

Mémoire de Master 1 Mathématiques Appliquées de l'Université de Paris Dauphine

Titre :

Provisionnement en assurance non-vie utilisant
les triangles des montants payés et encourus

Par : Keren Toledano - Waël Saadé - Pierre Lavieille

Directeur de Mémoire : Christophe Dutang
Numéro de groupe : A11

Confidentialité : ☒ Non ☐ Oui (Durée : ☐ 1 an ☐ 2 ans)

Table des matières

Table des matières	2
Introduction	3
1 La méthode classique Chain Ladder et le modèle de Mack	7
1.1 Présentation de la méthode déterministe Chain Ladder	7
1.1.1 Validation du modèle	8
1.1.2 Les limites de la méthode Chain Ladder	8
1.2 Le modèle de Mack et l'incertitude dans la prédiction	9
1.3 Mise en oeuvre de la méthode sur notre jeu de données	12
2 Munich Chain Ladder	18
2.1 Avant-propos	18
2.2 Explication de la méthode	20
2.2.1 Triangle des montants payés	21
2.2.2 Triangle des charges	23
2.3 Mise en oeuvre de la méthode sur notre jeu de données	25
3 Paid-incurred chain claims reserving method	30
3.1 Présentation du modèle lognormale	30
3.1.1 Modèle lognormale avec estimation fréquentiste	30
3.1.2 Modèle lognormale avec estimation bayésienne	31
3.1.3 Le cadre fréquentiste	32
3.1.4 La méthode PIC	35
3.1.5 Détails du résultat	36
3.1.6 Détails du calcul	39
3.2 Mise en oeuvre de la méthode PIC	39
3.2.1 Mise en oeuvre sur notre 1er jeu de données	39
3.2.2 Mise en oeuvre sur le 2 ème jeu de données	40
3.2.3 Focus sur le deuxième jeu de données	43
Conclusion	45
Bibliographie	47
A Annexe code R	48
A.1 Code R Méthode Chain Ladder	48
A.2 Code R Munich Chain Ladder	49
A.3 Focus sur le deuxième jeu de données	52

Introduction

Les provisions ont un poids essentiel dans le bilan d'un assureur. Elles représentent une part très importante du passif des compagnies d'assurance : effectivement, en 2002, au bilan de l'ensemble du marché français de l'assurance non-vie, les provisions techniques représentaient les $\frac{2}{3}$ du passif et la provision des sinistres en représentait 50 %. Leur évaluation a donc un fort impact sur le résultat de ces sociétés.

De plus, c'est sous contrainte première de devoir respecter leurs engagements vis-à-vis des assurés, que les sociétés évalueront leurs provisions techniques et plus précisément leurs réserves pour sinistres.

En effet, nous considérons un sinistre, contre lequel l'assuré est garanti pendant une certaine période, et survenu au cours de l'année n à une date dite de survenance.

C'est à compter de la date de déclaration du sinistre par l'assuré, que débutera le remboursement de l'assurance.

Cependant, ce remboursement ne s'effectue pas toujours en une fois, et peut se répartir en plusieurs années. Cet effet de cadencement oblige donc la compagnie d'assurance à se constituer des réserves pour pouvoir couvrir la sinistralité future. Or, comme le montant qui sera finalement payé pour le sinistre est inconnu au départ, la somme à mettre en réserve est également inconnue et il faut alors l'estimer.

D'après l'Article R 331 – 6 du code de l'Assurance : « *La provision pour sinistres à payer dite PSAP, est la valeur estimative des dépenses en principal et en frais, tant internes qu'externes, nécessaires au règlement de tous les sinistres survenus et non payés, y compris les capitaux constitutifs des rentes non encore mises à la charge de l'entreprise.* »

La PSAP représente ainsi la dette existante de l'assureur envers les assurés lorsque les sinistres ont été déclarés mais non encore réglés ou non encore rapportés.

L'enjeu pour l'assureur est de couvrir la sinistralité future et, de ce fait, se doit d'estimer au plus juste et de la façon la plus précise possible ses engagements.

En fait, cette provision imposée aux assurances par la réglementation se veut de protéger les assurés, souscripteurs et bénéficiaires des contrats : un trop faible provisionnement par exemple, pourrait conduire les compagnies à la faillite, et a contrario, une trop forte exigence en terme de provisionnement pourrait se révéler pénalisant pour l'assureur.

La compagnie se charge d'étudier sa provision au moment de l'inventaire dite date de fin d'exercice, du 31/12 de l'année n .

Pour un sinistre déclaré à l'année n , la provision à mettre en place chaque année ainsi que les montants des paiements réglés évoluent dans le temps. Ainsi, en fonction des nouvelles informations mais aussi des paiements effectués et à effectuer par la suite, la provision est sans cesse réévaluée.

Souvent, la provision s'avère décroissante au fil du temps.

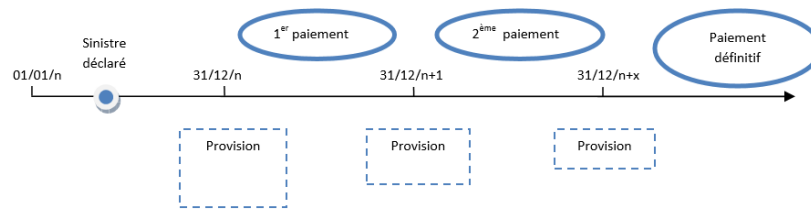


FIGURE 1: Paiements et provisions d'un sinistre

Le calcul de cette provision est crucial. Depuis que l'assurance existe, de nombreuses méthodes se sont développées autour de ce sujet.

Les méthodes traditionnellement nécessaires au calcul des montants de provision sont des méthodes dites statistiques qui reposent essentiellement sur les données historiques de la sinistralité.

On distingue alors les méthodes dites déterministes des méthodes dites stochastiques.

Ces méthodes sont d'autant plus performantes que :

- Le passé est régulier
- Le présent et le futur sont structurellement peu différents du passé
- La branche considérée est peu volatile
- Les données sont nombreuses
- Les données sont fiables

Etant donné que la qualité de la provision repose d'abord sur la qualité et la stabilité des données, certains risques à caractère « exceptionnel » et à données insuffisantes, tels que les risques atomique et spatial par exemple, ne pourront se prêter à l'utilisation de ces méthodes.

De plus, la survenance de sinistres catastrophiques de par leur « rareté » perturbent gravement le processus de provisionnement. Ils bénéficieront d'un traitement spécifique, et seront mis de côté lors de notre étude.

Les données d'étude sont présentées de la façon suivante : les sinistres sont regroupés et mis sous forme de triangles.

Ils sont rattachés à des années d'origine que sont leur année de survenance.

Les montants remboursés qui peuvent être exprimés annuellement ou en montants cumulés sont ensuite rangés par année de développement.

Les années de développement correspondent aux années au cours desquelles les sinistres sont remboursés par rapport à leur année d'origine.

