Provisionnement non vie

13 mai 2022

M1 Actuariat, année 2021 - 2022

Durée : 2h
La calculatrice est autorisée ainsi qu'une feuille manuscrite *seulement recto.*
Toutes les réponses doivent être soigneusement justifiées.

Exercice Nous considérons le triangle de paiements de sinistres incrémentaux, effectués par une compagnie d'assurance au titre des exercices de 1990 à 1999 :

	1	2	3	4	5	6	7	8	9	10
1990	5947	3721	896	208	206	63	65	15	11	16
1991	6347	3246	723	152	68	37	52	12	11	
1992	6269	2976	847	263	153	65	54	9		
1993	5863	2683	723	190	133	89	43			
1994	5779	2745	654	273	231	105				
1995	6185	2828	573	245	105					
1996	5600	2893	564	225						
1997	5288	2440	528							
1998	5291	2358								
1999	5676									

Le triangle des paiements cumulés est également disponible :

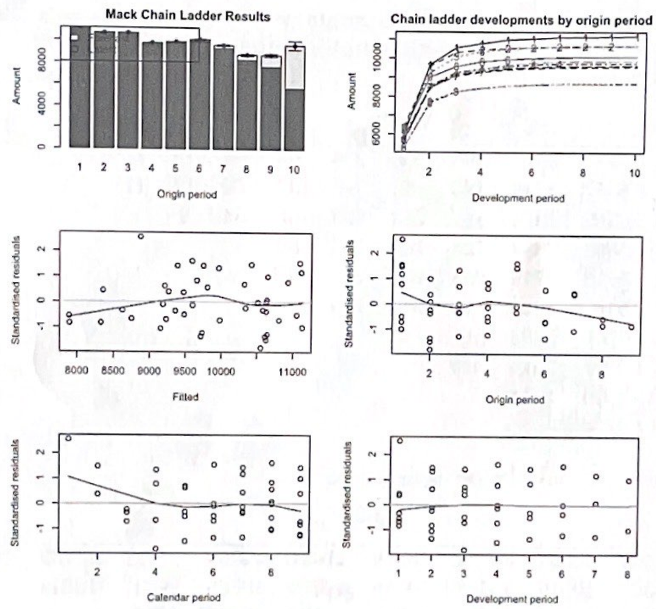
	1	2	3	4	5	6	7	8	9	10
1990	5947	9668	10564	10772	10978	11041	11106	11121	11132	11148
1991	6347	9593	10316	10468	10536	10573	10625	10637	10648	
1992	6269	9245	10092	10355	10508	10573	10627	10636		
1993	5863	8546	9269	9459	9592	9681	9724			
1994	5779	8524	9178	9451	9682	9787				
1995	6185	9013	9586	9831	9936					
1996	5600	8493	9057	9282						
1997	5288	7728	8256							
1998	5291	7649								
1999	5676									

À l'aide du logiciel R, nous avons estimé plusieurs modèles stochastiques dont les résultats sont présentés ci-dessous :

1) modèle de Mack (*N.B. les colonnes* f_j *et* σ_j^2 *ont été rajoutées*)

	Latest	Dev.To.Date	Ultimate	IBNR	Mack.S.E	CV(IBNR)	f_j	sigma^2_j
1	11,148	1.000	11,148	0.0	0.000	NaN	1.492496	18.29054
2					0.493			1.14097
3	10,636	0.998	10,662	26.1	0.646	0.0248	1.022862	0.2471363
4	9,724	0.996	9,759	34.7	3.025	0.0873	1.014850	0.394737
5	9,787	0.991	9,872	85.1	7.445	0.0875	1.006999	0.08638964
6	9,936	0.984	10,093	156.5	33.158	0.2118	1.005111	0.00379374
7	9,282	0.970	9,568	286.2	73.453	0.2566	1.001113	0.000691218
8	8,256	0.948	8,705	449.2	85.315	0.1899	1.001011	1.101271e-05
9	7,649	0.880	8,692	1,043.5	134.230	0.1286	1.001437	1.75458e-07
10	5,676	0.590	9,627	3,951.1	410.774	0.1040		

