

MÉTHODES PAR RÉGRESSION

KOROTOUMOU TRAORÉ, FRANCK VERMET

1. INTRODUCTION

Dans l'onglet "Régression" sont implémentées deux méthodes pour calculer le montant de provisions pour sinistres à payer, ainsi que les marges d'erreur associées :

- La méthode de Christophides,
- La méthode de GLM Poisson surdispersé.

Ces deux méthodes stochastiques sont basées sur des modèles paramétriques : les variables considérées sont les incréments, notés $X_{i,j}$, qui sont supposés indépendants. L'objectif est d'estimer les paramètres des lois choisies pour les v.a. $X_{i,j}$ et d'en déduire des informations sur les provisions par année de survenance : $R_i = \sum_{j=n-i+2}^n X_{i,j}$ et la provision totale $R = \sum_{i=2}^n R_i$.

Un point commun de ces méthodes est qu'elles peuvent être calibrées de façon similaire par régression.

Les références mentionnées dans la bibliographie ([1], [2], [3], [4], [5], [6], [7], [8], [9]) ont été utilisées pour cet onglet.

2. MODÈLE DE CHRISTOPHIDES

Le modèle de Christophides (1990) suppose que les incréments sont indépendants et suivent une loi lognormale, plus précisément que $Y_{ij} = \log(X_{ij})$ est de loi normale $\mathcal{N}(\alpha_i + \beta_j, \sigma^2)$. Ainsi les paramètres de la loi de X_{ij} dépendent à la fois de l'année de survenance i et de l'année de développement j .

Les paramètres peuvent être estimés facilement à partir du triangle supérieur par la méthode des moindres carrés (régression multilinéaire ordinaire), le modèle s'écrivant :

$$Y_{ij} = \sum_{k=1}^n \alpha_k \mathbf{1}_k(i) + \sum_{k=1}^n \beta_k \mathbf{1}_k(j) + \varepsilon_{i,j} = \alpha_i + \beta_j + \varepsilon_{i,j},$$

avec $\varepsilon_{i,j}$ de loi $\mathcal{N}(0, \sigma^2)$, où $\mathbf{1}_k(i) = 1$ si $k = i$ et 0 sinon. Pour des raisons d'identifiabilité du modèle, nous posons $\alpha_1 = \beta_1 = 0$.

A partir de ces estimations, on déduit aisément les estimations ponctuelles des provisions futures :

$$\hat{X}_{i,j} = \hat{\mathbb{E}}(X_{i,j}) = e^{\hat{\alpha}_i + \hat{\beta}_j + \frac{\hat{\sigma}^2}{2}} \text{ et } \hat{R} = \sum_{i=2}^n \sum_{j=n-i+2}^n \hat{X}_{i,j},$$

ainsi que l'estimation des variances :

$$\hat{V}(X_{i,j}) = e^{2(\hat{\alpha}_i + \hat{\beta}_j) + \hat{\sigma}^2} (e^{\hat{\sigma}^2} - 1) \text{ et } \hat{V}(R) = \sum_{i=2}^n \sum_{j=n-i+2}^n \hat{V}(X_{i,j}).$$

3. MODÈLE GLM POISSON SURDISPERSÉ

Le modèle GLM Poisson surdispersé a été introduit dans le contexte du provisionnement par Renshaw et Verrall en 1998 [9] et il a la particularité de donner les mêmes valeurs pour les provisions futures que le modèle de Chain Ladder. Il est par exemple décrit dans les mémoires de J.P. Boisseau [2], N.A. Dinh et G. Chau [6] et G.H. Phan Ngoc [8].