Le délai de développement traduit bien le fait que les sinistres ne sont pas tous réglés l'année d'origine du sinistre. Il correspond à l'étalement des différents paiements d'un sinistre.

Si l'on suppose que chaque sinistre aura été réglé en totalité au bout de $n+1$ années de développement, cette durée représente alors une borne maximale, ce qui explique la forme du triangle. On distingue le triangle des montants incrémentaux et le triangle des montants cumulés.

On introduit :

- i en ligne, l'année de survenance des sinistres
- j en colonne, l'année de développement
- $i+j$ (en diagonale), l'année calendaire de paiement (pour les incréments)

Le triangle des montants incrémentaux $X_{i,j}$, affiche les montants réglés pour les sinistres survenus à l'année i , payés à l'année $i+j$, et supposés non encore observés pour $i+j > n$.

		Délai de règlement					
		0	1	...	j	...	J
Année de survenance	0	$X_{0,0}$	$X_{0,1}$...	$X_{0,j}$...	$X_{0,J}$
	1	$X_{1,0}$					
					
	i	$X_{i,1}$					
					
	I	$X_{I,0}$					

Paievements de l'année calendaire I

FIGURE 2: Représentation d'un triangle de développement avec montants incrémentaux

Le triangle des montants cumulés affiche quant à lui les montants cumulés $C_{i,j}$ pour les sinistres survenus en année i et réglés entre l'année 0 et l'année $i+j$, et supposés non encore observés pour $i+j > n$.

On notera que $C_{i,j} = X_{i,0} + X_{i,1} + \dots + X_{i,n}$, pour l'année de survenance i , avec $X_{i,j}$ les incréments de paiements, pour l'année de développement j .

		Délai de règlement					
		0	1	...	j	...	J
Année de survenance	0	$C_{0,0}$	$C_{0,1}$...	$C_{0,j}$...	$C_{0,J}$
	1	$C_{1,0}$					
					
	i	$C_{i,1}$					
					
	I	$C_{I,0}$					

Montant cumulé des paiements pour l'année calendaire I

FIGURE 3: Représentation d'un triangle de développement avec montants cumulés

Les quantités analysées sont donc des paiements de sinistres survenus.

Ces derniers sont rapportés à des périodes annuelles, généralement du 01/01/n au 31/12/n, date d'inventaire à laquelle les provisions sont à estimer pour l'année à venir.

Les montants incrémentaux et cumulés mêlent dans les triangles les sinistres clos, les sinistres déclarés mais non clôturés ainsi que les sinistres dits tardifs c'est à dire, non encore déclarés.

Tout l'objectif des méthodes de provisionnement sera, à partir des données de paiements du triangle supérieur et éventuellement d'informations exogènes, de compléter la partie inférieure du triangles des paiements pour obtenir, au 31/12/n :

- La charge sinistre ultime $S_i = C_{i,n}$ de chaque année d'origine i où $i = 0, \dots, n$ avec $S_0 = C_{0,n}$ connue .

- La provision $R_i = C_{i,n} - C_{i,n-i}$ à constituer pour chaque année d'origine i avec $R_0 = 0$
- La provision globale

$$R = \sum_{i=1}^n R_i.$$

Ce mémoire portera alors sur le calcul de la provision pour sinistres à payer et plus précisément sur la provision à mettre en place face aux sinistres déjà survenus.

Différentes méthodes déterministe et stochastique sont envisagées dans ce mémoire : le modèle de Mack, la méthode Munich Chain Ladder et la méthode PIC proposée par Merz Wüthrich (2010).

L'étude de ces méthodes successives, mettra ainsi au jour une évolution et avancée des méthodes de calcul de la PSAP.

Nous présenterons donc dans le cadre de ce projet, un plan graduel, à l'image de ces méthodes à complexité croissante.

Chapitre 1

La méthode classique Chain Ladder et le modèle de Mack

1.1 Présentation de la méthode déterministe Chain Ladder

La méthode déterministe probablement la plus connue et la plus utilisée permettant d'estimer les réserves de sinistres est la méthode Chain Ladder.

Son succès vient de sa mise en pratique simple.

Elle peut s'appliquer sur des triangles de toute nature comme ceux de paiements cumulés , ou ceux de charges de sinistres.

Pour chaque année d'origine ou survenance i , et chaque année de développement j , on introduit les facteurs de développement individuels $f_{i,j} = \frac{C_{i,j+1}}{C_{i,j}}$.

Ils correspondent au passage d'une année de développement à l'autre .

Cette méthode s'appuie sur l'hypothèse suivante :

Pour j allant de 0 à $n-1$, les facteurs de développement $f_{i,j}$ sont indépendants de l'année de survenance i , autrement dit, pour j allant de 0 à $n-1$, on a $\frac{C_{0,j+1}}{C_{0,j}} = \frac{C_{1,j+1}}{C_{1,j}} = .. = \frac{C_{i,j+1}}{C_{i,j}} = \frac{C_{n-j-1,j+1}}{C_{n-j-1,j}}$.

On considère ainsi des coefficients de passage communs pour les années d'origine i et ceci pour chacune des années de développement j :

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

Dans la pratique, ces inégalités précédentes ne sont qu'approximativement vérifiées et on en choisit une approximation $\widehat{f_j}$ pour j allant de 0 à $n-1$.

En conjecturant que les cadences de règlement sont constantes pour toutes les années d'origine, et pour prévoir les règlements futurs, on applique les facteurs de développement aux montants de sinistres déjà connus. Grâce à ces facteurs, on en déduira :

- Les évaluations des charges ultimes $S_i = C_{i,n} = C_{i,n-i} \prod_{h=n-i}^{n-1} f_h$
Ils nous fourniront alors un rectangle complet de liquidation par $C_{i,j} = C_{i,n-i} \prod_{h=n-i}^{j-1} f_h$ où $i + j > n$.
- Les provisions par exercice de survenance $R_i = S_i$.
- $C_{i,n-i}$: cela correspond à la différence entre le montant total de sinistres et les montants réglés à la date d'inventaire.

- La provision globale

$$R = \sum_{i=1}^n R_i,$$

aussi appelée « réserve globale de Chain-Ladder ». Elle correspond à la différence entre le montant total à régler et le montant déjà remboursé.

Cependant, malgré sa simplicité d'utilisation, il est important d'effectuer certaines vérifications pour valider cette méthode.

1.1.1 Validation du modèle

Afin de pouvoir appliquer la méthode Chain Ladder, il nous faut vérifier les hypothèses sous-jacentes de cette méthode. Nous retenons à cet effet deux tests graphiques :

- L'examen de l'alignement des couples
- L'examen du triangle de développement

L'examen de l'alignement des couples : Si pour j fixé, il existe un paramètre f_j tel que $C_{i,j+1} = f_{i,j} C_{i,j}$ pour i allant de 0 à $n-j-1$, les $(n-j)$ couples $(C_{i,j}; C_{i,j+1})$ doivent être « sensiblement » alignés sur une droite passant par l'origine.
La représentation graphique correspondante prend la dénomination de CC plot.

