E[N_{i,t} | \theta]]$.

La généralisation des modèles gonflés à zéro (Boucher, Denuit et Guillén, 2006a) et des modèles à barrière (Boucher, Denuit et Guillén, 2006b) est basée sur l'hypothèse que les régresseurs x_i sont indépendants du temps, signifiant ainsi qu'ils sont identiques peu importe la période d'assurance t. Comme indiqué dans les papiers où sont présentées ces généralisations des modèles, cette hypothèse n'est pas aussi restrictive que cela peut laisser paraître. En effet, en pratique, plusieurs caractéristiques du risque ne changent pas dans le temps, certaines autres qui peuvent changer n'évoluent pas de manière aléatoire alors que les changements majeurs de caractéristiques amènent fréquemment le besoin de créer une nouvelle police. Afin de travailler avec les résultats obtenus par Boucher, Denuit et Guillén (2006a, 2006b), nous gardons cette hypothèse pour ce papier, mais une généralisation de cette hypothèse est assez directe, moyennant quelques petites transformations simples.

2. MODÈLES POUR DONNÉES DE PANEL

Cette section expose les 3 principaux modèles étudiés dans ce papier, tout en indiquant en quoi ils se différencient et comment ils peuvent être interprétés pour les données d'assurance. Afin de faciliter la lecture, les différents moments de ces modèles ne sont indiqués qu'en annexe.

2.1. Modèles de Poisson

Pour les données de panel, l'un des modèles de base pour modéliser des données de comptage est la distribution de Poisson avec effets aléatoires. Le facteur d'hétérogénéité θ_i ajouté au paramètre de moyenne peut être distribué selon une loi gamma générant une distribution jointe fermée pour le nombre de réclamations pour les périodes $t = 1, ..., T$. En effet, en supposant que le loi gamma est de moyenne 1 et de variance α, la distribution peut être exprimée comme (Hausman, Hall et Griliches, 1984)

$$\Pr[N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}]$$
$$= \left(\prod_{t=1}^{T} \frac{\lambda_i^{n_{i,t}}}{n_{i,t}!} \right) \frac{\Gamma(N_{i\bullet} + 1/\alpha)}{\Gamma(1/\alpha)} \left(\frac{1/\alpha}{T\lambda_i + 1/\alpha} \right)^{1/\alpha} (T\lambda_i + 1/\alpha)^{-N_{i\bullet}} \quad (2.1)$$

où, tel que décrit plus tôt, les variables explicatives sont incluses dans le modèle par la fonction $\lambda_i = \exp(x_i'\beta)$. Cette distribution, qui est très populaire dans de nombreux domaines statistiques (voir le chapitre 36 de Johnson, Kotz et Balakrishnan, 1996, pour un résumé), est connue sous les noms binomiale négative bivariée (BNMV) ou multinomiale négative. À noter, évidemment, que d'autres distributions peuvent être obtenues en choisissant différentes distributions des effets aléatoires, les plus connues étant la Poisson-inverse gaussienne ou la Poisson-lognormale (voir Boucher et Denuit, 2006, pour une application en assurance).

2.2. Modèles gonflés à zéro

Parce qu'il existe un grand nombre de valeurs à zéro dans l'analyse du nombre de réclamations à l'assureur, Boucher, Denuit et Guillén (2006b) ont eu l'idée de généraliser le modèle de Poisson gonflé à zéro (*Zero-Inflated Poisson*) pour données de panel. Les auteurs ont traité le terme supplémentaire à zéro ϕ_i de manière déterministe, tout en considérant un terme d'hétérogénéité θ_i dans le paramètre de moyenne de la loi de Poisson. En conditionnant sur ce θ_i et en utilisant la décomposition du binôme de Newton, la distribution peut s'exprimer comme :

$$\Pr[N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T} | \theta_i]$$
$$= \prod_{t=1}^{T} (I_{(n_{i,t}=0)}\phi_i + (1-\phi_i)\Pr[N_{i,t} = n_{i,T} | \theta_i])$$
$$= \sum_{j=0}^{T_0} \binom{T_0}{j} V_j^{P_{oi}}(N_{i\bullet} | \theta_i)\, \phi_i^{T_0 - j} (1-\phi_i)^{(T-T_0)+j}, \quad (2.2)$$

où la fonction $V^{Poi}(.)$ a la forme de la distribution de probabilité d'une loi de Poisson, telle qu'exprimée par l'équation suivante :

$$V_j^{P_{oi}}(N_{i\bullet}|\theta_i) = \frac{(\lambda_i\theta_i)^{N_{i\bullet}} \exp(-(T-T_0+j)\,\lambda_i\theta_i)}{\prod_t^T n_{i,t}!}. \quad (2.3)$$

Dans la cas où cet effet aléatoire est distribué selon une loi gamma de moyenne 1 et variance α, la distribution jointe s'exprime :

$$\Pr[N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}]$$

$$= \sum_{j=0}^{T_0} \binom{T_0}{j} V_j^{NB}(N_{i\bullet})\, \phi_i^{T_0-j}(1-\phi_i)^{(T-T_0)+j} \quad (2.4)$$

où le paramètre $\phi_i = \exp(y_i\,\gamma) / (1 + \exp(y_i\,\gamma))$ est modélisé avec des régresseurs y_i. Ces derniers représentent un sous-ensemble des régresseurs x_i, alors que γ est un vecteur de paramètres pour ces variables explicatives y_i. La fonction $V_j^{NB}(.)$, quant à elle, est exprimée comme suit :

$$V_j^{NB}(N_{i\bullet})$$

$$= \frac{\Gamma(N_{i\bullet}+1/\alpha)}{\Gamma(1/\alpha)\prod_t^T n_{i,t}!}\left(\frac{1/\alpha}{(T-T_0+j)\lambda_i+1/\alpha}\right)^{1/\alpha}\left(\frac{\lambda_i}{(T-T_0+j)\,\lambda_i+1/\alpha}\right)^{N_{i\bullet}}. \quad (2.5)$$

Cette distribution a été appelée *Multivariate Zero-Inflated Poisson gamma*, qui pourrait s'exprimer en français comme *modèle multivarié Poisson-gamma gonflé à zéro* (MP0-gamma).

Au lieu de généraliser une distribution Poisson gonflée à zéro et d'y ajouter des effets aléatoires, Boucher, Denuit et Guillén (2006a) ont utilisé la distribution de Poisson et y ont ajouté une hétérogénéité dégénérée. Formellement, cette distribution des effets aléatoires peut être exprimée comme suit :

$$g(\theta_i|\phi_i,\alpha) = \begin{cases} \phi_i & \text{pour } \theta_i = 0 \\ (1-\phi_i)\,f(\theta_i) & \text{pour } \theta_i \geq 0 \end{cases}, \quad (2.6)$$

où ϕ_i est un paramètre, différent de celui de l'équation (2.4), représentant une masse de point modélisé comme $\dfrac{\exp(y_i'\gamma)}{1+\exp(y_i'\gamma)}$ et $f(.)$ est une distribution d'effets aléatoire classique de moyenne 1 et de variance α. En supposant que celle-ci est gamma, les auteurs en ont déduit la distribution jointe suivante :

$$\Pr[N_{i,1}=n_{i,1},\ldots,N_{i,T}=n_{i,T}]$$

$$=I_{(T_0=T)}\phi_i+(1-\phi_i)\frac{\Gamma(N_{i\bullet}+1/\alpha)}{\Gamma(1/\alpha)\prod_{t=1}^{T}n_{i,t}!}\left(\frac{1/\alpha}{T\lambda_i+1/\alpha}\right)^{1/\alpha}\left(\frac{\lambda_i}{T\lambda_i+1/\alpha}\right)^{N_{i\bullet}}. \quad (2.7)$$

En utilisant le terme anglais exprimant cette distribution (*zero-inflated multivariate negative binomial*), ce modèle pourrait se traduire par *modèle gonflé à zéro de la binomiale négative multivariée* (0-BNMV).

2.3. Modèles à barrière

L'intérêt du modèle à barrière est basé sur le fait que la majorité des assurés rapporte moins de deux sinistres par années en assurance. En conséquence, une classification des assurés basée sur une distribution construite avec deux processus est intéressante. Premièrement, une distribution dichotomique différenciant les assurés avec ou sans réclamation alors que le second processus précise le nombre total de réclamations, conditionnellement au *succès* du premier processus.