	Totals
Latest:	92,742.00
Dev:	0.94
Ultimate:	98,789.65
IBNR:	6,047.65
Mack.S.E	462.82
CV(IBNR):	0.08



i) Calculer les quantités manquantes de la ligne 2 du tableau ci-dessus. Expliquer ce que chacune des quantités calculées représente et en donner une interprétation.

ii) Pourquoi $Mack.S.E = 0$ pour l'année 1 ? Comment interpréter le paramètre de volatilité $\hat{\sigma}_j^2$ grâce à la réécriture du modèle de Mack ?

iii) Quelle information nous apporte chacun des graphiques ci-dessus (10 lignes maximum) ?

iv) La provision semble suivre une loi asymétrique. Calculer un intervalle de confiance pour R_2, la provision pour les sinistres survenus l'année d'origine 2. Quelle est la provision suffisante dans 95% des cas pour les sinistres survenus en année 2 ?

v) Dans ce modèle, les estimateurs des facteurs de développement sont sans biais et non corrélés. Le montrer.

vi) Quelles sont les deux composantes de l'erreur quadratique moyenne de prédiction conditionnelle ?

2) modélisation LogNormale. Nous avons obtenu le tableau ci-dessous :

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   8.8481     0.1074  82.352  < 2e-16 ***
anF2         -0.2985     0.1049  -2.846  0.00726 **
anF3         -0.1007     0.1097  -0.918  0.36473
anF4         -0.1591     0.1149  -1.385  0.17469
anF5          0.0369     0.1212   0.305  0.76246
anF6         -0.2091     0.1291  -1.620  0.11403
anF7         -0.2069     0.1398  -1.480  0.14756
anF8         -0.3389     0.1558  -2.176  0.03621 *
anF9         -0.3209     0.1831  -1.752  0.08826 .
anF10        -0.2041     0.2470  -0.826  0.41422
devF2        -0.7150     0.1049  -6.818 5.69e-08 ***
devF3        -2.1705     0.1097 -19.790  < 2e-16 ***
devF4        -3.3274     0.1149 -28.953  < 2e-16 ***
devF5        -3.7989     0.1212 -31.354  < 2e-16 ***
devF6        -4.5296     0.1291 -35.093  < 2e-16 ***
devF7        -4.7396     0.1398 -33.900  < 2e-16 ***
devF8        -6.2516     0.1558 -40.137  < 2e-16 ***
devF9        -6.3009     0.1831 -34.405  < 2e-16 ***
devF10       -6.0755     0.2470 -24.592  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2225 on 36 degrees of freedom
Multiple R-squared:  0.9921,Adjusted R-squared:  0.9882
F-statistic: 251.9 on 18 and 36 DF,  p-value: < 2.2e-16
```

ce qui nous a permis de calculer les quantités suivantes :

	R_i	sep(R_i)	CV(R_i)
1	0	0	NaN
2	12.16825	8.598690	0.7066
3	26.66672	10.585304	0.3969
4	36.88204	9.95484	0.2699
5	109.5962	12.231702	0.1116
6	148.1379	9.446053	0.0637
7	278.3836	9.467239	0.0340
8	426.4271	8.241397	0.0193
9	1024.965	8.398766	0.0081
10	3998.089	9.495609	0.0023

R	6061.316
sep(R)	31.31507
CV(R)	0.0051

a) Détailler le calcul de la provision pour l'année d'origine 3.

b) Comment avons-nous calculé $sep(R_i)$? Et $sep(R)$?

