Provisionnement non vie

16 mai 2019

M1 Actuariat, année 2018-2019

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

Exercice 1 Nous considérons le triangle de paiements de sinistres cumulés, effectués par une compagnie d'assurance pour la branche RC Auto au titre des exercices de 1997 à 2006 :

	1	2	3	4	5	6	7	8	9	10
1997	22456	44807	51281	55672	57529	60585	60834	60923	61346	62150
1998	22954	46370	52168	55262	57106	57997	58585	59516	59852	
1999	21312	42351	49409	52324	54429	57219	58122	58479		
2000	20626	45155	52033	55421	57381	57945	59931			
2001	21037	44210	50875	54902	58481	59641				
2002	22708	43188	50563	54112	57718					
2003	18416	36312	42044	44272						
2004	19479	36895	42699							
2005	18006	35055								
2006	17559									

À l'aide du logiciel R, nous avons estimé plusieurs modèles. Les résultats sont présentés ci-dessous.

1) Modèle de Mack. Nous présentons le tableau issu de la commande *MackChainLadder* en R (*dans lequel 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  62,150       1.000   62,150      0        0       NaN   2.001898    183.83541899
2                                          179             1.152625      8.96676545
3  58,479       0.981   59,619  1,140      189     0.166   1.067721      6.72115929
4  59,931       0.973   61,573  1,642      553     0.337   1.045625     13.78961345
5  59,641       0.958   62,252  2,611    1,048     0.401   1.029695     24.13833238
6  57,718       0.930   62,034  4,316    1,708     0.396   1.015940      9.96984134
7  44,272       0.890   49,753  5,481    1,747     0.319   1.007756      3.21898570
8  42,699       0.833   51,235  8,536    1,888     0.221   1.006302      0.05069475
9  35,055       0.723   48,483 13,428    1,961     0.146   1.013106      0.27154444
10 17,559       0.361   48,616 31,057    3,258     0.105   1.000000

              Totals
Latest:   497,356.00
Dev:            0.88
Ultimate: 566,350.45
IBNR:      68,994.45
Mack.S.E    6,174.29
CV(IBNR):       0.09
```

→ Mack présente une erreur à 9%

i) Compléter la ligne 2 du tableau ci-dessus tout en interprétant les valeurs obtenues.

ii) Pourquoi pour évaluer l'erreur calculons-nous un MSECP (*Mean Square Error of Conditional Prediction*) et pas simplement un MSE (*Mean Square Error*) ?

iii) Il est possible de réécrire le modèle de Mack pour en donner une interprétation économétrique. Que représente $\hat{\sigma}_j^2$ dans cette réécriture du modèle ?

iv) Nous présentons dans les tableaux ci-dessous les résultats obtenus pour les 3 modèles de régression linéaire que nous avons estimé avec les C-C points.

```
Call: lm(formula = cij[, 1] ~ cij[, 2])
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 5063.0414  3013.1602    1.68   0.1368
cij[, 2]       0.3778     0.0721    5.24   0.0012 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 878.6 on 7 degrees of freedom
Multiple R-squared:  0.7968,Adjusted R-squared:  0.7678
F-statistic: 27.46 on 1 and 7 DF,  p-value: 0.001199
```

$\longrightarrow$ p.value du test de significativité $H_0 : a_j = 0$ vs $H_1 : a_j \neq 0$

```
Call: lm(formula = cij[, 2] ~ cij[, 3])
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.155e+03  2.524e+03  -0.854    0.426
cij[, 3]     9.117e-01  5.147e-02  17.714 2.08e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 560.4 on 6 degrees of freedom
Multiple R-squared:  0.9812,Adjusted R-squared:  0.9781
F-statistic: 313.8 on 1 and 6 DF,  p-value: 2.079e-06
```