La construction de la généralisation du modèle à barrière pour données de panel s'effectue en considérant deux distributions de probabilité $f_1(\cdot \mid \theta_{i,1})$ et $f_2(\cdot \mid \theta_{i,2})$ ayant des supports respectifs de {0, 1} et {0, 1, ...}. Une dépendance temporelle entre les contrats d'un même assuré est supposé lorsque des effets aléatoires individuels $\theta_{i,1}$ et $\theta_{i,2}$, pouvant s'identifier globalement par Θ_i, sont ajoutés à chaque processus. Ainsi, Boucher, Denuit et Guillén (2006a) ont choisi la distribution conditionnelle suivante pour $K_{i,t}$:

$$K_{i,t} \mid \Theta_i \sim Bernoulli\ (\theta_{i,1}), \quad i=1,\ldots,n, \quad t=1,\ldots,T.$$

et, dans le cas où $K_{i,t}=1$, la distribution conditionnelle suivante pour $N^*_{i,t}$:

$$N^*_{i,t} \mid \Theta_i, K_{i,t} \sim Poisson\ (\gamma_i\,\theta_{i,2}) \quad i=1,\ldots,n, \quad t=1,\ldots,T.$$

où pour $K_{i,t}=1$, les auteurs ont utilisé $N^*_{i,t}=N_{i,t}-1$, représentant une version translatée du nombre de réclamations effectuées, afin de simplifier les développements mathématiques et revenir à une expression connue de l'analyse des données de comptage. À noter que la variable $N^*_{i,t}$ est inutile dans le cas où $K_{i,t}=0$. L'introduction d'une fonction de score dans le paramètre $\gamma_i=\exp\,(y_i'\,\mu)$ est possible, ce qui permet d'inclure des caractéristiques du risque dans le deuxième processus. Conséquemment, la distribution jointe du nombre de réclamations de l'assuré i pour les périodes $t=1,\ldots,T$ peut s'exprimer comme suit :

$$\Pr[N_{i,1}=n_{i,1},...,N_{i,T}=n_{i,T}]$$

$$=\int\int\prod_{t=1}^{T}\theta_{i,1}^{k_{i,t}}(1-\theta_{i,1})^{1-k_{i,t}}\left(e^{-\lambda_i\theta_{i,2}}\frac{(\lambda_i\theta_{i,2})^{n_{i,t}^*}}{n_{i,t}^*!}\right)^{k_{i,t}}g(\theta_{i,1},\theta_{i,2})\,d\theta_{i,1}\,d\theta_{i,2} \quad (2.8)$$

où $g(\theta_{i,1}, \theta_{i,2})$ est la distribution jointe des effets aléatoires. L'utilisation d'une copule pour modéliser la dépendance entre les deux variables aléatoires est appropriée puisque son choix est indépendant des distributions marginales de chacun des effets aléatoires. Bien que d'autres choix soient possibles, Boucher, Denuit et Guillén (2006a) ont utilisé les distributions suivantes pour Θ_i :

$\theta_{i,1} \sim Beta\ (a_i, b), \quad i = 1, ..., n$

$\theta_{i,2} \sim Gamma\ (1 / \alpha, 1 / \alpha), \quad i = 1, ..., n$

où le paramètre $a_i = \exp(x_i\beta)$, ce qui permet d'inclure les caractéristiques de l'assuré dans le premier processus. Tout comme les modèles BNMV et gonflés à zéro, les paramètres de la loi gamma, quant à eux, sont sélectionnés afin que la moyenne de la distribution gamma soit égale à 1, ce qui ne biaise pas le paramètre de moyenne de la loi de Poisson.

2.3.1. Distributions jointes des effets aléatoires

Comme noté auparavant, la distribution jointe des effets aléatoires peut être déterminée par une copule, qui est un outil très puissant pour modéliser la dépendance (voir Nelsen, 1999, pour une revue). Ainsi, dans cette section, nous allons revoir les trois copules que Boucher, Denuit et Guillén (2006a) ont utilisées. Il est important de comprendre que l'utilisation d'un seul effet aléatoire, que ce soit pour le premier ou le second processus, ne peut pas être satisfaisant dans le contexte des assurances. En effet, l'usage exclusif d'un effet aléatoire sur la loi Bernoulli créerait une situation où la modification annuelle de la prime serait la même pour tous les assurés ayant réclamé au moins une fois, et ce, peu importe le nombre total de réclamations. De manière similaire, si les effets aléatoires n'étaient utilisés que pour le second processus, aucun rabais ne serait accordé aux assurés ne réclamant pas, alors qu'il serait accordé une réduction de primes à un assuré qui ne réclamerait qu'une seule fois![1]

En conséquence, il est évident qu'une analyse du modèle à barrière doit s'effectuer à l'aide de deux effets aléatoires, ceux-ci étant liés par une copule :

- La première copule utilisée est la copule d'indépendance, qui s'exprime simplement comme le produit des fonctions de densité :

$$g(\theta_{i,1}, \theta_{i,2}) = f_1(\theta_{i,1})\, f_2(\theta_{i,2})\ . \tag{2.9}$$

  L'utilisation de cette copule est davantage reliée à sa simplicité pour les calculs qu'à son interprétation car il ne semble pas raisonnable de croire à des effets aléatoires indépendants. En effet, il est intuitif de croire qu'un assuré qui réclame dans plusieurs périodes d'assurance aura sans doute plus de chance de réclamer souvent dans la même période. Toutefois, encore plus important, puisque l'introduction des effets aléatoires a été justifiée pour modéliser les caractéristiques inconnues du risque, il est évident que certains paramètres cachés touchent autant le premier processus que le second, créant ainsi une certaine dépendance.

- Pour modéliser cette dépendance, une copule gaussienne a été testée, générant ainsi la distribution jointe suivante pour les effets aléatoires :

$$g(\theta_{i,1}, \theta_{i,2}) = c^{Ga}(u, v)\, f_1(\theta_{i,1})\, f_2(\theta_{i,2}) \tag{2.10}$$

  La densité de la copule gaussienne s'exprime comme suit :

$$c^{Ga}(u,v) = \frac{1}{\sqrt{1-\rho^2 \exp}} \exp\left(-\frac{1}{2}\left(\frac{\rho^2\Phi^{-1}(u)^2 + \rho^2\Phi^{-1}(v)^2 - 2\rho\,\Phi^{-1}(u)\,\Phi^{-1}(v)}{1-\rho^2}\right)\right) \tag{2.11}$$

  où Φ est la fonction de distribution d'une loi normale(0, 1), $u = F_1(\theta_{i,1})$ et $v = F_2(\theta_{i,2})$.

- La dernière copule qui a été testée est la copule de Fréchet-Hoeffding. Elle correspond à la borne supérieure de toutes les copules, représentant la dépendance positive parfaite entre deux variables aléatoires, appelée comonoticité. En fait, l'estimation d'un modèle avec ce type de copule est simple puisqu'un seul effet aléatoire $\in$ est utilisé. En effet, en utilisant la transformation $u = \Phi(\in)$, cette copule génère des valeurs de la loi beta et de la loi gamma par $\theta_1 = F_1^{-1}(u)$ et $\theta_2 = F_2^{-1}(u)$, où $F_1(u)$ est la fonction de répartition de la loi beta (a_i, b) et $F_2(u)$ celle d'une gamma $(1 / \alpha, 1 / \alpha)$. Certaines propriétés de cette copule peuvent être obtenues en utilisant la copule gaussienne avec $\rho = 1$.

2.4. Comparaisons et interprétations des modèles

Les modèles gonflés à zéro et les modèles à barrière peuvent expliquer de manière intéressante les données d'assurance. Dans le cas du modèle gonflé à zéro MP0-gamma, le paramètre supplémentaire à zéro peut être interprété comme l'effet de la soif du bonus (Lemaire,1995) en assurance. Cette soif du bonus réfère à la situation où les assurés ne déclarent pas tous leurs accidents automobiles de peur que l'augmentation de leur prime annuelle soit plus haute que ce qu'ils ont demandé comme remboursement. Ainsi, l'idée du modèle gonflé à zéro est de distinguer le nombre de réclamations du nombre d'accidents par le paramètre supplémentaire ϕ_i. À ce sujet, Boucher et Denuit (2007) expliquent cette relation entre ces deux statistiques et les conséquences d'un changement de tarification sur la distribution du nombre de réclamations.

Le second modèle à zéro (0-BNMV) peut être vu comme une approximation du modèle MP0-gamma, où le paramètre supplémentaire ϕ_i représente une partie des assurés qui ne réclament jamais. Pour la même raison que le celle décrite dans le modèle précédent, il peut exister un type d'assurés qui n'osent pas réclamer à leur assureur, attendant probablement qu'un accident plus important justifie l'aide de l'assureur.

De l'autre côté, la justification de l'utilisation du modèle à barrière se situe au niveau de la distribution des données. En effet, puisque la très grande majorité (près de 99 %) des assurés ne réclament pas ou qu'une fois par année, il est pertinent de se baser plus profondément sur la modélisation des cas 0 et 1. Bien que le modèle à barrière ne soit pas d'origine actuarielle, il existe de nombreuses applications directes de ce modèle pour les données d'assurance. Ainsi, sa généralisation pour données de panel peut être pertinente pour l'analyse du nombre de couvertures d'assurance touchées ou le nombre de personnes blessées. De plus, cette modélisation du nombre de réclamations ressemble à l'analyse double effectuée pour le montant des réclamations, où les coûts limités et les coûts supérieurs à une certaine limite sont étudiés séparémment.

Néanmoins, le modèle à barrière a une interprétation beaucoup plus naturelle. En effet, afin de conserver leur bon dossier de conduite (bonus), plusieurs conducteurs préfèrent ne pas réclamer tous leurs sinistres. Ainsi, lorsqu'une réclamation a déjà été effectuée, il pourrait être logique de croire que le comportement des assurés n'est pas le même puisque leur bonus pour l'année en cours est déjà perdu.

Les deux modèles présentés dans cet article semblent particulièrement adaptés à la construction des systèmes de Bonus-Malus. En effet, pour illustration, même s'ils sont différents d'un pays à l'autre, nous pouvons rappeler ici grossièrement la construction du Bonus-Malus français qui illustre bien la situation :[2]

$$
\begin{aligned}
BM_t &= (1+\delta_t)^{N_\bullet}(1-\in_t)^{T-K_\bullet} \\
&= (1+\delta_t)^{N^*_\bullet}\left(\frac{1+\delta_t}{1-\in_t}\right)^{K_\bullet}(1-\in_t)^T \; .
\end{aligned}
$$

Nous pouvons remarquer que la variable $K_\bullet$ joue un rôle important dans la modélisation, ce qui correspond bien aux modèles gonflés à zéro et à barrière puisqu'il s'agit d'une des statistiques exhaustives des deux modèles. Il pourrait être intéressant d'utiliser un système de Bonus-Malus avec ce genre de distribution afin de vérifier si l'ajustement s'effectue plus facilement. Ce genre de recherches n'a pas encore été effectué et, de plus, il pourrait être intéressant de voir l'impact sur un tel système, en comparaison avec la loi de Poisson avec effets aléatoires.