c) Avez-vous un commentaire à faire sur l'applicabilité de cette méthode sur ce triangle ?

3) méthode Bootstrap avec modèle Log Gamma (loi de Gamma et fonction lien ln) :

	Latest	Mean Ultimate	Mean IBNR	IBNR.S.E	IBNR 75%	IBNR 95%
1	11,148	11,148	0.0	0.0	0.0	0.0
2	10,648	10,664	15.9	21.8	27.0	56.4
3	10,636	10,661	25.2	25.2	38.8	72.8
4	9,724	9,759	35.3	28.5	50.9	87.3
5	9,787	9,872	84.9	43.7	109.2	169.8
6	9,936	10,093	157.2	55.3	191.0	254.3
7	9,282	9,568	285.6	73.8	333.0	419.1
8	8,256	8,709	453.2	87.0	508.2	607.3
9	7,649	8,695	1,045.7	140.6	1,129.9	1,286.2
10	5,676	9,620	3,944.2	336.8	4,168.1	4,506.1

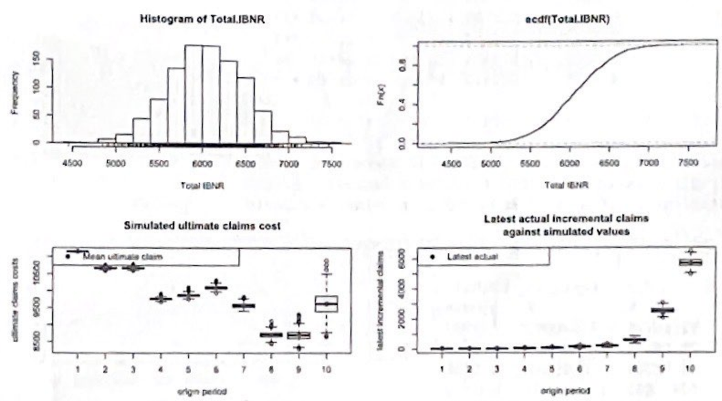
```
                    Totals
Latest:             92,742
Mean Ultimate:      98,789
Mean IBNR:           6,047
IBNR.S.E               430
Total IBNR 75%:      6,338
Total IBNR 95%:      6,717
```

a) Pouvez-vous détailler les différentes étapes qui ont permis le calcul de la provision par la méthode Bootstrap avec modèle Log Gamma ?

b) Que représente *Mean Ultimate* (2ème colonne) ?

c) Donner la provision suffisante dans 95% des cas pour les sinistres survenus l'année 2 et comparer le résultat avec celui obtenu dans le modèle lognormale.

d) À l'aide des graphiques ci-dessous, que pouvez-vous conclure sur la performance de la méthode ? Pourquoi ?



e) On nous demande d'estimer un quantile extrême. La méthode Bootstrap est-elle adaptée pour ce type d'estimation ? Que proposez-vous ?

4) Au finale, au vu des résultats obtenus, quelle méthode stochastique, parmi celles vues plus haut, auriez-vous adopté ? Pourquoi ? Argumenter.

5) Nous disposons maintenant d'une information supplémentaire représentée par le volume de prime (acquise) par année d'origine ainsi que d'un avis d'expert concernant l'évolution des rapports sinistres à primes sur les prochaines annés. Quelle méthode proposez-vous afin de calculer la provision ? La détailler (sans faire les calculs).

Exercice :

1) i) Ligne 2 :

Latest : $C_{2,9} = 10\,648$ ← dernier montant connu pour les sinistres survenus l'année 2.

Dev. to. date $= \frac{\text{Latest}}{\text{Ultimate}} = 0{,}99857$ ← Cadence de règlement pour l'année 2.

Ultimate $= C_{2,10} = \hat{f}_9 \, C_{2,9} = 10663{,}3$ ← Charge ultime pour l'année 2 estimée par le modèle de Mack.

IBNR : $C_{2,10} - C_{2,9} = 15{,}3$ ← provis° estimée pour l'année 2.

$CV(IBNR) = \frac{\text{Mack.S.E}}{\text{IBNR}} = 0{,}0322$ ← Coeff de variat°

$$\hat{f}_j = \frac{\sum_{i=0}^{m-j-1} C_{i,j+1}}{\sum_{i=0}^{m-j-1} C_{ij}} \implies \hat{f}_2 = 0{,}972712$$ ← facteur de dvpt pour l'année 2.

ii) Car $\text{Mack.S.E}_i = \sqrt{MSEP(\hat{R}_i)} = \hat{C}_{im} \sqrt{\sum_{j=m-i+1}^{m-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2}\left[\frac{1}{\hat{C}_{ij}} + \frac{1}{\sum_{k=1}^{m-j} C_{kj}}\right]}$