Soient $(Y_i, X_i^{(1)}, \dots, X_i^{(p)})$ des vecteurs aléatoires indépendants. Les modèles GLM, introduits en 1972 par J. Nedler et R. Wedderburn, supposent que les lois conditionnelles des variables à expliquer Y_i sachant les variables explicatives $X_i = (X_i^{(1)}, \dots, X_i^{(p)})$ suivent une loi dont la fonction de densité $f(x, \mu_i, \varphi)$ (par rapport à la mesure de référence, qui est la mesure de Lebesgue dans le cas continu et la loi de comptage dans le cas discret) appartient à la famille exponentielle :

$$f(x, \mu_i, \varphi) = \exp\left(\frac{\mu_i x - b(\mu_i)}{a(\varphi)} + c(x, \varphi)\right),$$

où :

- les μ_i sont des paramètres de la loi,
- φ est un paramètre de dispersion,
- b est une fonction suffisamment régulière de classe C^2 deux fois dérivables (par exemple $b(u) = \exp(u)$ pour la loi de Poisson et $b(u) = \exp(u^2/2)$ pour la loi normale),
- c est une fonction ne dépendant pas de μ_i .

On peut montrer que $m_i = \mathbb{E}[Y_i|X_i] = b'(\mu_i)$ et $\mathbb{V}(Y_i|X_i) = a(\varphi)b''(\mu_i)$. On appelle fonction de variance $V(m_i) = b''(\mu_i) = b''(b'^{-1}(m_i))$, qui lie l'espérance et la variance de la loi, puisque $\mathbb{V}(Y_i|X_i) = a(\varphi)V(m_i)$.

On définit la fonction de lien canonique $g = b'^{-1}$ telle que $g(m_i) = \mu_i$ et on suppose que $g(m_i) = \sum_{j=1}^p \beta_j X_i^{(j)}$. L'estimation des paramètres β_j se fait par maximum de vraisemblance.

Pour Y de loi de Poisson $\mathcal{P}(\lambda)$, on a pour $x \in \mathbb{N}$,

$$f(x, \lambda) = e^{-\lambda} \frac{\lambda^x}{x!} = \exp(-\lambda + x \log(\lambda) - \log(x!)).$$

Ainsi, on peut poser $\mu = \log(\lambda)$, $a(\varphi) = 1$, $b(\mu) = \exp(\mu) = \lambda$, $c(x, \varphi) = -\log(x!)$. On a alors $g(\cdot) = \log(\cdot)$ et on retrouve $\mathbb{E}[Y] = \mathbb{V}(Y) = \lambda$.

Dans le cas de la régression GLM Poisson, on a alors $\mathbb{E}[Y_i|X_i] = \mathbb{V}[Y_i|X_i] = \lambda_i = \exp(\sum_{j=1}^p \beta_j X_i^{(j)})$.

Dans le cas qui nous intéresse ici, Renshaw et Verrall supposent que les incréments X_{ij} sont indépendants, de moyenne $m_{i,j}$, avec $m_{i,j} = \mu + \alpha_i + \beta_j$, où $\alpha_1 = \beta_1 = 0$. Il s'agit du modèle GLM Poisson où les variables à expliquer sont les X_{ij} et les variables explicatives les indicatrices $\mathbf{1}_k(i)$ et $\mathbf{1}_k(j)$.

Il peut sembler curieux de modéliser des montants par une loi de Poisson, qui ne charge que les valeurs entières. En fait, Renshaw et Verrall montrent que sous l'hypothèse $\sum_{i=1}^{n-j+1} X_{ij} \geq 0$, pour tout j , l'estimation des paramètres par vraisemblance conditionnelle et la prédiction des valeurs futures conduisent aux mêmes formules que la méthode de Chain Ladder, ce qui légitime le fait d'utiliser cette approche, même pour des tableaux à valeurs non entières.

Par contre, une contrainte forte pour la loi de Poisson est que l'espérance et la variance sont égales. Si cette hypothèse n'est pas vérifiée par les données, on peut considérer la distribution de Poisson surdispersée (ou quasi-Poisson), qui n'est pas une vraie loi, mais vérifie $\mathbb{V}(X) = \Phi \mathbb{E}[X]$, où $\Phi > 1$

est un paramètre de surdispersion. Plus précisément, X est dite de loi de Poisson surdispersée de paramètres (λ, Φ) si X/Φ est de loi $\mathcal{P}(\lambda/\Phi)$.