3. DISTRIBUTION PRÉDICTIVE

À chaque période d'assurance, le ou les effets aléatoires, exprimés par le terme Θ_i peuvent être mis à jour selon l'historique de perte, révélant ainsi partiellement l'effet des caractéristiques non observées sur la fréquence de réclamations. Formellement, la distribution prédictive peut être trouvée en utilisant le développement suivant :

$$
\begin{aligned}
&\Pr[N_{i,T+1} = n_{i,T+1} \,|\, \Pi_{i,T}] \\
&= \int \Pr[N_{i,T+1} = n_{i,T+1} \,|\, \Theta_i] \left(\frac{\left[\prod_{t=1}^{T} \Pr[N_{i,t} = n_{i,T+1} \,|\, \Theta_i]\right] g(\Theta_i)}{\int \left[\prod_{t=1}^{T} \Pr[N_{i,t} = n_{i,t} \,|\, \Theta_i]\right] g(\Theta_i)\, d\Theta_i} \right) d\Theta_i \\
&= \int \Pr[N_{i,T+1} = n_{i,T+1} \,|\, \Theta_i]\, g(\Theta_i \,|\, \Pi_{i,T})\, d\Theta_i \; ,
\end{aligned}
\tag{3.1}
$$

où $g(\Theta_i \,|\, \Pi_{i,T})$ est la distribution (jointe) *a posteriori* des effets aléatoires Θ_i.

3.1. Prédicateur optimal quadratique

Les distributions *a posteriori* et *prédictive* utilisent la théorie bayésienne et sont fortement liées à la théorie de la crédibilité. (Bülhmann, 1967; Bülhmann et Straub, 1970; Hachemeister, 1975; Jewell, 1975). La correction *a posteriori* appliquée sur la prime de l'assuré est basée sur une distribution de perte spécifique. Le choix le plus populaire pour la fonction de perte est l'erreur quadratique. Pour cette situation, le minimum de :

$$E\,[(\mu_{i,T+1}(\theta) - \psi\,(\Pi_{i,T}))^2] \tag{3.2}$$

pour toutes les fonctions mesurables $\psi : \mathbb{R}^T \to \mathbb{R}$ est obtenu pour :

$$\psi^*(\Pi_{i,T}) = \psi^*_{i,T} = E\,[N_{i,T+1} \mid \Pi_{i,T}] \tag{3.3}$$

Ce qui peut être prouvé relativement facilement. Evidemment, d'autres fonctions de pertes peuvent être envisagées. La différence absolue est une alternative bien connue, de laquelle il peut être prouvé que la médiane atteint le minimum de l'équation de perte. La fonction de perte développée par Young (1998) où un terme de deuxième ordre est ajouté à la fonction de perte quadratique, ou encore le modèle de Young et DeVylder (2000) qui a favorisé l'ajout d'un terme qui encourage l'estimateur de crédibilité à être près d'une constante pour les queues de distributions sont d'autres possibilités. Parce que la distribution de perte quadratique pénalise autant les surcharges que les sous-charges, la distribution de perte exponentielle est une alternative intéressante car elle permet de corriger ce potentiel défaut de l'erreur au carré (Ferreira, 1977; Denuit et Dhaene, 2001 ou encore Boucher et Denuit, 2007, qui l'utilise pour le modèle MP0-gamma). Malgré tout, dans ce papier, nous focalisons sur cette perte quadratique, mais le développement avec d'autres fonctions de perte pourrait être intéressant.

3.1.1. Modèles de Poisson

Il est bien connu que la distribution *a posteriori* d'un modèle de Poisson avec effets aléatoires gamma$(1 / \alpha, 1 / \alpha)$ est aussi gamma avec paramètres $T\,\lambda_i + 1 / \alpha$ et $N_{i\bullet} + 1 / \alpha$. En conséquence, l'espérance prédictive d'un tel modèle est égal à :

$$E\,[N_{i,T+1} \mid N_{i,1}, \ldots, N_{i,T}] = \lambda_i \, \frac{N_{i\bullet} + 1/\alpha}{T\lambda_i + 1/\alpha}\,. \tag{3.4}$$

À noter que ce moment ne dépend que de la somme du nombre de réclamations passées. De plus, lorsque le nombre de périodes d'expérience T tend vers l'infini, la prime prédictive de l'assuré i converge vers son nombre moyen de réclamations annuelles.

3.1.2. Modèles gonflés à zéro

Dans l'estimation de la prime prédictive pour les modèles gonflés à zéro, la construction suivant l'équation (3.1) est aussi effectué, à l'aide de la distribution de θ_i. Ainsi, selon le modèle, Boucher, Denuit et Guillén (2006b) ont montré les diverses primes prédictives de ces modèles.

- Distribution *MP0-gamma*

$$E\left[N_{i,T+1} \middle| n_{i,1},\dots,n_{i,T}\right] = \lambda_i \sum_{j=0}^{T_0} \frac{(N_{i\bullet} + 1/\alpha)\, H(j)}{(T+1-T_0+j)\,\lambda_i + 1/\alpha} \tag{3.5}$$

avec :

$$H(j) = \frac{\binom{T_0}{j} \phi_i^{T_0-j} (1-\phi_i)^{(T+1-T_0)+j} \left((T-T_0+j)\,\lambda_i + 1/\alpha\right)^{-\left(\sum_t^T n_t + 1/\alpha\right)}}{\sum_{k=0}^{T_0} \binom{T_0}{k} \phi_i^{T_0-j} (1-\phi_i)^{(T-T_0)+k} \left((T-T_0+k)\,\lambda_i + 1/\alpha\right)^{-\left(\sum_t^T n_t + 1/\alpha\right)}}, \tag{3.6}$$

Contrairement au modèle classique BNMV, la moyenne prédictive ne dépend pas seulement du nombre de réclamations passées, mais aussi du nombre de périodes d'assurance dans lesquelles l'assuré a réclamé.

- Distribution *0-BNMV*

Ce modèle est particulier car il génère deux formes de prime prédictive, dépendant de l'historique de perte. Dans le cas où l'assuré a réclamé au moins une fois dans les T dernières années, l'espérance est égale à :

$$E\left[N_{i,T+1} \middle| N_{i,1},\dots,N_{i,T}, T_0 \neq T\right] = \lambda_i \frac{N_{i\bullet} + 1/\alpha}{T\lambda_i + 1/\alpha}, \tag{3.7}$$

qui est similaire à la prime prédictive obtenue pour le modèle BNMV. A l'inverse, la prime pour les assurés n'ayant pas réclamé est égale à :

$$E\,[N_{i,T+1}\,|N_{i,1},...,N_{i,T},T_0=T]=\lambda_i\,\frac{(1-\phi_i)\left(\dfrac{1/\alpha}{T\lambda_i+1/\alpha}\right)^{1+1/\alpha}}{\phi_i+(1-\phi_i)\left(\dfrac{1/\alpha}{T\lambda_i+1/\alpha}\right)^{1/\alpha}}. \quad (3.8)$$

Ainsi, bien que les équations des primes prédictives ne dépendent que de $N_{i\bullet}$, le modèle dépend aussi de T_0.

3.1.3. Modèles à barrière

Puisque le modèle à barrière utilise deux effets aléatoires tel que $\Theta_i = (\theta_{i,1}, \theta_{i,2})$, la forme de la distribution *a posteriori* est beaucoup plus complexe :

$$g(\theta_{i,1},\theta_{i,2}\,|\mathrm{II}_{i,T})$$
$$=\frac{\theta_{i,1}^{K_{i\bullet}+a_i-1}(1-\theta_{i,1})^{T-K_{i\bullet}}\ \theta_{i,2}^{N_{i\bullet}-K_{i\bullet}+1/\alpha_i-1}\ e^{\theta_{i,1}(K_{i\bullet}\lambda_i+1/\alpha)}\ c(F_1(\theta_{i,1}),F_2(\theta_{i,2}))}{\int_0^1\int_0^\infty \theta_{i,1}^{K_{i\bullet}+a_i-1}(1-\theta_{i,1})^{T-K_{i\bullet}}\ \theta_{i,2}^{N_{i\bullet}-K_{i\bullet}+1/\alpha_i-1}\ e^{\theta_{i,1}(K_{i\bullet}\lambda_i+1/\alpha)}\ c(F_1(\theta_{i,1}),F_2(\theta_{i,2}))\,d\theta_{i,1}d\theta_{i,2}}$$

où $c(.)$ est la densité de la copule déterminant la dépendance entre les effets aléatoires. Tout comme le modèle gonflé à zéro, la distribution *a posteriori* ne dépend pas uniquement du nombre de réclamations passées $N_{i\bullet}$, mais aussi du nombre de périodes d'assurance avec au moins une réclamation $K_{i\bullet}$. Toutefois, par opposition à ces modèles gonflés à zéro, la distribution prédictive de $N_{i,T+1}$ ne peut pas s'exprimer de manière analytique pour toutes les copules.

Toutefois, dans le cas de la copule d'indépendance, chacun des processus du modèle à barrière avec effets aléatoires modélisé peut être exprimé indépendamment de l'autre. En conséquence, les deux processus peuvent aussi être étudiés séparément lors de l'analyse prédictive. Comme prouvé par Boucher, Denuit et Guillén (2006b), l'espérance prédictive de ce modèle est égale à :

$$E\,[N_{i,T+1}\,|\mathrm{II}_{i,T}]=\left(\frac{K_{i\bullet}+a_i}{K_{i\bullet}\gamma_i+1/\alpha}\right)\left(\frac{\gamma_i N_{i\bullet}+(1+\gamma_i)/\alpha}{T+b+a_i}\right). \quad (3.9)$$

Tout comme les modèles gonflés à zéro, la moyenne prédictive dépend aussi du nombre de périodes d'assurance dans lesquelles l'assuré a réclamé. Néanmoins, lorsque $T \to \infty$ et que $K_{i\bullet}$ est relativement grand, tout comme le modèle BNMV, la prime de l'assuré ne dépend plus de $K_{i\bullet}$, mais tend vers sa moyenne annuelle de sinistres réclamés.