$\implies \text{Mack.S.E}_1 = 0$

Ici,

$$\hat{\sigma}_j^2 = \begin{cases} \frac{1}{m-j-1} \sum_{i=1}^{m-j} C_{ij} \left(\frac{C_{i,j+1}}{C_{ij}} - \hat{f}_j\right)^2 & \text{si } j \leq m-2 \\ \min\left\{\frac{\hat{\sigma}_{m-2}^4}{\hat{\sigma}_{m-3}^2} ; \min\left(\hat{\sigma}_{m-3}^2 ; \hat{\sigma}_{m-2}^2\right)\right\} & \text{si } j = m-1 \end{cases}$$

On peut réécrire le modèle de Mack comme : $C_{i,j+1} = f_j C_{ij} + \sigma_j^2 \sqrt{C_{ij}}\, \varepsilon_{ij}$

L'interprétation économétrique est que le 1er terme correspond à une prévision et le second à l'erreur.

Ici σ_j^2 représente la variance estimée de l'erreur

iii) Ok

iv) $IC_{R_2} = [\hat{R}_2 \pm 2 se(\hat{R}_2)]$
$= [6047{,}65 \pm 2 \times 462{,}82]$
$= [5122{,}01 ; 6973{,}29]$

$R_{2,95\%} = \hat{R}_2 + 1{,}96\, se(\hat{R}_2) = 6954{,}78$

v) Notons $T_j = \{ C_{ik} \mid k \leq j,\ i+j \leq n+1 \}$ $\quad j \in [\![1,n]\!]$

On a $\mathbb{E}[C_{i,j+1} \mid T_j] = \mathbb{E}[C_{i,j+1} \mid C_{i1}, \dots, C_{ij}]$
$= f_j C_{ij}$

$$\mathbb{E}[\hat{f}_j \mid T_j] = \mathbb{E}\left[\frac{\sum_{i=1}^{n-j} C_{i,j+1}}{\sum_{i=1}^{n-j} C_{ij}} \,\middle|\, T_j \right]$$

$$= \frac{\sum_{i=1}^{n-j} \mathbb{E}[C_{i,j+1} \mid T_j]}{\sum_{i=1}^{n-j} C_{ij}}$$

$$= f_j$$

$\Rightarrow$ Par conditionnement, $\mathbb{E}[\hat{f}_j] = \mathbb{E}[\mathbb{E}[\hat{f}_j \mid T_j]] = f_j$

Donc $\hat{f}_j$ est donc un estimateur sans biais de f_j.

De +, pour $j < k$,

$$\mathbb{E}[\hat{f}_j \hat{f}_k] = \mathbb{E}[\mathbb{E}[\hat{f}_j \hat{f}_k \mid T_k]]$$

$$= \mathbb{E}[\hat{f}_j \, \mathbb{E}[\hat{f}_k \mid T_k]]$$

$$= \mathbb{E}[\hat{f}_j] f_k$$
$$= \mathbb{E}[\hat{f}_j] \mathbb{E}[\hat{f}_k]$$

Donc les estimateurs de facteur de dvpt sont sans biais et non corrélés.

vi) $MSECP(\hat{R}_i) = \mathbb{E}[(\hat{R}_i - R_i)^2 \mid T_i]$

$\Rightarrow MSECP(\hat{R}_i) = MSE(\hat{R}_i) + Var(R_i)$

2) a) $\hat{R}_3 = \sum_{i+j-1>n} \hat{\mu}_{ij}$ $\quad \hat{\mu}_{ij} = \exp\left(\mu + \alpha_i + \beta_j + \frac{\sigma^2}{2}\right)$ $\quad i = 3,\ j \in \{9, 10\},\ n = 10$

$\Rightarrow \hat{R}_3 = \exp\left(\mu + \alpha_3 + \frac{\sigma^2}{2}\right)\left[e^{\beta_9} + e^{\beta_{10}}\right]$

$= \exp\left(8{,}8481 - 0{,}1007 + \frac{0{,}2225^2}{2}\right)\left[e^{-6{,}3009} + e^{-6{,}0755}\right]$

$= 26{,}66728$

b) $Sep(R_i) = \sqrt{Var(R_i)} = \sqrt{Var(\sum X_{ij})}$
$= \sqrt{\sum Var(X_{ij})}$ $\quad$ et $Var(X_{ij}) = e^{2m_{ij} + \sigma^2}(e^{\sigma^2} - 1) = e^{2(\mu + \alpha_i + \beta_j + \frac{\sigma^2}{2})}(e^{\sigma^2} - 1)$

X_{ij} ⫫

Idem pour $Sep(R) = \sqrt{Var(R)}$

$$= \sqrt{Var\left(\sum_i R_i\right)}$$

$$= \sqrt{\sum_i Var(R_i)}$$

$$= \sqrt{\sum_i \sum_j Var(X_{ij})}$$