Le modèle GLM avec loi de Poisson surdispersée conduit alors dans notre cas aux hypothèses suivantes : les incréments $X_{i,j}$ sont indépendants et de loi de Poisson surdispersée de paramètres $(\mu_{i,j}, \Phi)$, avec $\Phi > 0$. Il existe des paramètres réels $\mu, \alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n$ tels que $\mu_{i,j} = \exp(\mu + \alpha_i + \beta_j)$, avec les conditions d'identifiabilité $\alpha_1 = \beta_1 = 0$. Dans ce cas, les paramètres du modèle sont estimés par maximisation de la quasi-vraisemblance (on peut utiliser `glm(., family=quasipoisson)` dans R).

On a ensuite les estimations ponctuelles des provisions futures :

$$\hat{X}_{i,j} = \hat{\mathbb{E}}(X_{i,j}) = e^{\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j} \text{ et } \hat{R} = \sum_{i=2}^n \sum_{j=n-i+2}^n \hat{X}_{i,j},$$

ainsi que l'estimation des variances :

$$\hat{\mathbb{V}}(X_{i,j}) = \hat{\Phi} e^{\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j} \text{ et } \hat{\mathbb{V}}(R) = \sum_{i=2}^n \sum_{j=n-i+2}^n \hat{\mathbb{V}}(X_{i,j}).$$

Par additivité de la loi de Poisson surdispersé, nous obtenons sous l'hypothèse d'indépendance des incréments que R suit une loi de Poisson surdispersé de paramètres $(\sum_{i=2}^n \sum_{j=n-i+2}^n \mu_{i,j}, \Phi)$. Nous pourrions à partir de là utiliser les estimations des paramètres pour en déduire un quantile pour la provision totale R . Cependant, il faut aussi tenir compte de l'erreur d'estimation de R et pas seulement de sa variance pour quantifier la volatilité totale.

4. ESTIMATION DE LA VOLATILITÉ DES RÉSERVES

Comme nous l'avons vu dans l'onglet sur le modèle de Mack, l'erreur de prédiction se mesure par la *mean square error prediction* (MSEP) :

$$MSEP(\hat{R}) = \mathbb{E}[(\hat{R} - R)^2 | D],$$

où $D = \{C_{i,j}, i + j \leq n + 1\}$. Ce terme peut se décomposer en deux, par approximation :

$$MSEP(\hat{R}) \simeq \mathbb{E}[(\hat{R} - \mathbb{E}[R])^2 | D] + \mathbb{E}[(R - \mathbb{E}[R])^2 | D],$$

où le premier terme est l'erreur d'estimation, car \hat{R} a été estimé à partir du triangle supérieur, et le second terme est la variance classique de la variable R , appelée *process variance*. On peut approcher l'erreur d'estimation par la variance de \hat{R} , d'où la formule :

$$MSEP(\hat{R}) \simeq \mathbb{V}(\hat{R}) + \mathbb{V}(R),$$

où nous avons omis le conditionnement par rapport à D dans la notation de la variance.

Nous avons vu dans les deux sections précédentes comment estimer $\mathbb{V}(R)$ pour les méthodes de Christophides et GLM Poisson surdispersé.

Calculer l'erreur d'estimation est moins évident. Une façon de l'estimer est d'utiliser une méthode de Bootstrap : le principe est de simuler de nouvelles données similaires aux données observées, à partir desquelles on déduit un échantillon de réalisations de \hat{R} . On peut alors calculer une variance empirique comme estimateur de $\mathbb{V}(\hat{R})$.

La simulation se fait de la façon suivante, par re-échantillonnage des résidus. Nous l'écrivons dans le cas du modèle GLM Poisson surdispersé, il se fait de manière similaire pour le modèle de Christophides.