Les modèles utilisant les copules gaussienne et de Fréchet-Hoeffding ne peuvent pas s'exprimer de manière close, tout comme leur distribution prédictive. Ainsi, Boucher, Denuit et Guillén (2006a) ont utilisé des simulations Monte Carlo par chaînes de Markov (MCMC) pour estimer ces primes prédictives. Le principe des simulations MCMC est de simuler des réalisations d'une chaîne de Markov ayant une distribution stationnaire correspondant à la distribution jointe des effets aléatoires (Smith et Roberts, 1993; ou Scollnik, 2001, pour une introduction avec applications en sciences actuarielles). Puisque cette méthode n'est pas très rapide, ni facile ou accessible, il nous semble pertinent d'étudier une autre manière d'évaluer ces primes.

3.2. Estimation par crédibilité linéaire

La crédibilité linéaire est une théorie utilisée afin d'obtenir la prime d'assurance d'un assuré en fonction d'une moyenne pondérée entre son expérience passée et sa prime *a priori*. L'avantage d'une telle approche est qu'il n'est plus nécessaire de calculer la distribution prédictive exacte d'une distribution et que les hypothèses de la crédibilité linéaire dépendent beaucoup moins des particularités des modèles étudiés.

3.2.1. Crédibilité univariée

La forme la plus ancienne et la plus connue de la théorie de la crédibilité est la prime de Bühlmann (1967), qui n'utilise que les informations sur les $n_{i,t}$, $t = 1, ..., T$ pour prédire la prime. Ainsi, cette prime prédictive a la forme suivante :

$$P^*_{i,T+1} = \nu_i \bar{N}_i + \pi_i P_{i,0} \tag{3.10}$$

où $P^*_{i,T+1}$ est la prime prédictive au temps $T + 1$, $P_{i,0}$ est la prime *a priori*. En minimisant l'erreur quadratique, Bühlmann (1967) a montré que les paramètres ν_i et π_i sont égaux à :

$$\nu_i = \frac{Cov(\bar{N}_i, N_{i,T+1})}{Var(\bar{N}_i)} \tag{3.11}$$

$$\pi_i = 1 - \nu_i \ . \tag{3.12}$$

Bien évidemment, parce que cette prime prédictive n'utilise pas toutes les statistiques pertinentes au modèle, tel que les $k_{i,t}$, le modèle est plutôt limité dans ses prédictions. Le but est donc d'utiliser un modèle de crédibilité linéaire bivariée utilisant non seulement les $n_{i,t}$ mais aussi ces $k_{i,t}$.

3.2.2. Crédibilité bivariée

En utilisant un type de primes de crédibilité utilisant aussi le nombre de périodes d'assurance avec réclamations $\bar{K}_i$, un modèle de crédibilité bivariée peut s'exprimer comme suit :

$$P^{**}_{i,T+1} = \delta_i \bar{K}_i + \tau_i \bar{N}_i + \omega_i \tag{3.13}$$

où les deux paramètres δ_i et τ_i peuvent être vus comme des variables explicatives provenant respectivement de la régression de la prime future avec $\bar{K}_i$ et $\bar{N}_i$. Ainsi, ces paramètres décomposent la prime future de l'assuré *i* en deux statistiques, expliquant plus en détails l'impact d'une réclamation sur la prime d'assurance. Il peut être montré que les paramètres de ce modèle de crédibilité linéaire bivariée sont égaux à :

$$\delta_i = \frac{Cov[\bar{K}_i, P_{i,T+1}]\, Var[\bar{N}_i] - Cov[\bar{N}_i, P_{i,T+1}]\, Cov[\bar{K}_i, \bar{N}_i]}{Var[\bar{N}_i]\, Var[\bar{K}_i] - Cov[\bar{K}_i, \bar{N}_i]^2} \tag{3.14}$$

$$\nu_i = \frac{Cov[\bar{N}_i, P_{i,T+1}]\, Var[\bar{K}_i] - Cov[\bar{K}_i, P_{i,T+1}]\, Cov[\bar{K}_i, \bar{N}_i]}{Var[\bar{N}_i]\, Var[\bar{K}_i] - Cov[\bar{K}_i, \bar{N}_i]^2} \tag{3.15}$$

$$\omega_i = P_{i,0}\,(1 - \nu_i) - \delta_i\, E[\bar{K}_i]\ . \tag{3.16}$$

Le modèle classique de la crédibilité linéaire de Bühlmann (1967), representé à l'équation (3.10), génère des résultats égaux à l'espérance prédictive pour les distributions conjuguées (Jewell, 1975), comme pour le cas Poisson avec effets aléatoires gamma. Puisque la crédibilité linéaire ne se base que sur les deux premiers moments de la distribution des effets aléatoires, l'approximation linéaire d'un modèle de Poisson avec effets aléatoires a la même forme peu importe la distribution des effets aléatoires (inverse-Gaussienne, lognormale ou autres). Évidemment, l'approximation de la crédibilité linéaire par P^*_{T+1} ne génèrera pas des primes prédictives exactes pour ces autres modèles.

À l'opposé, lorsque la crédibilité linéaire bivariée P^{**}_{T+1} est utilisée avec le nombre de périodes d'assurance avec réclamations, l'estimateur de l'espérance prédictive ne sera pas le même pour ces différents modèles de Poisson avec effets aléatoires. En effet, la prime de crédibilité ne dépendra pas seulement des deux premiers moments de la distribution d'hétérogénéité, $E\,[\theta]$ et $Var\,[\theta]$, mais aussi des expressions $E\,[\theta e^{-\lambda\theta}]$, $E\,[e^{-\lambda\theta}]$ ou $E\,[e^{-2\lambda\theta}]$. En conséquence, les paramètres obtenus par les équations (3.14) à (3.16) n'ont clairement pas la même forme si les effets aléatoires sont gamma, inverse-Gaussienne ou lognormale.

L'approximation par crédibilité linéaire peut être vue comme une projection de la vraie prime prédictive sur l'espace de Hilbert créée par la prime de crédibilité P^{**}_{T+1}. Conséquemment, la prime de crédibilité utilisant les équations (3.14) et (3.15) devrait aussi générer une prime de crédibilité exacte pour le modèle Poisson-gamma. Pour obtenir ce résultat, δ_i doit être égal à 0, révélant l'égalité suivante :

$$Cov\,[\bar{K}_i, P_{i,T+1}]\ Var\,[\bar{N}_i] = Cov\,[\bar{N}_i, P_{i,T+1}]\ Cov\,[\bar{K}_i, \bar{N}_i]\,. \quad (3.17)$$

Pour un modèle de Poisson avec effet aléatoire, ce résultat ne tient que lorsque cet effet aléatoire est gamma. En effet, en développant cette dernière égalité, le résultat suivant est obtenu :

$$\frac{E[e^{-\lambda_i\theta}]}{E[\theta e^{-\lambda_i\theta}]} = (1+\lambda_i\theta)\,. \quad (3.18)$$

Cette égalité est vérifiée pour une hétérogénéité gamma, mais pas pour une hétérogénéité inverse-Gaussienne. En effet, dans cette situation, le résultat suivant est généré :

$$\frac{E[e^{-\lambda_i\theta}]}{E[\theta e^{-\lambda_i\theta}]} = \sqrt{1+2\lambda_i\theta}\,. \quad (3.19)$$

Un résultat indiquant que l'égalité de l'équation (3.17) ne tient pas pour un effet aléatoire lognormal peut aussi être prouvé. En conséquence, sauf pour le modèle avec hétérogénéité gamma, le modèle à crédibilité linéaire bivariée utilisant le nombre de périodes avec réclamations ne génère pas la même résultat que le modèle classique de Bülhmann (même si les deux primes sont très proches...).

Pour les modèles analysés dans cet article, il semble évident que la crédibilité bivariée améliore l'approximation puisque $\bar{K}_i$ est l'une des statistiques exhaustives de la distribution prédictive. Une autre approximation de crédibilité mérite d'être mentionnée. En effet, une prime de crédibilité utilisant seulement les informations sur $\bar{K}_i$, une statistique potentiellement exhaustive pour certains modèles, et signifiant un $\nu_i = 0$ est intéressante :

$$P^{***}_{i,T=1} = \delta_i\bar{K}_i + \omega_i \quad (3.20)$$

avec :

$$\delta_i = \frac{Cov(\bar{K}_i, N_{i,T+1})}{Var(\bar{K}_i)} \tag{3.21}$$

$$\omega_i = P_{i,0} - \delta_i\, E\,[\bar{K}_i] \tag{3.22}$$

Parce que $\bar{K}_i$ n'est pas un bon estimateur de $P_{i,T+1}$, le complément de crédibilité ω_i n'est pas seulement égal à $1 - \delta_i$. En effet, ce dernier doit corriger cette erreur d'approximation.

Pour alléger la structure du papier, comme décrit précédemment, les moments nécessaires aux calculs des primes de crédibilité sont inscrits en annexe.