L'examen du triangle de développement : Le triangle de développement dit d-triangle, est formé des facteurs de développement individuels : $f_{i,j} = \frac{C_{i,j+1}}{C_{i,j}}$ pour $i + j \leq n - 1$.
L'objectif est de visualiser les variations des facteurs individuels par colonne du d-triangle $(f_{i,j})_i$. L'hypothèse est alors vérifiée si pour $j = 0, \dots, n-2$, les $f_{i,j}$ sont sensiblement constants : cela traduira alors une faible volatilité ou variance.

Si tel n'est pas le cas, son examen permet d'effectuer des choix alternatifs de facteurs .

Cet examen permet en particulier de mettre en évidence des colonnes faisant apparaître des valeurs extrêmes ou des structures non aléatoires .

1.1.2 Les limites de la méthode Chain Ladder

Bien que connue et facile à mettre en place, la méthode Chain Ladder fait apparaître des insuffisances. Il est utile de noter que pour conserver une approche de cadences de règlement constantes sur les années d'origine, il est nécessaire d'introduire de la flexibilité dans le choix des facteurs de développement, autant de flexibilité que n'en permet la méthode Chain Ladder standard.

De manière générale, les résultats peuvent être considérés comme inexacts dès lors que l'on observe dans les données :

- Des irrégularités (absence de stabilité) des cadences de règlement au sein par exemple d'exercices plus récents
- Un nombre de contrats trop faible
- Une fréquence de sinistres trop faible
- Une dispersion élevée des montants de sinistres
- Une hausse ou baisse de l'inflation qui génèrerait respectivement un sous ou sur provisionnement

Par ailleurs, le caractère déterministe de la méthode Chain Ladder réduit considérablement son champ d'activités.

L'une des raisons pour laquelle on a recours aux méthodes stochastiques est de répondre au besoin de mesurer l'incertitude présente dans les triangles, ainsi que dans les résultats issus des méthodes déterministes.

Le principe des méthodes stochastiques repose sur une modélisation paramétrée du triangle de liquidation, les paramètres du modèle étant estimés à l'aide des données du triangle supérieur.

On présuppose donc que les données du rectangle de liquidation sont des variables aléatoires réelles observées au 31/12/n.

Basée sur les résultats obtenus par les méthodes déterministes, l'approche stochastique permet :

- D'explicitier les hypothèses utilisées dans le modèle
- De valider partiellement ces hypothèses par une analyse des résidus ou erreurs
- De détecter des irrégularités potentielles dans le triangle
- D'élargir le champ des paramètres de provisionnement pertinents, les paramètres de provisionnement étant des paramètres non directement accessibles, caractéristiques de la distribution "prédictive" mise sur notre variable des provisions R , tels que son espérance, sa variance, son écart type, son risque de prédiction (MSEP), ou encore sa Value at risk, sa probabilité d'insuffisance etc...
- D'évaluer la variabilité des estimations de ces paramètres de provisionnement, prévue par le modèle
- De construire des intervalles de confiance pour ces paramètres de provisionnement
- De simuler, à l'aide de méthodes de Monte-Carlo, la sinistralité d'exercices futurs

1.2 Le modèle de Mack et l'incertitude dans la prédiction

On peut toutefois noter que l'approche stochastique est soumise au risque d'erreur de spécification, consistant à utiliser un modèle «inexact» produisant des résultats erronés.

Cependant, compte tenu de l'importance pratique de la méthode Chain Ladder, l'intérêt était d'obtenir une méthode stochastique reproduisant les évaluations de celle-ci, bien qu'un modèle reproduisant exactement les résultats de Chain Ladder ne serait pas forcément une garantie d'optimalité dans le modèle stochastique.

Le modèle récursif de Mack se présente ainsi comme la version stochastique de la méthode Chain Ladder : il a non seulement le privilège de fournir le même « best estimate » des provisions que la méthode Chain Ladder, mais a de plus l'avantage d'estimer la volatilité de l'estimateur des provisions techniques, c'est-à-dire, l'erreur de prédiction des provisions : le modèle de Mack nous fournit ainsi davantage d'informations sur les provisions à constituer.

De plus, il a pour intérêts :

- D'intégrer des indicateurs de risque de prédiction déduits d'une hypothèse de volatilité au sein du triangle.
- D'être non paramétrique au sens où aucune hypothèse de distribution n'est faite sur les composantes du triangle
- D'être conditionnel au sens où les espérances sont déterminées sachant les réalisations du triangle supérieur

Enfin, ce modèle s'applique à des montants cumulés : nombres, paiements, charges sinistres etc...
Ce modèle non paramétrique repose alors sur trois hypothèses :

H1 : L'indépendance des années d'origine
 $\{C_{i,1}, \dots, C_{i,n}\}$ et $\{C_{k,1}, \dots, C_{k,n}\}$ sont indépendants pour $k \neq i$.

H2 : Il existe f_j telle que $E[C_{i,j+1}|C_{i,1}, \dots, C_{i,j}] = f_j C_{i,j}$ pour $0 \leq i \leq n-1, 0 \leq j \leq n-2$

H3 : Il existe σ_j tel que $\text{Var}(C_{i,j+1}|C_{i,1}, \dots, C_{i,j}) = \sigma_j^2 C_{i,j}$ pour $0 \leq i \leq n-1, 0 \leq j \leq n-2$.

Soit $D = \{C_{i,j} | i+j \leq n\}$ l'ensemble des variables observables.

Sous les hypothèses ci-dessus, pour $i \geq 1$, on a $E(C_{i,n}|D) = C_{i,n-i+1}f_{n-i+1} \dots f_{n-1}$.
En effet, par l'hypothèse d'indépendance **H1**, on a :

$$E(C_{i,n}|D) = E(C_{i,n}|C_{i,1}, \dots, C_{i,n-i+1}),$$

puis, par double conditionnement, on obtient :

$$E(C_{i,n}|D) = E[E(C_{i,n}|C_{i,1}, \dots, C_{i,n-1}) | C_{i,1}, \dots, C_{i,n-i+1}],$$

avec :

$$E(C_{i,n}|C_{i,1}, \dots, C_{i,n-1}) = f_{n-1}C_{i,n-1},$$

On en déduit :

$$E(C_{i,n}|D) = f_{n-1}E(C_{i,n-1}|C_{i,1}, \dots, C_{i,n-i+1}),$$

par récurrence, on a :

$$\begin{aligned} E(C_{i,n}|D) &= f_{n-1} \times \dots \times f_{n-i+1} E(C_{i,n-i+1}|C_{i,1}, \dots, C_{i,n-i+1}) \\ &= f_{n-1} \times \dots \times f_{n-i+1} C_{i,n-i+1}. \end{aligned}$$

Pour $j = 0, \dots, n-2$, nous obtenons comme facteurs de la méthode Chain Ladder :

$$\hat{f}_j = \frac{\sum_{i=0}^{n-j-1} C_{i,j+1}}{\sum_{i=0}^{n-j-1} C_{i,j}},$$

Bien que dépendants du même triangle de données, nous allons démontrer que ce sont non seulement des estimateurs sans biais des paramètres f_j , mais qu'ils sont de plus décorrélés.