```
Call: lm(formula = cij[, 3] ~ cij[, 4])
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 4.039e+03  2.572e+03    1.57    0.177
cij[, 4]    8.606e-01  4.829e-02   17.82 1.02e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 481.6 on 5 degrees of freedom
Multiple R-squared:  0.9845,Adjusted R-squared:  0.9814
F-statistic: 317.6 on 1 and 5 DF,  p-value: 1.021e-05
```

Au vu de ces résultats serait-il plus raisonnable d'appliquer la méthode London Chain plutôt que la méthode de Mack ? Oui ? Non ? Pourquoi ?

v) Montrer que dans le modèle de Mack, l'estimateur du ratio sinistres à primes pour l'année d'origine i, $\widehat{\left(\frac{S}{P}\right)}_i = \frac{\hat{C}_{in}}{P_i}$, est un estimateur sans biais. Calculer ensuite l'erreur d'estimation.

2) Modélisation Log Poisson. Pour ce modèle, nous avons calculé les quantités suivantes :

	CV(R_i)
1	NaN
2	0.57
3	0.48
4	0.40
5	0.32
6	0.25
7	0.22
8	0.17
9	0.14
10	0.09

CV(R)
0.06

i) Donner l'estimation de la provision globale.

ii) Comparer les résultats du modèle log Poisson aux résultats obtenus par le modèle de Mack à la fois par année d'origine et globalement.

iii) Calculer le $MSECP(\hat{R}_i)$ dans un modèle log Poisson (on passera par la décomposition du MSECP en deux composantes et on fera l'hypothèse que $\hat{R}_i$ est un estimateur sans biais de R_i).

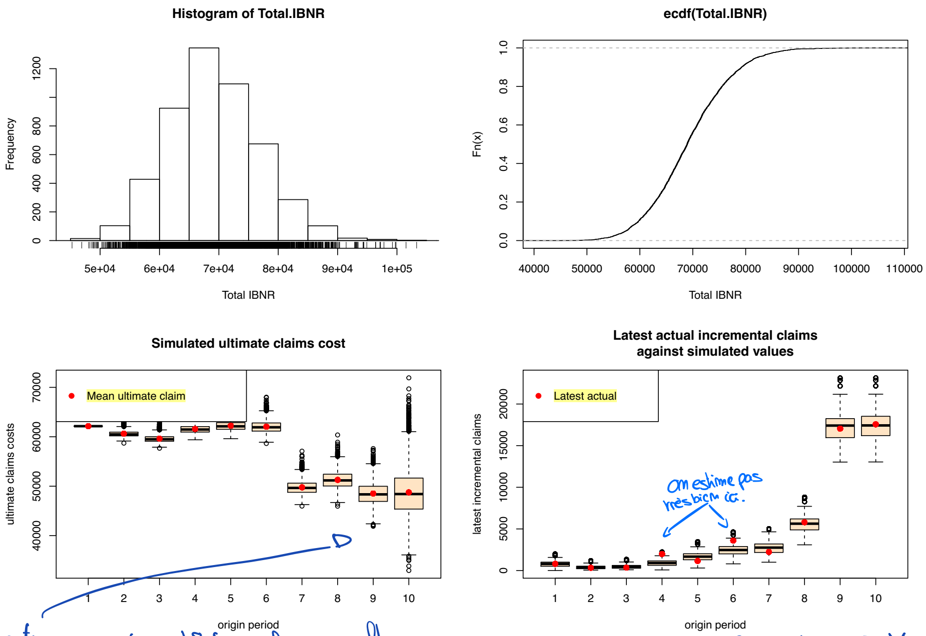
3) Bootstrap (B=5000) sur un modèle log-Gamma. Nous avons obtenu les résultats suivants :

	Latest	Mean Ultimate	Mean IBNR	IBNR.S.E	IBNR 75%	IBNR 95%
1	62,150	62,150	0	0	0	0
2	59,852	60,636	784	628	1,090	1,995
3	58,479	59,625	1,146	728	1,534	2,547
4	59,931	61,578	1,647	845	2,136	3,198
5	59,641	62,246	2,605	1,024	3,226	4,476
6	57,718	62,040	4,322	1,252	5,083	6,603
7	44,272	49,770	5,498	1,394	6,346	7,990
8	42,699	51,283	8,584	1,761	9,730	11,639
9	35,055	48,517	13,462	2,312	14,930	17,581
10	17,559	48,758	31,199	4,887	34,093	39,966

	Totals
Latest:	497,356
Mean Ultimate:	566,602
Mean IBNR:	69,246
IBNR.S.E	7,642
Total IBNR 75%:	74,184
Total IBNR 95%:	82,184

Hist pour l'estimat° de la prov globale $\hat{R}$

fdr empirique de R.



Charge ult. →

on estime pas très bien ici.

Variabilité + importante pour les années les + récentes car - de données

Estimat° des charges ult.

Sinon pour les autres ça va.

a) Au vu des graphiques, pouvons-nous considérer cette méthode satisfaisante ?

b) Les résultats obtenus par Bootstrap sont-ils meilleurs par rapport à ceux obtenus par les deux méthodes précédentes ?

c) Pourquoi est-il important pour une compagnie d'assurance de ne pas sous-évaluer ni sur-évaluer la provision ? Est-il plus gênant de sur ou sous évaluer la provision à votre avis ?

d) Pourquoi la technique du bootstrap est-elle couplée aux MLG afin de calculer une provision ?

e) Le Bootstrap est-il adapté pour l'estimation d'un quantile très élevé ? Oui ? Non ? Pourquoi ?

4) Nous disposons maintenant d'un triangle tronqué : on suppose que seulement les paiement pour les sinistres survenus de 2002 à 2006 ont été observés (on ignore les paiements qui ont été effectués de 1997 à 2001). Par contre, les sinistres continuent à se développer sur 10 ans. Que faire pour calculer les ultimes à 10 ans et la provision ? Sans faire les calculs détailler la méthode que vous proposez d'utiliser.