4. EXEMPLE NUMÉRIQUE

Afin d'illustrer et comparer les primes de crédibilité linéaires aux moyennes prédictives exactes, nous travaillons avec les résultats obtenus dans les papiers de Boucher, Denuit et Guillén (2006a, 2006b). Ces données correspondent à un échantillon d'un portefeuille d'une grande compagnie d'assurance espagnole. Seules les données de la responsabilité civile touchant aux véhicules de particuliers ont été utilisées. Les primes prédictives ont été calculées pour diverses configurations d'historique de réclamations, sur une période d'assurance de 10 ans. Malgré que les deux papiers utilisent la même base de données, les données utilisées pour calculer les primes futures ne peuvent pas être comparées d'un modèle à l'autre. En effet, dans Boucher, Denuit et Guillén (2006a), les auteurs utilisent les 7 premières années de la base de données, alors que Boucher, Denuit et Guillén (2006b) n'utilisent que les 6 premières. L'exercice réfère donc principalement à l'analyse de l'approximation de la théorie de la crédibilité linéaire.

4.1. Modèles gonflés à zéro

Le type d'assurés qui a été sélectionné pour le calcul des primes prédictives des modèles gonflés à zéro correspond à un profil de risque moyen, ayant les paramètres suivants selon les deux modèles analysés.

TABLEAU 4.1
PARAMÈTRES UTILISÉS POUR LES MODÈLES GONFLÉS À ZÉRO

Paramètre	Modèles	
	MP0-gamma	0-BNMV
ϕ_i	0,2028	0,0262
λ_i	0,0841	0,0677
α	0,8304	0,7678

Le tableau suivant indique les résultats numériques obtenus en utilisant les primes calculées de manière exacte et le modèle de crédibilité bivariée pour le modèle *MP0-gamma*.

TABLEAU 4.2
PRIME DE CRÉDIBILITÉ POUR LE MODÈLE MP0-GAMMA

Méthodes	$K_.$	Somme des sinistres rapportés ($N_.$)					
		0	1	2	3	4	10
Primes exactes	0	0,0434	–	–	–	–	–
	1	–	0,0789	0,1151	0,1515	0,1882	0,4150
	2	–	–	0,1138	0,1498	0,1860	0,4088
	3	–	–	–	0,1482	0,1839	0,4029
	4	–	–	–	–	0,1818	0,3972
	10	–	–	–	–	–	0,3672
Crédibilité bivariée	0	0,0436	–	–	–	–	–
	1	–	0,0786	0,1142	0,1498	0,1854	0,3989
	2	–	–	0,1135	0,1491	0,1847	0,3983
	3	–	–	–	0,1485	0,1841	0,3977
	4	–	–	–	–	0,1835	0,3971
	10	–	–	–	–	–	0,3933

Pour les situations les plus communes, i.e. pour $N_{\bullet} \leq 4$, le modèle de crédibilité bivariée approxime relativement bien la structure de primes. Néanmoins, l'approximation est beaucoup moins bonne pour les autres situations qui peuvent qualifiées de plus extrêmes. En effet, la forme de la prime prédictive en fonction de $\bar{K}_i$ ou $\bar{N}_i$ est très complexe et il semble que l'approximation linéaire de ces deux variables ne génère pas un ajustement intéressant. En analysant plus en détail les résultats obtenus pour ce modèle de crédibilité, nous voyons que δ = -0,0063 et ν = 0,3560, signifiant que l'impact de $\bar{K}_i$ reste très marginal dans le calcul de la prime. Bien que cela semble justifié pour la majorité des cas (la partie du gauche du tableau), cela n'est pas valide pour les autres situations où l'on peut remarquer une différence de primes considérable selon $\bar{K}_i$.

Ainsi, s'il devenait essentiel d'évaluer la prime de crédibilité linéaire pour un tel modèle, la prime de crédibilté de Bühlmann serait sans doute plus intéressante puisque moins complexe, et générant quasiment les mêmes résultats que la prime de crédibilité bivariée. Notons ici que nous travaillons avec des données ne dépendant que légèrement de $\bar{K}_i$, ainsi, il pourrait être intéressant de refaire l'exercice pour une couverture d'assurance où cette dépendance est plus forte.

Le tableau suivant, quant à lui, illustre les primes de crédibilité obtenues pour le modèle 0-BNMV. Par opposition à l'autre modèle gonflé à zéro de ce papier, la crédibilité linéaire utilisée n'est pas bivariée, mais plutôt univariée. En effet, la prime prédictive ne dépend directement de $\bar{K}_i$. Comme mentionné plus tôt dans le papier, l'estimateur de crédibilité linéaire d'une distribution Poisson avec effets aléatoires gamma (modèle BNMV) génère des primes qui sont égales à l'espérance prédictive du modèle. Néanmoins, pour ce modèle 0-BNMV, le modèle de Bühlmann de l'équation (3.11) ne génère pas de résultats exacts, comme l'indique le tableau suivant.

TABLEAU 4.3
PRIME DE CRÉDIBILITÉ POUR LE MODÈLE 0-BNMV

	Somme des sinistres rapportés ($N_{\bullet}$)					
Modèles	**0**	**1**	**2**	**3**	**4**	**10**
Primes exactes	0,0426	0,0787	0,1129	0,1471	0,1813	0,3864
Crédibilité univariée	0,0429	0,0778	0,1128	0,1477	0,1826	0,3923

Dans cet exemple numérique, la valeur de ν_i de l'équation (3.11) génère une valeur de 0,3495. Même si les primes ne sont pas tout-à-fait exactes, elles restent néanmoins très similaires. Une simple petite transformation du modèle permet toutefois d'obtenir des primes exactes pour l'approximation linéaire. En effet, si le modèle 0-BNMV était utilisé comme un modèle BNMV standard, les primes de crédibilité seraient identiques aux primes exactes pour les assurés avec au moins une réclamation. Pour les assurés n'ayant pas réclamé, cette prime de crédibilité linéaire n'a qu'à être divisée par un facteur de correction. En utilisant l'équation (3.8) cette correction se calcule directement :

$$1+\frac{\phi_i}{1-\phi_i}\left(\frac{1/\alpha}{T\lambda_i+1/\alpha}\right)^{-1/\alpha} .$$

En conséquence, il est intéressant de noter que le modèle 0-BNMV se veut une alternative intéressante au modèle classique MVNB. En effet, non seulement parce que ses primes de crédibilité sont moins sévères que le modèle classique, mais aussi parce que la prime de crédibilité classique de Bülhmann peut être utilisée moyennant une légère correction pour un groupe d'assurés spécifiques.

4.2. Modèles à barrière

Pour le modèle à barrière, le profil d'assuré choisi par Boucher, Denuit et Guillén (2006a) correspond à un risque moyen (par rapport à la base de données analysée). Plus spécifiquement, les paramètres utilisés pour le calcul des primes sont les suivants, où les moments joints ont été obtenus numériquement (tableau 4.4).

TABLEAU 4.4
PARAMÈTRES UTILISÉS POUR LES MODÈLES À BARRIÈRE

Paramètre	Copule du modèle à barrière		
	Indépendance	Gaussienne	Fréchet-Hoeffding
a_i	1,3019	1,3102	1,3192
b	19,9640	19,9568	20,0836
γ_i	0,0770	0,0434	0,0410
α	0,8122	0,7518	0,8818
ρ	–	0,8424	1,0000
$E[\theta_{i,1}\ \theta_{i,2}]$	0,0612	0,1092	0,1153
$E[\theta^2_{i,1}\ \theta_{i,2}]$	0,0063	0,0162	0,0176
$E[\theta_{i,1}\ \theta^2_{i,2}]$	0,1109	0,3454	0,3472
$E[\theta^2_{i,1}\ \theta^2_{i,2}]$	0,0115	0,0641	0,0682

Le premier modèle à barrière réfère à la distribution supposant que les effets aléatoires sont indépendants. Bien qu'il existe une manière d'obtenir une prime exacte en utilisant une formule fermée (équation (3.9)), il reste pertinent d'analyser la justesse de l'approximation de la crédibilité linéaire bivariée. Le tableau 4.5 illustre la prime exacte et la prime de crédibilité linéaire obtenues pour ce modèle.

L'approximation du modèle par la crédibilité linéaire est satisfaisante et les écarts entre l'approximation et le résultat exact sont minces. Pour les assurés ayant une expérience de sinistres *normale*, la prime de crédibilité est même très proche de la prime exacte. Les différences surviennent pour les situations plus extrêmes, c'est-à-dire dans les situations où l'assuré a réclamé plus de 3 fois dans la même année, ou dans le cas où 10 réclamations ont été effectuées.