Théorème 1.2.1 *Les estimateurs sont sans biais.*

Démonstration 1.2.2 : Pour $j = 0, \dots, n-1$, $D_j = \{C_{ih} : h \leq j, i+j \leq n\}$ la partie du triangle supérieure limitée au délai de règlement j . On a :

$$E(C_{i,j+1}|D_j) = E(C_{i,j+1}|C_{i1}, \dots, C_{ij}) = f_j C_{ij},$$

puis :

$$E(\hat{f}_j|D_j) = E\left[\frac{\sum_{i=1}^{n-j} C_{i,j+1}}{\sum_{i=1}^{n-1} C_{ij}} | D_j\right] = \frac{\sum_{i=1}^{n-1} E(C_{ij+1}|D_j)}{\sum_{i=1}^{n-1} C_{ij}} = f_j,$$

et, par conditionnement :

$$E(\hat{f}_j) = E[E(\hat{f}_j|D_j)] = f_j,$$

On a bien \hat{f}_j , estimateur sans biais de f_j pour $j < k$.

Théorème 1.2.3 *Les estimateurs sont décorrélés.*

Démonstration 1.2.4 : $E(\hat{f}_j, \hat{f}_k) = E[E(\hat{f}_j, \hat{f}_k | D_k)] = E[\hat{f}_j E(\hat{f}_k | D_k)] = f_j f_k$.

Par l'égalité précédente, $E(\hat{f}_k | D_k) = f_k$.

On en déduit : $\text{Cov}(\hat{f}_j, \hat{f}_k) = E(\hat{f}_j \hat{f}_k) - E(\hat{f}_j) E(\hat{f}_k) = 0$.

D'où la non corrélation des \hat{f}_j .

L'absence de biais se transmet également aux valeurs "calculées" à l'aide des facteurs standards de la méthode Chain Ladder :

$$\hat{C}_{i,n} = C_{i,n-i+1} \hat{f}_{n-i+1} \times \dots \times \hat{f}_{n-1} \text{ pour } E(C_{i,n} | D) = C_{i,n-i+1} f_{n-i+1} \times \dots \times f_{n-1},$$

Plus généralement :

$$\hat{C}_{i,j} = C_{i,n-i+1} \hat{f}_{n-i+1} \times \dots \times \hat{f}_{j-1} \text{ pour } E(C_{i,j} | D) = C_{i,n-i+1} f_{n-i+1} \times \dots \times f_{j-1},$$

Il en est de même de :

$$\hat{R}_i = \hat{C}_{i,n} - C_{i,n-i+1} \text{ pour } E(R_i | D) = E(C_{i,n} | D) - C_{i,n-i+1},$$

$$\hat{R} = \sum_{i=0}^n \hat{R}_i \text{ pour } E(\hat{R} | D).$$

Sous les 3 hypothèses ci-dessus, on peut montrer de manière similaire que les (σ_j^2) proposés par Mack, et estimés par :

$$\begin{cases} \hat{\sigma}_j^2 = \frac{1}{n-j-1} \sum_{i=0}^{n-j-1} C_{i,j} \left(\frac{c_{ij+1}}{c_{ij}} - \hat{f}_j \right)^2, 0 \leq j \leq n-2 \\ \hat{\sigma}_{n-1}^2 = \min(\hat{\sigma}_{n-2}^4 / \hat{\sigma}_{n-3}^2, \min(\hat{\sigma}_{n-3}^2, \hat{\sigma}_{n-2}^2)) \end{cases}$$

sont aussi des estimateurs sans biais.

On peut en déduire que ces estimateurs sont optimaux au sens où ils sont sans biais.

Il est à noter que ces estimateurs coïncident avec ceux de la méthode Chain Ladder standard.

On se propose maintenant de mesurer l'incertitude de la prédiction de C_{in} par \widehat{C}_{in} . Pour cela, on utilise l'erreur quadratique moyenne conditionnelle que l'on notera

$$MSEP(\widehat{C}_{in}) = E\left((C_{in} - \widehat{C}_{in})^2 | D\right),$$

espérance en laquelle seule C_{in} est aléatoire.

Le MSEP (Mean Square Error of Prediction) va donc mesurer l'écart entre la variable et sa valeur estimée, conditionnellement aux informations déjà connues.

Remarquons que l'erreur de prédiction de la provision par année d'exercice $MSEP(\widehat{R}_i)$ est telle que : $MSEP(\widehat{R}_i) = MSEP(\widehat{C}_{in})$ car $R_i - \widehat{R}_i = C_{in} - \widehat{C}_{in}$.

Ainsi, sous les hypothèses **H1** à **H3**, l'erreur de prédiction par exercice, $MSEP(\widehat{R}_i)$, peut être estimé pour $i = 2, \dots, n$ par :

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

$MSEP(\hat{R})$ pourra être estimé par :

$$\widehat{MSEP}(\hat{R}) = \sum_{i=2}^n \left\{ \widehat{MSEP}(\hat{R}_i) + \hat{C}_{in} \left(\sum_{k=1+1}^n \hat{C}_{kn} \right) \sum_{j=n-1+1}^{n-1} \frac{2\hat{\sigma}_j^2}{\hat{f}_j^2 \sum_{n=1}^{n-1} C_{ni}} \right\}. \quad (1.1)$$

En outre, disposant des estimations pour la moyenne de la variable R_i , notée \hat{R}_i , nous avons dorénavant accès à son écart-type $\sqrt{\widehat{MSEP}(\hat{R}_i)}$.

Sous des hypothèses portant sur le choix d'une distribution "prédictive" conditionnelle pour R , T.Mack construit des intervalles de prédiction pour la provision c'est à dire des intervalles de confiance autour des \hat{R}_i . On peut cependant noter le caractère arbitraire d'un tel choix, à l'opposé du caractère non paramétrique de cette approche.

1.3 Mise en oeuvre de la méthode sur notre jeu de données

Application de la méthode Chain Ladder :

Dans le cadre de ce projet, plusieurs jeux de données nous ont été proposés. Chacun d'entre eux regroupe 6 triangles.

Mettons à present en oeuvre la méthode Chain Ladder sur notre premier jeu de données.