- (1) Estimation initiale des paramètres sur le triangle supérieur et déduction d'un triangle supérieur d'incrémentes moyens $\hat{\mu}_{i,j}$.
- (2) Calcul des résidus de Pearson : $r_{i,j} = \frac{X_{i,j} - \hat{\mu}_{i,j}}{\sqrt{\hat{\mu}_{i,j}}}$.
- (3) Boucle à réaliser M fois : pour $k = 1, \dots, M$,
 - (a) Re-échantillonnage des résidus et obtention d'un nouveau triangle supérieur de résidus $r_{i,j}^{(k)}$.
 - (b) Calcul d'un nouveau triangle supérieur d'incrémentes : $X_{i,j}^{(k)} = \hat{\mu}_{i,j} + r_{i,j}^{(k)} \sqrt{\hat{\mu}_{i,j}}$.
 - (c) Utilisation du modèle pour re-estimer les paramètres et calcul d'une nouvelle provision totale $R^{(k)}$.
- (4) Estimation de $\mathbb{V}(\hat{R})$ empiriquement à partir de l'échantillon $\{R^{(k)}, k = 1, \dots, M\}$.

Cette méthode est dite semi-simulatoire, car la MSEP est obtenue en additionnant la variance de R (*process variance*), obtenue en section 3, et la variance de \hat{R} (erreur d'estimation), obtenue par simulation. Dans l'application, pour des raisons de temps d'exécution, la valeur de M est fixée à 1000.

Il peut arriver que des incréments obtenus dans le point (3.b) de l'algorithme de bootstrap soient négatifs. Dans ce cas, le triangle est rejeté et un nouveau re-échantillonnage des résidus est réalisé, car le modèle GLM Poisson surdispersé ne peut s'appliquer dans ce cas. Cela peut avoir comme incidence de surestimer les quantiles inférieurs, cependant ce sont plutôt les quantiles élevés qui nous intéressent en provisionnement.

Une fois obtenues les estimations de la provision totale moyenne et de la volatilité totale, il est possible de calculer un quantile, en choisissant de façon usuelle comme loi de référence une loi normale ou lognormale. C'est ce que nous proposons dans l'application.

RÉFÉRENCES

- [1] Y. Angoua. *Provisionnement non-vie*. Support de cours de Master 2 de l'EURIA, 2017.
- [2] J. Boisseau. *Solvabilité 2 et mesure de volatilité dans les provisions pour sinistres à payer*. Mémoire d'actuariat du CEA, 2010.
- [3] C. Bonnefoy and D. Devictor. *Calcul stochastique de provisions techniques d'assurance non-vie*. Mémoire d'actuariat de l'ENSAE, 2007.
- [4] A. Charpentier. *Computational Actuarial Science with R*. CRC Press, 2012.
- [5] A. Charpentier and C. Dutang. *L'actuariat avec R*. 2012.
- [6] N. Dinh and G. Chau. *Mesures de provision cohérentes et méthodes lignes à lignes pour des risques non-vie*. Mémoire d'actuariat de l'Université de l'ENSAE, 2012.
- [7] M. Louis. *Evaluation des provisions techniques non-vie dans le contexte de Solvabilité II*. Mémoire d'actuariat de l'Université de Strasbourg, 2010.
- [8] H. Phan Ngoc. *Provisionnement stochastique adapté aux spécialités de la réassurance non-vie*. Mémoire d'actuariat de l'ISFA, 2015.

- [9] A. Renshaw and J. Verral. A stochastic model underlying the chain-ladder technique. *British Actuarial Journal*, 4(4) :903–923, 1998.

(Korotoumou Traoré, Franck Vermet) EURIA, UNIVERSITÉ DE BRETAGNE OCCIDENTALE, 6, AVENUE VICTOR LE GORGEU, CS 93837, F-29238 BREST CEDEX 3, FRANCE
E-mail address, Franck Vermet: `Franck.Vermet@univ-brest.fr`