Tout comme l'analyse de la sévérité, bien détaillée dans Goulet et Forgues (2006), il est aussi possible d'étudier une autre forme de prime de crédibilité pour ce modèle. Parce le montant total des réclamations correspond au produit de la fréquence et de la sévérité, Gerber (1972) a proposé d'utiliser un estimateur de crédibilité de Bühlmann pour la fréquence et un second pour la sévérité. La même

TABLEAU 4.5
PRIME DE CRÉDIBILITÉ POUR LA COPULE D'INDÉPENDANCE

Méthode	$K_\bullet$	Somme des sinistres rapportés ($N_\bullet$)					
		0	1	2	3	4	10
Primes exactes	0	0,0448	–	–	–	–	–
	1	–	0,0790	0,0833	0,0876	0,0920	0,1180
	2	–	–	0,1128	0,1187	0,1246	0,1598
	3	–	–	–	0,1465	0,1538	0,1972
	4	–	–	–	–	0,1800	0,2309
	10	–	–	–	–	–	0,3786
Crédibilité bivariée	0	0,0448	–	–	–	–	–
	1	–	0,0788	0,0846	0,0904	0,0962	0,1308
	2	–	–	0,1129	0,1186	0,1244	0,1590
	3	–	–	–	0,1469	0,1526	0,1872
	4	–	–	–	–	0,1809	0,2155
	10	–	–	–	–	–	0,3849

approche pourrait être intéressante pour les modèles à barrière. Puisque la théorie de la crédibilité linéaire donne des résultats exacts pour les distributions conjuguées (Jewell, 1975), ces primes de crédibilité pour les combinaison Bernoulli-beta ($\nu_{i,1}$) et Poisson-gamma ($\nu_{i,2}$) génèrent exactement leur moyenne prédictive. Dans cette situation, à l'aide des équations (3.11) et (3.12), il peut être prouvé que :

$$\nu_{i,1} = \frac{T}{T + a_i + b} \tag{4.1}$$

$$\nu_{i,2} = \frac{K_i}{K_i + \gamma_i}, \tag{4.2}$$

ce qui génère une prime prédictive correspondant à celle obtenue à l'équation (3.9). Par contre, en développant l'équation de la prime de crédibilité, nous pouvons voir que l'expression n'est pas la même que celle exprimée à l'équation (3.13) puisque :

$$\begin{aligned} P_{i,T+1} &= (\nu_{i,1}\bar{K}_i + (1-\nu_{t,1})\,\mu_{i,1})\,(\nu_{i,2}\,\bar{N}_i + (1-\nu_{i,2})\,\mu_{i,2}) \\ &= \nu_{i,1}\,\nu_{i,2}\,\bar{K}_i\,\bar{N}_i + (1-\nu_{i,1})\,\nu_{i,2}\,\mu_{i,1}\,\bar{N}_i \\ &\quad + (1-\nu_{i,2})\,\nu_{i,1}\,\mu_{i,2}\,\bar{K}_i + (1-\nu_{i,1}-\nu_{i,2}+\nu_{i,1}\,\nu_{i,2})\,\mu_i \end{aligned} \tag{4.3}$$

où les primes *a priori* sont définies comme :

$$\mu_{i,1} = \frac{a_i}{a_i + b},\ \mu_{i,2} = 1+\gamma_i,\ \mu_i = \mu_{i,1}\mu_{i,2} = P_{i,0}\ .$$

Les situations de corrélation entre les effets aléatoires sont intéressantes à analyser car elles permettent de voir si l'approximation des primes par l'utilisation des premiers et seconds moments du modèle capture bien la corrélation des Θ_i (tableau 4.6).

Pour les assurés n'ayant pas réclamé ou seulement réclamé une fois au cours des 10 dernières années, le modèle de crédibilité s'ajuste bien. Par contre, pour les autres situations, des écarts assez importants peuvent être observés. En effet, dans le cas d'assurés ayant effectué plusieurs réclamations dans la même année d'assurance, le modèle de crédibilité linéaire bivariée ne semble pas bien capturer l'effet des Θ_i sur la prime future.

En observant plus attentivement les différences entre les primes exactes et les primes de crédibilité, nous pouvons constater que la régression sur les $K_\bullet$ et les $N_\bullet$ n'est pas suffisante car un terme en $K_\bullet \times N_\bullet$ semble nécessaire pour modéliser la contagion (dépendance) des effets aléatoires. En effet, nous pouvons voir que bien que l'approximation est plutôt juste pour les éléments de la diagonale, elle est plutôt médiocre pour les éléments situés dans le coin supérieur droit du tableau. Dans le cas du modèle avec effets aléatoires indépendants, nous pouvions voir que pour un $\bar{N}_i$ fixe, la relation entre $\bar{K}_i$ et la prime prédictive était proportionnelle. L'introduction de la corrélation entre les effets aléatoires change ce lien puisqu'il est plus *rentable* pour un assuré d'avoir toutes ses réclamations situées dans la même période qu'échelonnées sur plusieurs contrats d'assurance. Similairement, tout porte à croire que l'introduction d'un terme croisé améliorerait la prédiction.

TABLEAU 4.6
PRIMES DE CRÉDIBILITÉ POUR LA COPULE GAUSSIENNE

Méthode	$K_{\bullet}$	Somme des sinistres rapportés ($N_{\bullet}$)					
		0	1	2	3	4	10
Primes exactes	0	0,0441	–	–	–	–	–
	1	–	0,0776	0,1063	0,1336	0,1598	0,3217
	2	–	–	0,1113	0,1403	0,1687	0,3341
	3	–	–	–	0,1444	0,1750	0,3442
	4	–	–	–	–	0,1774	0,3529
	10	–	–	–	–	–	0,3770
Crédibilité bivariée	0	0,0441	–	–	–	–	–
	1	–	0,0773	0,1155	0,1536	0,1918	0,4208
	2	–	–	0,1105	0,1487	0,1868	0,4159
	3	–	–	–	0,1437	0,1819	0,4109
	4	–	–	–	–	0,1769	0,4059
	10	–	–	–	–	–	0,3762

Le modèle avec copule d'indépendance a pu générer une bonne approximation du modèle. La présence d'un élément croisé entre $K_{\bullet}$ et $N_{\bullet}$, tel qu'exprimé par l'équation (4.3) n'est peut-être pas étranger à cela. Toutefois, dans le cas où il existe une dépendance entre chaque processus ou dans le cas des modèles gonflés à zéro, il est évident que ce modèle de crédibilité linéaire n'est pas approprié car théoriquement faux. En effet, cette façon d'exprimer la prime suppose l'indépendance entre le nombre de réclamations et le nombre de périodes avec réclamations.

De prochaines recherches pourraient permettre de voir si l'ajout d'un tel terme dans l'approximation de la prime améliore l'ajustement. Il est à noter que l'addition de cette statistique créera le besoin de calculer des moments du troisième ou même du quatrième ordre.

Comme dernier exemple numérique, nous analysons les résultats obtenus pour la copule représentant la borne supérieure de dépendance entre les effets aléatoires (tableau 4.7).

TABLEAU 4.7
PRIMES DE CRÉDIBILITÉ POUR LA COPULE DE FRÉCHET-HOEFFDING

Méthode	$K_{\bullet}$	Somme des sinistres rapportés ($N_{\bullet}$)					
		0	1	2	3	4	10
Primes exactes	0	0,0441	–	–	–	–	–
	1	–	0,0775	0,1137	0,1500	0,1861	0,4044
	2	–	–	0,1106	0,1469	0,1826	0,4006
	3	–	–	–	0,1438	0,1795	0,3965
	4	–	–	–	–	0,1764	0,3928
	10	–	–	–	–	–	0,3704
Crédibilité bivariée	0	0,0441	–	–	–	–	–
	1	–	0,0771	0,1176	0,1580	0,1985	0,4412
	2	–	–	0,1101	0,1506	0,1910	0,4337
	3	–	–	–	0,1431	0,1835	0,4263
	4	–	–	–	–	0,1761	0,4188
	10	–	–	–	–	–	0,3741

Sans surprise, le modèle de crédibilité linéaire bivariée n'ajuste pas bien les données et un terme croisé semble aussi nécessaire afin de modifier le lien entre la prime future et $\bar{K}_i$ pour un $\bar{N}_i$ fixe.

5. CONCLUSION

Ce papier nous a permis de réviser les modèles gonflés à zéro et les modèles à barrière. Les différences d'interprétations entre les modèles ont été pointées, tout comme l'impact de l'effet aléatoire dans leur construction. Alors que les modèles gonflés à zéro sont valides avec un seul effet aléatoire, il n'est pas possible d'utiliser un modèle à barrière sans l'utilisation de deux termes d'hétérogénéité.

Plus spécifiquement, pour ces modèles, nous avons pu voir que le choix de la copule pour modéliser la dépendance entre les effets aléatoires a une importance primordiale pour le calcul des primes prédictives.

L'approximation des primes prédictives pour les modèles gonflés à zéro a été satisfaisante. Par opposition, bien que nous ayons pu trouver une façon d'exprimer les primes prédictives de manière exacte avec la théorie de la crédibilité linéaire pour le modèle ayant des effets aléatoires indépendants, nous avons pu remarquer que cette théorie n'est pas efficace pour approximer la structure de dépendance des effets aléatoires avec d'autres copules. Puisque chaque réclamation $k_{i,t}$ modifie l'effet aléatoire $\theta_{i,1}$ et que celui-ci est corrélé à $\theta_{i,2}$ (et vice-versa), il nous semble logique de présumer qu'une meilleure approximation de la prime résulterait de l'ajout d'un terme croisé entre $K_\bullet$ et $N_\bullet$.

Des recherches futures touchant non seulement à ce type d'approximation par la théorie de la crédibilité linéaire semblent être une voie intéressante, particulièrement pour les modèles à barrières qui ne génèrent pas une forme fermée d'espérance prédictive.

ANNEXE

L'expression des moments nécessaires à l'évaluation des coefficients de crédibilité est présentée dans cette annexe.