Les données que nous possédons sont les triangles `fretrilauto9605` et `fretri2auto9605` du package CAS datasets contenant des triangles payés et encourus d'assureurs français, en non-vie, entre 1996 et 2005. Plus précisément, deux garanties en automobile sont disponibles : la garantie dommages matériels et la garantie dommages corporels. On retrouve à chaque fois deux triangles par catégorie, celui des montants encourus et celui des montants payés. Il y a 3 catégories : body, damage et total. Nous nous intéresserons principalement à la catégorie totale. Les données commencent à partir de l'année d'accident 1996 et correspondent à la situation en 2006. On notera cependant que l'on ne dispose d'aucune information avant 1999, il manque donc 6 valeurs sur l'arbre.

Nous nous concentrerons dans un premier temps sur le premier jeu de données, `fretrilauto9605`.

Nous nous attarderons dans cette section sur le triangle total des montants payés, relatif à notre premier jeu de données :

	0	1	2	3	4	5	6	7	8	9
1996				186 908	190 860	194 217	198 879	199 726	200 403	201 022
1997			186 967	195 488	201 015	204 518	206 394	207 930	209 766	
1998		170 092	183 403	189 427	194 477	198 577	202 297	206 513		
1999	112 074	167 919	179 126	186 632	191 689	195 838	201 103			
2000	111 317	155 731	166 886	173 593	182 013	186 220				
2001	112 656	160 725	176 065	183 339	188 226					
2002	107 608	147 991	159 641	165 503						
2003	100 534	137 948	146 703							
2004	101 198	135 817								
2005	100 197									

TABLE 1.1: Données Francaises `fretrilauto9605 total paid` : montant cumulé $C_{i,j}$

A partir du tableau des charges payées 1.1 on calcule les facteurs de developpement. On en rappelle la formule :

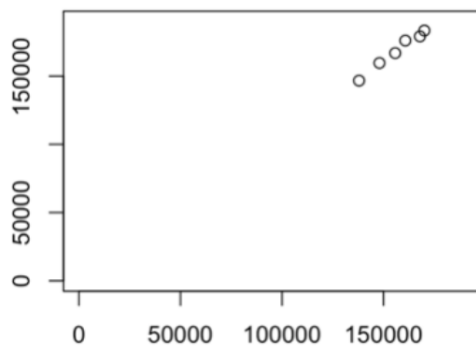
$$\forall j : 1 \leq j \leq n \quad \text{and} \quad f_j = \frac{\sum_{i=0}^{n-j-1} C_{i,j+1}}{\sum_{i=0}^{n-j-1} C_{i,j}}, \quad (1.2)$$

j	0	1	2	3	4	5	6	7	8	9
f_j	1.404	1.076	1.040	1.029	1.020	1.019	1.010	1.006	1.003	1.000

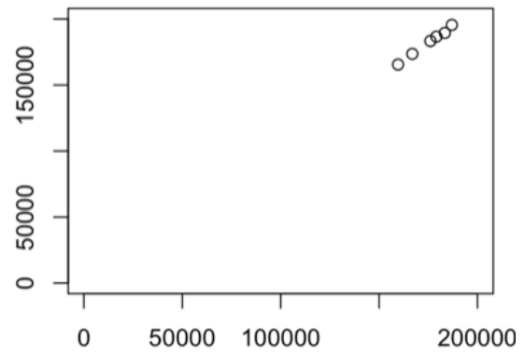
TABLE 1.2: Facteurs de développement

Vérification des hypothèses :

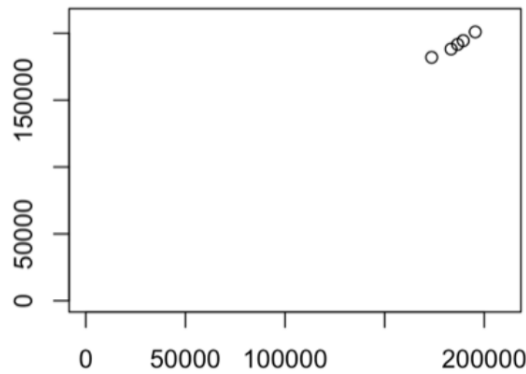
Examen de l'alignement des couples : Voici les C-C plots :



CC plot pour j=1



CC plot pour j=2



CC plot pour j=3

FIGURE 1.1: CC Plots

On observe que les couples $(C_{i,j}, C_{i,j+1})_i$ sont sensiblement alignés sur une droite passant par l'origine : l'hypothèse n'est pas remise en cause.

Examen du triangle de développement : Voici le tableau des facteurs de développement. On l'accompagne du tableau des statistiques de base pour chacune des colonnes j : la moyenne, l'écart-type et le

coefficient de variation (C_v). Ce dernier est un indicateur normalisé de volatilité : il est le quotient de l'écart type et de la moyenne.

$f_{i,j}$	0	1	2	3	4	5	6	7	8
1996				1.021	1.018	1.024	1.004	1.003	1.003
1997			1.046	1.028	1.017	1.009	1.007	1.009	
1998		1.078	1.033	1.027	1.021	1.087	1.021		
1999	1.498	1.067	1.042	1.027	1.022	1.027			
2000	1.399	1.072	1.040	1.048	1.023				
2001	1.427	1.095	1.041	1.027					
2002	1.375	1.078	1.037						
2003	1.372	1.063							
2004	1.342								

TABLE 1.3: Tableau des facteurs de developpement

	0	1	2	3	4	5	6	7	8
\hat{f}_j	1.404	1.076	1.040	1.029	1.020	1.019	1.010	1.006	1.003
$\hat{\sigma}_j$	0.05	0.01	0.00	0.01	0.00	0.01	0.01	0.00	0.00
C_v	0.04	0.01	0.00	0.01	0.00	0.01	0.01	0.00	0.00

TABLE 1.4: Tableau de statistiques par colonne

Nous avons représenté les variations des facteurs individuels par colonne du triangle de développement pour $j = 0$ à 4.

On constate une faible volatilité en $j=0$, mais qui semble s'atténuer pour les autres valeurs de j .

Les variations du triangle de développement, presque négligeables, ne laissent donc pas apparaître de structure aléatoire. L'hypothèse n'est pas remise en cause.

De plus, le tableau de statistiques confirme une volatilité globalement faible et décroissante avec j . Cela renforce bien l'hypothèse selon laquelle les coefficients sont sensiblement constants par colonne.