Exercice 1 :

1) i) Sur la ligne 2,

Latest $= C_{2,9} = 59852$ — Cela correspond au dernier montant connu pour les sinistres survenus l'année 2 (ici 1989).

Dev.To.Date $= \dfrac{\text{Latest}}{\text{Ultimate}} = 0{,}98706$ — Cela correspond au vecteur des cadences de règlement

Ultimate $= C_{2,10} = C_{2,9} f_9 = 60636{,}42$ — Cela correspond au vecteur des charges ultimes estimées par le modèle de Mack.

IBNR = Ultimate − Latest = 427 — Cela correspond au vecteur des provis° estimées par année d'origine.

$$\text{MackSE} = \sqrt{MSEP(\hat{R}_i)} = \sqrt{C_{2,10}^2 \frac{\sigma_9^2}{f_9^2}\left[\frac{1}{C_{2,9}} + \frac{1}{C_{1,9}}\right]} = 179$$

Variat° de l'erreur de l'estimat°

$$CV(IBNR) = \frac{\text{MackSE}}{\text{IBNR}} = 0{,}2283$$

Coeff de variation

$$\widehat{MSEP}(\hat{R}_i) = \hat{C}_{in}^2 \sum_{j=n-i+1}^{n-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2}\left[\frac{1}{\hat{C}_{ij}} + \frac{1}{\sum_{k=1}^{n-j} C_{kj}}\right]$$

$$f_j = \frac{\sum_{i=0}^{n-j-1} C_{i,j+1}}{\sum_{i=0}^{n-j-1} C_{ij}}$$

$$\hat{\sigma}_j^2 = \frac{1}{n-j-1}\sum_{i=0}^{j} C_{ij}\left(\frac{C_{ij+1}}{C_{ij}} - \hat{f}_j\right)^2$$

ii) Car on ne compare pas avec un paramètre mais une variable aléatoire.
On compare $\hat{R}_i$ avec R_i or avec le MSE on compare $\hat{R}_i$ avec le paramètre $\mathbb{E}[R_i]$.

iii) On peut le réécrire : $C_{ij+1} = \underbrace{f_j C_{ij}}_{\text{prévision}} + \underbrace{\sigma_j^2 \sqrt{C_{ij}}\,\varepsilon_{ij}}_{\text{erreur}}$ → économétrie

Où σ_j^2 représente la variance estimée de l'erreur
et $\hat{\sigma}_j^2 = \dfrac{1}{n-j-1}\displaystyle\sum_{i=1}^{n-j}\left(\frac{C_{ij+1} - f_j C_{ij}}{\sqrt{C_{ij}}}\right)^2$

Réécriture du modèle

$$C_{i,j+1} = f_j C_{ij} + \sigma_j^{\mathbf{2}}\sqrt{C_{ij}}\epsilon_{ij}$$

avec ϵ_{ij} i.i.d. centrés et de variance unitaire. On estime les paramètres inconnus par la méthode des MCP, i.e. en cherchant à résoudre :

$$\min \sum_{i=1}^{n-j} \frac{1}{C_{ij}}(C_{i,j+1} - f_j C_{ij})^2$$

Les résidus standardisés

$$\hat{\epsilon}_{ij} = \frac{C_{i,j+1} - \hat{f}_j C_{ij}}{\sqrt{C_{ij}}}$$

nous donnent une idée simple pour estimer le paramètre de volatilité

$$\hat{\sigma}_j^{\mathbf{2}} = \frac{1}{n-j-1}\sum_{i=1}^{n-j}\left(\frac{C_{i,j+1} - \hat{f}_j C_{ij}}{\sqrt{C_{ij}}}\right)^2$$

iv) p.value = 0.1368 $\Rightarrow$ Le test de significativité $\Rightarrow$ on accepte $H_0 \Rightarrow a_1 = 0$
= 13,68 % > 5 %

de même $a_2 = 0$ et $a_3 = 0$ $\Rightarrow$ Mack est bien adapté car l'ordonnée à l'origine n'est pas significativement $\neq$ de 0.

Rappel : en fonct° des p.value on pourrait coupler Mack et London Chain par certaines colonnes.

$\rightarrow$ Non car toutes les p.value > 5 % $\Rightarrow$ l'ordonnée à l'origine n'est pas significativement $\neq$ de 0.

v) Montrons que $\mathbb{E}\left[\frac{\hat{C}_{in}}{P_i}\right] = \frac{C_{in}}{P_i}$ et pour l'erreur d'estimation voir slide 22/54.

2) i) En Log-Poisson, la provision globale coïncide avec la provision globale de Mack

$\Rightarrow$