A.1. Généralités

Plus généralement, les équations suivantes sont utilisées afin d'obtenir les moments nécessaires :

$$Var[N_{i,t}] = E\big[Var[N_{i,t}|\Theta]\big] + Var\big[E[N_{i,t}|\Theta]\big] \tag{A.1}$$

$$Var[\bar{N}_i] = \frac{E\big[Var[N_{i,t}|\Theta]\big]}{T} + Var\big[E[N_{i,t}|\Theta]\big] \tag{A.2}$$

$$Var[K_{i,t}] = E\big[Var[K_{i,t}|\Theta]\big] + Var\big[E[K_{i,t}|\Theta]\big] \tag{A.3}$$

$$Var[\bar{K}_i] = \frac{E\big[Var[K_{i,t}|\Theta]\big]}{T} + Var\big[E[K_{i,t}|\Theta]\big] \tag{A.4}$$

$$Cov[\bar{N}_i, \bar{K}_i] = \frac{E\big[Cov[K_{i,t}, N_{i,t}|\Theta]\big]}{T} + Cov\big[E[K_{i,t}|\Theta], E[N_{i,t}|\Theta]\big] \tag{A.5}$$

$$Cov[\bar{N}_i, P_{i,T+1}] = Var\big[E[N_{i,t}|\Theta]\big] \tag{A.6}$$

$$Cov\big[E[K_{i,t}|\Theta], E[N_{i,t}|\Theta]\big] = Cov[\bar{K}_i, P_{i,T+1}] \tag{A.7}$$

A.2. Distribution BNMV

Bien que les moments de cette distribution soit bien connus, il est utile d'en rappeler quelques expressions, particulièrement pour les moments impliquant les variables $K_{i,t}$ ou $\bar{K}_i$:

$$E[N_{i,t}] = E[\bar{N}_i] = \lambda_i \tag{A.8}$$

$$E[K_{i,t}] = E[\bar{K}_i] = 1 - \left(\frac{1/\alpha}{\lambda_i + 1/\alpha}\right)^{1/\alpha} \tag{A.9}$$

$$E\big[Var[N_{i,t}|\Theta]\big] = \lambda_i \tag{A.10}$$

$$Var\big[E[N_{i,t}|\Theta]\big] = \lambda_i^2\alpha \tag{A.11}$$

$$E\big[Var[K_{i,t}|\Theta]\big] = \left(\frac{1/\alpha}{2\lambda_i + 1/\alpha}\right)^{1/\alpha} - \left(\frac{1/\alpha}{\lambda_i + 1/\alpha}\right)^{2/\alpha} \tag{A.12}$$

$$Var\left[E[K_{i,t}|\Theta]\right] = \left(\frac{1/\alpha}{\lambda_i + 1/\alpha}\right)^{1/\alpha} - \left(\frac{1/\alpha}{2\lambda_i + 1/\alpha}\right)^{1/\alpha} \tag{A.13}$$

$$Cov[\bar{K}_i, P_{i,T+1}] = \lambda_i \left(\left(\frac{1/\alpha}{\lambda_i + 1/\alpha}\right)^{1/\alpha} - \left(\frac{1/\alpha}{\lambda_i + 1/\alpha}\right)^{1/\alpha+1}\right) \tag{A.14}$$

$$E\left[Cov[K_{i,t}, N_{i,t}|\Theta]\right] = \lambda_i \left(\frac{1/\alpha}{\lambda_i + 1/\alpha}\right)^{1/\alpha+1} \tag{A.15}$$

A.3. Modèles gonflés à zéro

Le calcul des moments des deux modèles gonflés à zéro est beaucoup plus simple à obtenir. Ainsi, les sous-sections suivantes indiquent ces différents moments.

A.3.1.Distribution MP0-gamma

$$E[N_{i,t}] = E[\bar{N}_i] = \lambda_i(1-\phi_i) \tag{A.16}$$

$$E[K_{i,t}] = E[\bar{K}_i] = 1 - \phi_i - (1-\phi_i)\left(\frac{1/\alpha}{\lambda_i + 1/\alpha}\right)^{1/\alpha} \tag{A.17}$$

$$E\left[Var[N_{i,t}|\Theta]\right] = (1-\phi_i)\,\lambda_i + \phi(1-\phi)\,\lambda_i^2(\alpha+1) \tag{A.18}$$

$$Var\left[E[N_{i,t}|\Theta]\right] = (1-\phi_i)^2\lambda_i^2\alpha \tag{A.19}$$

$$\begin{aligned} E\left[Var[K_{i,t}|\Theta]\right] &= (1-\phi)\left(1 - \left(\frac{1/\alpha}{\lambda_i + 1/\alpha}\right)^{1/\alpha}\right) \\ &\quad - (1-\phi)^2\left(1 - 2\left(\frac{1/\alpha}{\lambda_i + 1/\alpha}\right)^{1/\alpha} + \left(\frac{1/\alpha}{2\lambda_i + 1/\alpha}\right)^{1/\alpha}\right) \end{aligned} \tag{A.20}$$

$$Var\left[E[K_{i,t}|\Theta]\right] = (1-\phi_i)^2\left(\left(\frac{1/\alpha}{2\lambda_i + 1/\alpha}\right)^{1/\alpha} - \left(\frac{1/\alpha}{\lambda_i + 1/\alpha}\right)^{2/\alpha}\right) \tag{A.21}$$

$$Cov[\bar{K}_i, P_{i,T+1}] = (1-\phi_i)^2\lambda_i\left(\left(\frac{1/\alpha}{\lambda_i + 1/\alpha}\right)^{1/\alpha} - \left(\frac{1/\alpha}{\lambda_i + 1/\alpha}\right)^{1/\alpha+1}\right) \tag{A.22}$$

$$E\left[Cov[K_{i,t} N_{i,t}|\Theta]\right] = (1-\phi_i)^2\lambda_i\left(\frac{1/\alpha}{\lambda_i + 1/\alpha}\right)^{1/\alpha+1} + (1-\phi_i)\phi_i\lambda_i \tag{A.23}$$

A.3.2. Distribution 0-BNMV

$$E[N_{i,t}] = E[\bar{N}_i] = \lambda_i(1-\phi_i) \tag{A.24}$$

$$E[K_{i,t}] = E[\bar{K}_i] = 1-(1-\phi_i)\left(\frac{1/\alpha}{\lambda_i+1/\alpha}\right)^{1/\alpha} \tag{A.25}$$

$$E\big[Var[N_{i,t}|\Theta]\big] = \lambda_i(1-\phi_i) \tag{A.26}$$

$$Var\big[E[N_{i,t}|\Theta]\big] = \lambda_i^2\big((1-\phi_i)(\alpha+1)-(1-\phi_i)^2\big) \tag{A.27}$$

$$E\big[Var[K_{i,t}|\Theta]\big] = (1-\phi_i)\left(\left(\frac{1/\alpha}{\lambda_i+1/\alpha}\right)^{1/\alpha} - \left(\frac{1/\alpha}{2\lambda_i+1/\alpha}\right)^{1/\alpha}\right) \tag{A.28}$$

$$Var\big[E[K_{i,t}|\Theta]\big] = (1-\phi_i)\left(\frac{1/\alpha}{2\lambda_i+1/\alpha}\right)^{1/\alpha} - (1-\phi_i)^2\left(\frac{1/\alpha}{\lambda_i+1/\alpha}\right)^{2/\alpha} \tag{A.29}$$

$$Cov[\bar{K}_i, P_{i,T+1}] = \lambda_i\left((1-\phi_i)^2\left(\frac{1/\alpha}{\lambda_i+1/\alpha}\right)^{1/\alpha} - (1-\phi_i)\left(\frac{1/\alpha}{\lambda_i+1/\alpha}\right)^{1/\alpha+1}\right) \tag{A.30}$$

$$E\big[Cov[K_{i,t}N_{i,t}|\Theta]\big] = \lambda_i(1-\phi_i)\left(\frac{1/\alpha}{\lambda_i+1/\alpha}\right)^{1/\alpha+1} \tag{A.31}$$

A.4. Modèles à barrière

Puisque les différentes copules associées aux modèles à barrière génèrent des distributions jointes différentes pour les effets aléatoires, les moments sont exprimés en utilisant ces moments joints. Ces moments joints ne peuvent pas être calculés analytiquement, mais de simples estimations numériques suffisent pour obtenir leurs valeurs.

Le calcul des différents moments nécessaires à l'évaluation des primes de crédibilité s'effectue en suivant souvent la même méthode. Par exemple, en supposant que $f(.|\lambda)$ est la fonction de probabilité d'une loi Poisson avec paramètre λ, les deux premiers moments du modèle se calculent comme suit :

$$E[N_{i,t}|\Theta_i] = \sum_{j=1}^{\infty} j\theta_{i,1} f(j-1|\gamma_i\theta_{i,2}) \tag{A.32}$$

$$= \theta_{i,1}\sum_{k=0}^{\infty}(1+k)f(k|\gamma_i\theta_{i,2})$$

$$= \theta_{i,1} + \gamma_i\theta_{i,1}\theta_{i,2} \tag{A.33}$$

$$E[N_{i,t}^2 | \Theta_i] = \theta_{i,1} \sum_{k=0}^{\infty} (1+k)^2 f(k | \gamma_i \theta_{i,2})$$
$$= \theta_{i,1}(\gamma_i^2 \theta_{i,2}^2 + 3\gamma_i \theta_{i,2} + 1) . \quad \text{(A.34)}$$

Plusieurs autres moments de la distribution à barrière se calculent sensiblement de la même manière puisque la variable aléatoire $K_{i,t}$ ne prend que des valeurs égales à 0 ou 1. Ainsi, pour continuer dans le développement des moments :

$$\begin{aligned} Cov[\bar{K}_i, P_{i,T+1}] &= Cov\big[E[\bar{K}_i | \Theta_i],\ E[P_{i,T+1} | \Theta_i]\big] + E\big[Cov[\bar{K}_i, P_{i,T+1} | \Theta_i]\big] \\ &= \frac{1}{T} \sum_{t=1}^{T} Cov\big[E[K_{i,t} | \Theta_i],\ E[P_{i,T+1} | \Theta_i]\big] + 0 \\ &= Cov[\theta_{i,1}, \theta_{i,1} + \gamma_i \theta_{i,1} \theta_{i,2}] \\ &= E[\theta_{i,1}^2] + \gamma_i E[\theta_{i,1}^2 \theta_{i,2}] - E[\theta_{i,1}]^2 - \gamma_i E[\theta_{i,1}]\, E[\theta_{i,1}\theta_{i,2}] \quad \text{(A.35)} \end{aligned}$$

où le second terme de la première équation est égal à 0 car les $K_{i,t}$, $t = 1, ..., T$ sont indépendants de $P_{i,T+1}$ lorsque conditionnés sur les effets aléatoires Θ_i. Suivant la même construction, la covariance suivante peut être calculée :

$$\begin{aligned} Cov[\bar{N}_i, P_{i,T+1}] &= \frac{1}{T} \sum_{t=1}^{T} Cov\big[E[N_{i,t} | \Theta_i],\ E[P_{i,T+1} | \Theta_i]\big] \\ &= E[\theta_{i,1}^2] + 2\gamma_i E[\theta_{i,1}^2 \theta_{i,2}] + \gamma_i^2 E[\theta_{i,1}^2 \theta_{i,2}^2] - E[\theta_{i,1}]^2 - 2\gamma_i E[\theta_{i,1}]\, E[\theta_{i,1}\theta_{i,2}] \quad \text{(A.36)} \end{aligned}$$