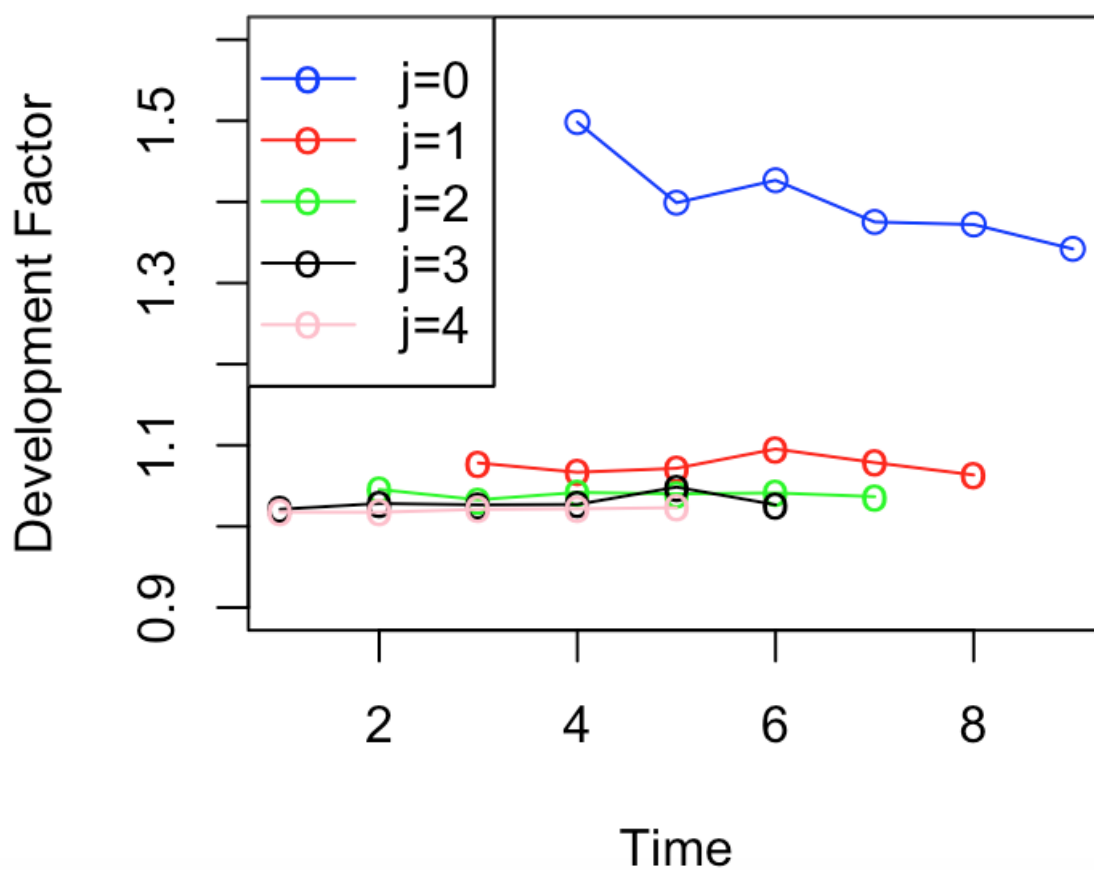


FIGURE 1.2: Facteurs de développement en fonction du temps

En appliquant la méthode Chain Ladder, on complète le triangle des montants payés :

	0	1	2	3	4	5	6	7	8	9
1996	118 989	167 063	179 750	186 908	190 860	194 217	198 879	199 726	200 403	201 022
1997	123 767	173 770	186 967	195 488	201 015	204 518	206 394	207 930	209 766	210 414
1998	121 147	170 092	183 403	189 427	194 477	198 577	202 297	206 513	207 786	208 427
1999	112 074	167 919	179 126	186 632	191 689	195 838	201 103	203 288	204 541	205 172
2000	111 317	155 731	166 886	173 593	182 013	186 220	189 865	191 927	193 110	193 706
2001	112 656	160 725	176 065	183 339	188 226	192 013	195 771	197 898	199 117	199 732
2002	107 608	147 991	159 641	165 503	170 384	173 812	177 214	179 138	180 242	180 799
2003	100 534	137 948	146 703	152 545	157 043	160 203	163 339	165 113	166 130	166 643
2004	101 198	135 817	146 132	151 951	156 432	159 579	162 702	164 470	165 483	165 994
2005	100 197	140 677	151 361	157 388	162 030	165 290	168 525	170 355	171 405	171 935

TABLE 1.5: Tableau des montants cumulés ultimes

A partir de ce tableau on peut alors calculer pour chacune des années : la charge ultime des sinistres avec le développement actuel ainsi que les provisions :

	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Ult	201 022	210 414	208 427	205 172	193 706	199 732	180 799	166 643	165 994	171 935
Dev.Totale	100.00 %	99.69 %	99.08 %	98.02 %	96.14 %	94.24 %	91.54 %	88.03 %	81.82 %	58.28 %
Prov	0	648	1 915	4 069	7 486	11 506	15 296	19 940	30 177	71 738

TABLE 1.6: Tableau récapitulatif de la méthode Chain Ladder

En additionnant les provisions relatives à chacune des années, on obtient une provision globale de 162 775 ainsi qu'un développement total de 91.4 %

Pour mieux visualiser l'évolution par année :

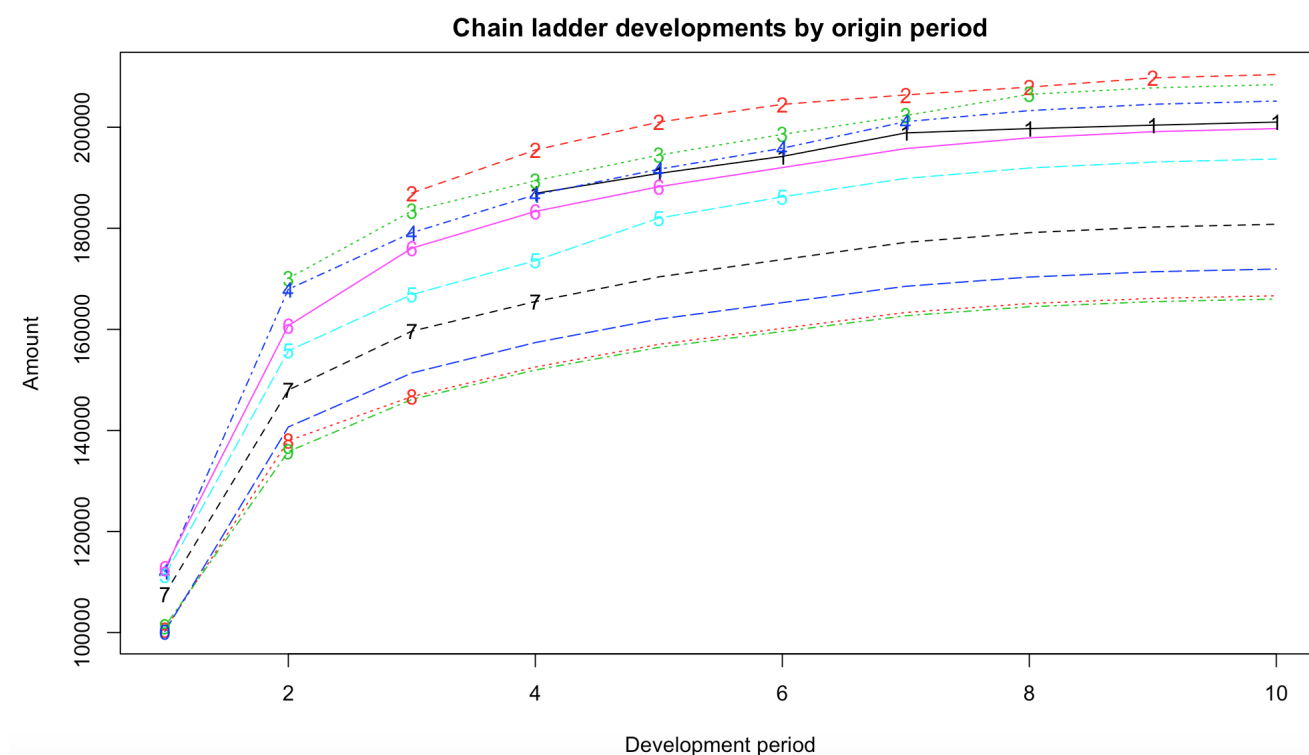


FIGURE 1.3: Évolution graphique des montants payés

Calcul de l'incertitude :