c) Les var anF4, anF5, anF6, anF7, anF9 et anF10 ont une p.Value $> 5\%$ $\Longrightarrow$ non significatives

3) a) Etapes de la procédure bootstrap

1) on estime les paramètres du modèle de régression, ce qui permet d'obtenir les valeurs prévues par le modèle $\hat{\mu}_{ij} = g^{-1}(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j)$ et le best estimate

$$\widehat{E(R)} = \hat{R} = \sum_{i+j>n} \hat{\mu}_{ij}$$

2) On calcule les résidus de Pearson (r_{ij}^P), $i + j \leq n$, par

$$r_{ij}^P = \frac{x_{ij} - \hat{\mu}_{ij}}{\sqrt{V(\hat{\mu}_{ij})}}$$

et on estime (si nécessaire) le paramètre de dispersion ϕ à l'aide des résidus de Pearson

$$\hat{\phi} = \frac{1}{t-p} \sum_{i+j \leq n} (r_{ij}^P)^2$$

On peut maintenant commencer le bootstrap :

3) on répète B fois $(b = 1, \dots, B)$ les opérations suivantes:
   - on obtient un échantillon bootstrap de résidus $(r_{ij}^{*b})_{i+j \leq n}$ par rééchantillonage des résidus initiaux du modèle
   - on détermine le triangle d'increments bootstrappés associés $(x_{ij}^{*b})_{i+j \leq n}$ par

$$x_{ij}^{*b} = \hat{\mu}_{ij} + r_{ij}^{*b} \sqrt{V(\hat{\mu}_{ij})}$$

   - à partir de ces nouvelles données de triangle supérieur, on obtient à nouveau les valeurs prévues $(\hat{\mu}_{ij}^{*b})_{i+j>n}$ et on calcule le best estimate $\hat{R}^{*b} = \sum_{i+j>n} \hat{\mu}_{ij}^{*b}$
   - on stocke $\hat{R}^{*b}$ et on répète le bootstrap

4) on peut maintenant utiliser le B- échantillon bootstrap $(\hat{R}^{*1}, \dots, \hat{R}^{*B})$ pour calculer par exemple les quantités suivantes :

- $E_{boot}(\hat{R}) = \frac{1}{B} \sum_{b=1}^{B} \hat{R}^{*b}$
- le risque d'estimation

$$MSE_{boot}(\hat{R}) = \frac{1}{B-1} \sum_{b=1}^{B} [\hat{R}^{*b} - E_{boot}(\hat{R})]^2$$

- le risque de processus

$$Var_{boot}(R) = \hat{\phi} \sum_{i+j>n} V(\hat{\mu}_{ij})$$

- l'erreur de prédiction

$$sep_{boot}(\hat{R}) = \sqrt{MSEPC_{boot}(\hat{R})} = \sqrt{MSE_{boot}(\hat{R}) + Var_{boot}(R)}$$

b) Moyenne des charges ultimes

c) $R_{2,95\%} = 56{,}4$

Dans modèle log normale, $R_{2,95\%} = \hat{R}_2 + 1{,}96 \times se(R_2)$

$$= 12{,}1825 + 1{,}96 \times 8{,}598690$$

$$= 29{,}02168$$

Sous Log Normale, $R_{2,95\%}$ est $51{,}46\%$ + faible qu'en dans modèle Log Gamma.

d) Je vois rien

e) Non car le Bootstrap travaille conditionnellement à l'échantillon observé

Or cette échantillon est éloigné des queues du distribution

$\longrightarrow$ Pas adapté

$\Longrightarrow$ Technique des valeurs extrêmes.

4) BLABLA

On choisit Bootstrap car lognormale faussée car Var non signif

et Mack moins bonne.

5) Bornhuetter-Ferguson

$$R_i^{BF} = (1-\hat{p}_{n-i})\,\underbrace{\hat{S}_i}_{\text{charge ultime apriori}} \qquad \text{avec } \hat{S}_i = \underbrace{\phi_i}_{\text{avis d'expert.}} P_i$$