Le calcul de la covariance entre $\bar{K}_i$ et $\bar{N}_i$ est néanmoins un peu plus complexe car, malgré le conditionnement aux effets aléatoires Θ_i, il existe une dépendance entre certains termes :

$$Cov[\bar{K}_i, \bar{N}_i] = Cov\big[E[\bar{K}_i | \Theta_i],\ E[\bar{N}_i | \Theta_i]\big] + E\big[Cov[\bar{K}_i, \bar{N}_i | \Theta_i]\big] \quad \text{(A.37)}$$

où le premier élément se calcule comme suit :

$$\begin{aligned} Cov\big[E[\bar{K}_i | \Theta_i],\ E[\bar{N}_i | \Theta_i]\big] &= Cov[\theta_{i,1}, \theta_{i,1} + \gamma_i \theta_{i,1} \theta_{i,2}] \\ &= E[\theta_{i,1}^2] + \gamma_i E[\theta_{i,1}^2 \theta_{i,2}] - E[\theta_{i,1}]^2 - \gamma_i E[\theta_{i,1}]\, E[\theta_{i,1}\theta_{i,2}] \quad \text{(A.38)} \end{aligned}$$

et le second selon le développement suivant :

$$E\left[Cov[\bar{K}_i, \bar{N}_i | \Theta_i]\right] = \frac{1}{T^2} \sum_{j=1}^{T} \sum_{t=1}^{T} E\left[Cov[K_{i,j}, N_{i,t} | \Theta_i]\right]$$

$$= \frac{1}{T} E\left[Cov[K_{i,q}, N_{i,q} | \Theta_i]\right]$$

$$= \frac{1}{T}\Big(E\left[E[K_{i,q} N_{i,q} | \Theta_i]\right] - E\left[E[K_{i,q} | \Theta_i]\right] \; E\left[E[N_{i,q} | \Theta_i]\right]\Big)$$

$$= \frac{E[\theta_{i,1}] + \gamma_i E[\theta_{i,1}\theta_{i,2}] - E[\theta_{i,1}^2] - \gamma_i E[\theta_{i,1}^2 \theta_{i,2}]}{T} \qquad \text{(A.39)}$$

où le moment joint entre $K_{i,q}$ et $N_{i,q}$ se détermine selon le même raisonnement que celui utilisé pour $E[N_{i,t} \mid \Theta_i]$:

$$E[K_{i,q}, N_{i,q} | \Theta_i] = \sum_{m=1}^{1} \sum_{j=1}^{\infty} m j \theta_{i,1} f(j-1 | \gamma_i \theta_{i,2})$$

$$= \theta_{i,1} + \gamma_i \theta_{i,1} \theta_{i,2} \quad . \qquad \text{(A.40)}$$

Ainsi, en utilisant les résultats précédents et des développements similaires, les moments suivants peuvent être trouvés :

$$Var[\bar{K}_i] = E[\theta_{i,1}^2] - E[\theta_{i,1}]^2 + \frac{E[\theta_{i,1}] - E[\theta_{i,1}^2]}{T} \qquad \text{(A.41)}$$

$$Var[\bar{N}_i] = E\left[Var[\bar{N}_i]\right] + Var\left[E[\bar{N}_i]\right] \qquad \text{(A.42)}$$

$$E\left[Var[\bar{N}_i | \Theta]\right]$$
$$= \frac{\gamma_i^2 E[\theta_{i,1}\theta_{i,2}^2] + 3\gamma_i E[\theta_{i,1}\theta_{i,2}] + E[\theta_{i,1}] - E[\theta_{i,1}^2] - 2\gamma_i \, E[\theta_{i,1}^2 \theta_{i,2}] - \gamma_i^2 E[\theta_{i,1}^2 \theta_{i,2}^2]}{T} \qquad \text{(A.43)}$$

$$Var\left[E[\bar{N}_i | \Theta]\right]$$
$$= E[\theta_{i,1}^2] + 2\gamma_i E[\theta_{i,1}^2\theta_{i,2}] + \gamma_i^2 E[\theta_{i,1}^2\theta_{i,2}^2] - E[\theta_{i,1}]^2 - 2\gamma_i \, E[\theta_{i,1}] E[\theta_{i,1}\theta_{i,2}] - \gamma_i^2 E[\theta_{i,1}\theta_{i,2}]^2. \qquad \text{(A.44)}$$

Références

Bühlmann, H. (1967). « Experience Rating and Credibility ». *ASTIN Bulletin*, 4 : 199-207.

Bühlmann, H. et Straub, E. (1970). « Glaubwürdigkeit für Schadensätze ». *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, 70 :111-133.

Boucher, J.-P. et Denuit, M. (2006). « Fixed versus Random Effects in Poisson Regression Models for Claim Counts: Case Study with Motor Insurance ». *ASTIN Bulletin* 36 : 285-301.

Boucher, J.-P. et Denuit, M. (2007). « Credibility Premiums for the Zero-Inflated Model and New Hunger for Bonus Interpretation ». *Insurance: Mathematics and Economics*, sous presse.

Boucher, J.-P., Denuit, M. et Guillén, M. (2006a). *Independent and Correlated Random Effects for Hurdle Models Applied to Panel Data Count*. Document de travail, Université du Québec à Montréal, disponible au http://www.math.uqam.ca/actuariat/boucher/HurdlePanel.pdf.

Boucher, J.-P., Denuit, M. et Guillén, M. (2006b). *Number of Accidents or Number of Claims? An Approach with Zero-inflated Poisson Models for Panel Data*. Document de travail, Université du Québec à Montréal, disponible au http://www.math.uqam.ca/actuariat/boucher/ZIPanel.pdf.

Denuit, M. et Dhaene, J. (2001). « Bonus-Malus Scales using Exponential Loss Functions ». *German Actuarial Bulletin*, 25 : 13-27.

Denuit, M., Maréchal, X., Pitrebois, S. et Walhin, J.-F. (2007). *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Scales*. Wiley : New York.

Dionne, G. et Vanasse, C. (1989). « A Generalization of Automobile Insurance Rating Models: The Negative Binomial Distribution with Regression Component ». *ASTIN Bulletin*, 19 : 199-212.

Dionne, G. et Vanasse, C. (1992). « Automobile Insurance Ratemaking in the presence of asymmetrical information ». J*ournal of Applied Econometrics*, 7 : 149-165.

Ferreira, J. (1977). *Identifying Equitable Insurance Premiums for Risk Classes: an Alternative to the Classical Approach*. Lecture presentée à la 23e rencontre internationale de l'Institut of Management Sciences, Athène, Grèce.

Gerber, H. (1972). « Discussion of Hewitt (1971) ». *Proceedings of the Casulaty Actuarial Society*, 58 : 25-27.

Goulet, V., Forgues, A., et Lu, J. (2006). « Credibility for Severity Revisited ». *North American Actuarial Journal*, 10 : 49-62.

Hachemeister, C. (1975). « Credibility for Regression Models with Applications to Trend ». dans *Credibility: Theory and Applications*, ed. P. M. Kahn, Academic Press, New York, 129-163.

Hausman, J., Hall, B. et Griliches, Z. (1984). « Econometric Models for Count Data with Application to the Patents-R and D Relationship ». *Econometrica*, 52 :909-938.

Jewell, W. (1975). « The Use of Collateral Data in Credibility Theory: A Hierarchical Model ». *Giornale dell Istituto Italiano degli Attuari*, 38 : 1-16.

Johnson, N., Kotz, S. et Balakrishnan, N. (1996). *Discrete Multivariate Distributions*. New York : Wiley, 2nd ed.

Lemaire, J. (1995). *Bonus-Malus Systems in Automobile Insurance*. Boston : Kluwer Academic Publishers.

Nelsen, R. (1999). *An Introduction to Copulas*. Springer-Verlag, New York.

Scollnik, D. (2001). « Actuarial Modeling With MCMC and BUGS ». *North American Actuarial Journal*, 5 : 96-125.

Smith, A. et Roberts, G. (1993). « Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods » (avec discussion). *Journal of the Royal Statistical Society, Series B*, 55 : 3-23.

Young, V. (1998). « Credibility using a Loss Function from Spline Theory: Parametric Models with a One-Dimensional Sufficient Statistic ». *North American Actuarial Journal*, 2 : 101-117.

Young, V. et De Vylder, F. (2000). « Credibility in Favor of Unlucky Insureds. *North American Actuarial Journal*, 4 : 107-113.

Notes

1. À l'exception peu vraisemblable où le portefeuille analysé ne comprendrait que des assurés ayant réclamé au maximum une fois par année.

2. Voir Denuit *et al.* (2007) pour une explication plus détaillée de ce système.