Appliquons désormais les formules de Mack à notre jeu de données pour déterminer l'incertitude de notre provision :

	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Prov	0	648	1 915	4 069	7 486	11 506	15 296	19 940	30 177	71 738
Incertitude	0	501	1 092	2 329	2 821	2 926	3 318	3 272	3 839	8 480
Taux d'incertitude		0.77	0.57	0.57	0.38	0.25	0.22	0.16	0.13	0.12

TABLE 1.7: Tableau relatif à l'incertitude

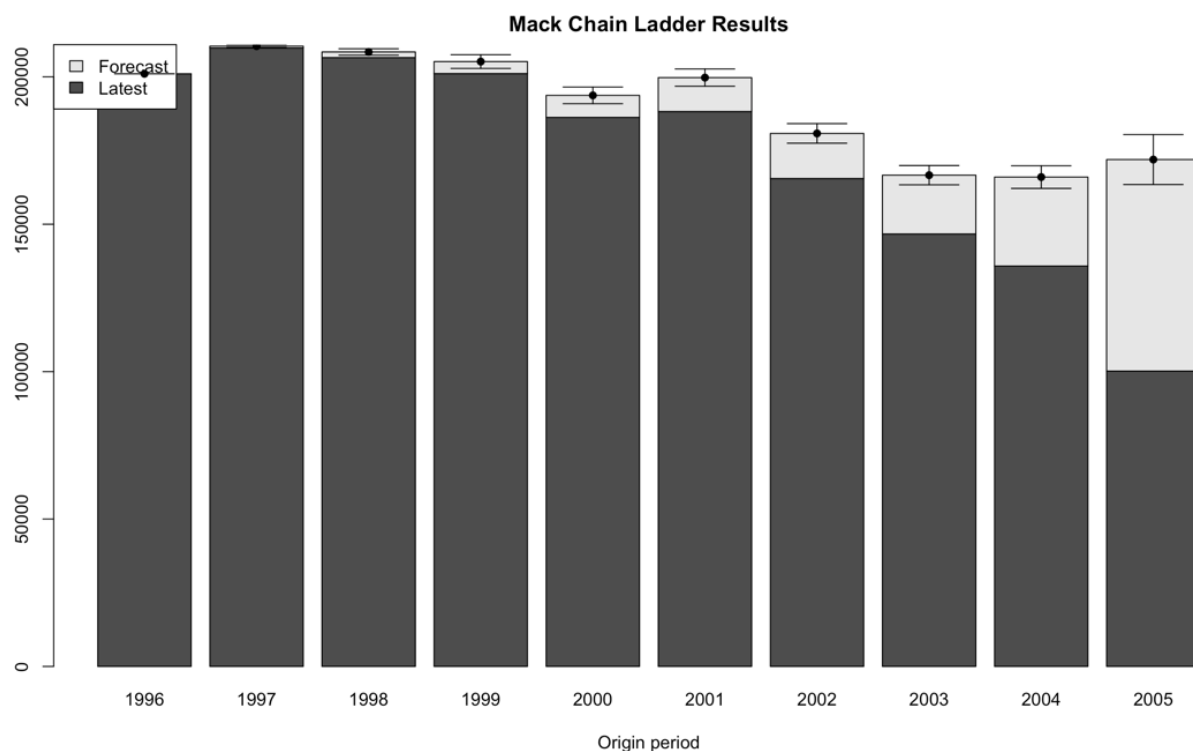


FIGURE 1.4: Montants payés et incertitudes

Ici, l'incertitude totale est de 15 605, soit une incertitude liée à la provision globale proche de 9 %. La figure 1.4, ci-contre, nous permet de mieux visualiser l'incertitude et de mieux comprendre les informations issues du tableau 1.7 : la partie foncée représente les montants déjà payés tandis que la partie claire représente les provisions estimées par la méthode Chain Ladder.

De plus, au sommet de chaque colonne, se trouve une boîte à moustache qui mesure l'incertitude par année d'exercice. Plus la boîte à moustache empiète sur la partie claire des provisions estimées, plus le taux d'incertitude augmente. Par exemple, à l'année 2000 où la boîte à moustache recouvre presque toute la partie claire, le taux d'incertitude est 3 fois plus grand que celui à l'année 2005 où la boîte à moustache recouvre faiblement la partie claire.

Chapitre 2

Munich Chain Ladder

2.1 Avant-propos

La méthode Chain Ladder a certes l'avantage d'être simple à comprendre et à mettre en oeuvre, mais fait néanmoins ressortir un problème crucial en provisionnement : la non convergence usuelle des provisions de la méthode Chain Ladder issues du triangle des paiements d'une part, et de celui des charges d'autre part.

		Délai						
Exercice	i	0	1	2	3	4	5	Provisions
1988	0						4456	0
1989	1					4730	4752	22
1990	2				5420	5430	5456	36
1991	3			6020	6046	6057	6086	66
1992	4		6794	6872	6902	6914	6947	153
1993	5	5217	7204	7287	7318	7332	7367	2150
							Total	2427

FIGURE 2.1: Paiements cumulés brut de recours et provisions

		Délai						
Exercice	i	0	1	2	3	4	5	Provisions
1988	0						4456	0
1989	1					4750	4750	20
1990	2				5470	5456	5456	36
1991	3			6131	6101	6085	6085	65
1992	4		7087	6920	6886	6869	6869	75
1993	5	7353	7173	7004	6970	6952	6952	1735
							Total	1931

FIGURE 2.2: Charges cumulées prédites et provisions

Nous appliquons dans un premier temps la méthode Chain Ladder au triangle total des montants

payés (2.1) pour obtenir le rectangle de liquidation des paiements cumulés, et en déduisons les provisions associées.

Puis, dans un second temps, nous appliquons une nouvelle fois la méthode Chain Ladder au triangle des sinistres survenus (2.2) pour obtenir le tableau des charges ultimes estimées.

En faisant la différence entre les charges ultimes estimées et la diagonale du triangle des paiements cumulés, nous en retirons alors les provisions correspondantes.