ii) 1 $\rightarrow$ Non car valeur connue $\Rightarrow$ pas d'erreur d'estimat°

Ceux de Log-Poisson sont + faibles, meilleure estimat° en Log-Poisson.

Dissocier années + récentes puis globalem^t.

iii) Slide 14

$\rightarrow MSECP(\hat{R}_i) = \mathbb{E}[(\hat{R}_i - R_i)^2 \mid T_i]$ et on a $\mathbb{E}[\hat{R}_i] = R_i$.

4) a) Globalem^t, elle semble être OK, mais à part que incertitude importante pour les années récentes et 4^e graph...

b) On a pas le coeff de variat° $\rightarrow$ on les calcule $\frac{\text{IBNR.S.E}}{\text{MEAN IBNR}}$

c) Sous-évalue $\rightarrow$ Pb pour faire face aux engagem^ts.
Sur-évaluer $\rightarrow$ argent non utilisé

+ gênant de sous-évaluer

d) Bootstrap s'applique sur échantillon iid
Nos paiements incrémentaux sont indep mais pas forcément iid
=> couplage MVG.

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

Or cet échantillon est éloigné des queues de distributᵒ → pas adapté => technique des valeurs extrêmes

4) Lissage pour f_5 à f_9 puis on extrapole.

TAIL FACTOR.

On veut calculer les ultimes or en l'état on ne peut pas calculer les facteurs de dvpt
→ On adopte des techniques de lissage
On calcule f_1 à f_4 avec nos données
→ On les lisse
Puis on peut extrapoler pour obtenir les facteurs de dvpt.

4) Nous disposons maintenant d'un triangle tronqué : on suppose que seulement les paiement pour les sinistres survenus de 2002 à 2006 ont été observés (on ignore les paiements qui ont été effectués de 1997 à 2001). Par contre, les sinistres continuent à se développer sur 10 ans. Que faire pour calculer les ultimes à 10 ans et la provision ? Sans faire les calculs détailler la méthode que vous proposez d'utiliser.

Exercice 1 :

1) i) Ligne 2 :

Latest : $C_{2,9} = 59852$ ← Dernier montant connu pour les sinistres de l'année 2 (ici 1998)

Dev. to. date : $\frac{\text{Latest}}{\text{Ultimate}} = \frac{59852}{60636{,}42} = 0{,}987$ ← Cadence de règlement pour l'année 2

Ultimate : $C_{2,10} = C_{2,9} \times f_9 = 59852 \times 1{,}013106$
$= 60636{,}42$ ← Charge ultime estimée par le modèle de Mack pour l'année 2.

$IBNR = \text{Ultimate} - \text{Latest} = 784{,}42$ ← provision estimée pour l'année 2.

$\text{Mack SE} = \sqrt{MSEP(\hat{R}_i)} = \sqrt{C_{2,10}^2 \frac{\sigma_9^2}{f_9^2}\left[\frac{1}{C_{2,9}} + \frac{1}{C_{1,9}}\right]} = 179$ ← Variat° de l'erreur de l'estimation.

Rappel : $\widehat{MSEP}(\hat{R}_i) = \hat{C}_{im}^2 \sum_{j=m-i+1}^{m-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2}\left[\frac{1}{\hat{C}_{ij}} + \frac{1}{\sum_{k=1}^{m-j} C_{kj}}\right]$

$CV(IBNR) = \frac{\text{Mack SE}}{IBNR} = 0{,}2282$ ← Coeff de variat°

$\hat{f}_j = \frac{\sum_{i=0}^{m-j-1} C_{i,j+1}}{\sum_{i=0}^{m-j-1} C_{i,j}}$ et $\hat{\sigma}_j^2 = \frac{1}{m-j-1}\sum_{i=0}^{j} C_{ij}\left(\frac{C_{ij+1}}{C_{ij}} - \hat{f}_j\right)^2$

ii) Car ici, on ne compare pas $\hat{R}_i$ avec le paramètre $E[R_i]$ (cas du MSE) mais on compare $\hat{R}_i$ avec la variable aléatoire R_i (Cas du MSECP).

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

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 avec $\hat{\sigma}_j^2 = \frac{1}{m-j-1}\sum_{i=1}^{m-j}\left(\frac{C_{ij+1} - \hat{f}_j C_{ij}}{\sqrt{C_{ij}}}\right)^2$

iv) p.value = 13,68% > 5% ⟹ Pour le test de significativité, on accepte $H_0 : a_1 = 0$

p.value = 42,6 % > 5% ⟹ Pour le test de significativité, on accepte $H_0 : a_2 = 0$

p.value = 17,7 % > 5% ⟹ Pour le test de significativité, on accepte $H_0 : a_3 = 0$

⟹ Mack est bien adapté car l'ordonnée à l'origine n'est pas significativement ≠ de 0.

Donc non car toutes les p.values sont > 5% ⟹ l'ordonnée à l'origine n'est pas significativem^t ≠ de 0.

v) Montrons que $\mathbb{E}\left[\frac{\hat{C}_{im}}{P_i}\right] = \frac{C_{im}}{P_i}$

$$\mathbb{E}\left[\frac{1}{P_i} C_{i,m-i+1} \hat{f}_{m-i+1} \cdots \hat{f}_{m-1}\right] = \frac{1}{P_i} C_{i,m-i+1} \mathbb{E}\left[\hat{f}_{m-i+1} \cdots \hat{f}_{m-1}\right]$$

$$= \frac{1}{P_i} C_{i,m-i+1} f_{m-i+1} - f_{m-1} = \frac{1}{P_i} C_{im}$$

$$\widehat{MSEP}\left(\left(\frac{\hat{S}}{P}\right)_i\right) = \left(\frac{1}{P_i}\right)^2 \widehat{MSEP}(\hat{C}_{im})$$

$$\frac{Mack\ SE}{R_i} = CV(R_i)$$

2) i) En log-Poisson, la provision global coïncide avec la provision globale de Mack. $\Rightarrow R = 68\,994$

ii) Pour l'année 1 $\rightarrow$ Pas d'erreur d'estimation $\Rightarrow CV(R_1) = \text{Nan}$.

Les CV en Log-Poisson sont + faibles $\Rightarrow$ meilleure estimat°

Toutefois il faut être conscient que les années récents ont un CV faible tandis que celui des années + ancienne est + grand. (Log Poisson mieux à partir de l'année 4)

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

$$= \mathbb{E}[(\hat{R}_i - \mathbb{E}[R_i])^2 | T_i] + \mathbb{E}[(R_i - \mathbb{E}[R_i])^2 | T_i]$$

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

Or $\hat{R}_i$ est un ESB de $R_i \Rightarrow MSE(\hat{R}_i) = 0$

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

3) a) Globalement cette methode semble satisfaisante mis à part qu'il y a une incertitude importante pour les années les + récentes et les Latest actual incremental claims pour années 4, 5 et 6 en dehors des box plots.

b) Les CV ne sont pas donnés, on les calcule avec $\frac{\text{IBNR.S.E}}{\text{MEAN IBNR}}$

$$\Rightarrow CV(R_2) = 0,801$$
$$CV(R_3) = 0,635$$
$$CV(R_4) = 0,513$$
$$CV(R_5) = 0,393 \quad X$$
$$CV(R_6) = 0,2897 \quad X$$
$$CV(R_7) = 0,25355 \quad X$$

$$CV(\hat{R}) = 0,11036$$

$CV(R_8) = 0{,}205$ ✗
$CV(R_9) = 0{,}1717$
$CV(R_{10}) = 0{,}1566$

Mack chain Ladder meilleur sauf pour années : 5,6,7,8 et Log Poisson meilleur partout.

⟶ Non

c) Si on sous évalue ⟹ Pb pour faire face au engagement, risque de défaut

sur évalue ⟹ Argent non utilisé.

Il est + gênant de sous-évaluer car ⟹ risques.

d) Car la méthode Bootstrap s'applique sur échantillon iid.

Or nos paiements incrémentaux sont ⊥ mais pas forcément iid
⟹ Couplage MLG.

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

Or cette échantillon est éloigné des queues de distribution
⟶ Pas adapté
⟹ Technique des valeurs extrêmes.

4) Lissage pour f_5 à f_9 puis on extrapole.

TAIL FACTOR.

On veut calculer les ultimes or en l'état on ne peut pas calculer les facteurs de dvpt
⟶ On adopte des techniques de lissage
On calcule f_1 à f_4 avec nos données
⟶ on les lisse
Puis on peut extrapoler pour obtenir les facteurs de dvpt.