On obtient donc :

$R(\text{paiements}) = 2427$ euros et $R(\text{survenus}) = 1931$ euros.

Il en résulte que l'écart relatif entre ces deux provisions est de 20,4 %. Cet écart n'est pas négligeable.

On note, pour i, j allant de 1 à n , P_{ij} les paiements cumulés de l'année d'origine i après j années de développement, et I_{ij} la charge sinistre de l'année d'origine i vue à la fin de la $j^{\text{ème}}$ année. Ces quantités sont connues pour $i + j \leq n + 1$ et seront projetées au-delà.

Pour $j = 1, \dots, n - 1$, les facteurs de développement standards de la méthode Chain Ladder pour les triangles (P) et (I) sont

$$f_j^{(P)} = \frac{\sum_{i=1}^{n-j} P_{i,j+1}}{\sum_{i=1}^{n-j} P_{i,j}} \quad \text{et} \quad f_j^{(I)} = \frac{\sum_{i=1}^{n-j} I_{i,j+1}}{\sum_{i=1}^{n-j} I_{i,j}}$$

et pour $j > n - i + 1$, $P_{ij+1} = f_j^{(P)} P_{ij}$, $I_{ij+1} = f_j^{(I)} I_{ij}$ sont les projections de la méthode Chain Ladder.

Dans la méthode Munich Chain Ladder, un rôle fondamental est joué par les ratios $\frac{P}{I}$ (qui représente la part de la charge estimée effectivement payée), et $\frac{I}{P}$ déduits des paiements (P_{ij}) et de charges (I_{ij}).

On s'intéresse donc aux ratios $\frac{P}{I}$:

$$\left(\frac{P}{I}\right)_{ij} = \frac{P_{ij}}{I_{ij}} \quad i, j = 1, \dots, n.$$

Les triangles supérieurs des montants payés et des charges sinistres donnent le triangle supérieur des ratios $\frac{P}{I}$. On observe sur le graphique une convergence attendue vers 1 en cours de développement.

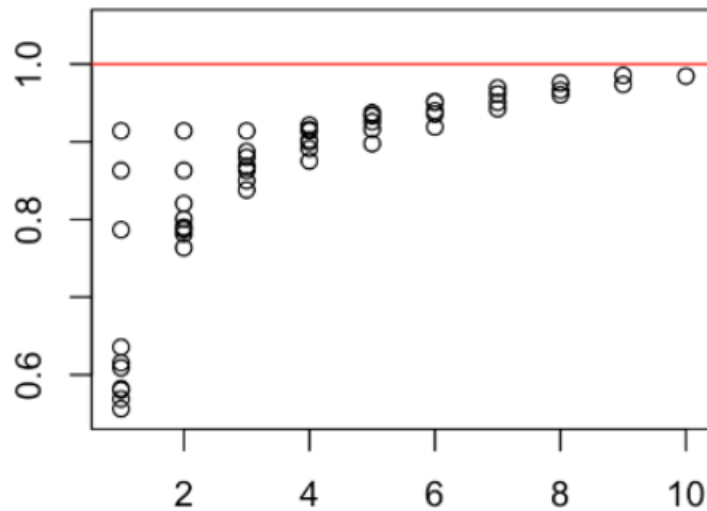


FIGURE 2.3: Ratios $\frac{P}{I}$ par délai (triangle)

On applique respectivement la méthode Chain Ladder aux rectangles P (2.1) et I (2.2) pour obtenir les rectangles de liquidation associés.

La représentation graphique des ratios $\frac{P}{I}$ montre que la convergence vers 1 est perdue après projection

de la sinistralité. C'est là encore l'une des limites de l'efficacité de la méthode Chain Ladder standard et l'un des problèmes que la méthode Munich Chain Ladder tente de résoudre.

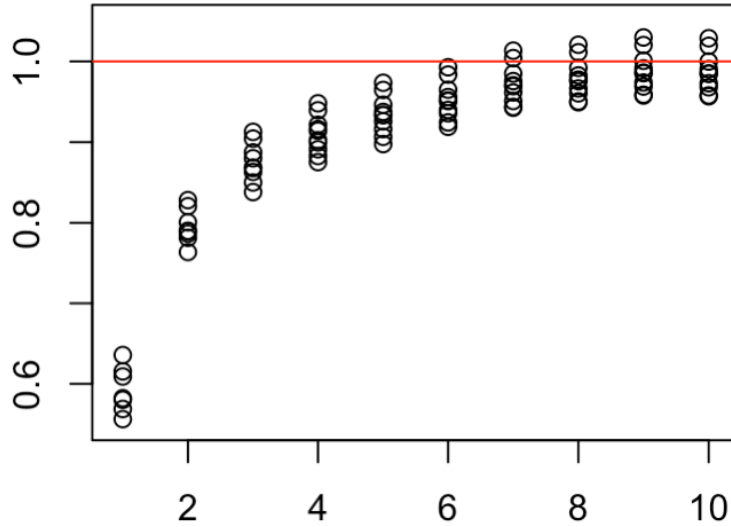


FIGURE 2.4: Ratios $\frac{P}{I}$ par délai après application de Mack (rectangle)

La méthode Munich Chain Ladder a été présentée sous une forme initiale par G.Quarg au XXX^e colloque ASTIN de Berlin (2003), et a été publiée sous sa forme actuelle par G.Quarg et T.Mack en 2004.

2.2 Explication de la méthode

La méthode Munich Chain Ladder présente une différence majeure par rapport à la méthode de Mack.

Premièrement, dans cette dernière méthode, les facteurs de développement sont indépendants des années d'origine. De plus, ceux-ci ne dépendent que du triangle auquel la méthode est appliquée. La Munich Chain Ladder, quant à elle, prend en compte la corrélation entre les paiements et les charges par année d'origine pour les mêmes développements. Ainsi, les facteurs de développement varient d'une année d'origine à l'autre et dépendent des données des triangles des montants payés et des charges. Cependant, la corrélation entre paiements et charges n'est pas directe.

On lui substitue l'approche suivante : pour les exercices passés totalement déroulés, un ratio $\frac{P}{I}$ relativement faible, était suivi par des facteurs de développement relativement élevés pour les P, ou faibles pour les I, ou les deux.

Inversement, un ratio $\frac{P}{I}$ relativement élevé, était suivi par des facteurs de développement relativement faibles pour les P, ou élevés pour les I, ou les deux.

L'idée de base de la Munich Chain Ladder consiste donc à corrélérer les facteurs de développement standards de la méthode Chain Ladder avec les ratios $\frac{P}{I}$ du même délai c'est à dire pour $j = 1, \dots, n-1$, les facteurs individuels $f_{ij}^{(P)} = \frac{P_{i,j+1}}{P_{i,j}}$ aux ratios $(\frac{P}{I})_{ij}, i = 1, \dots, n-j+1$ et les facteurs individuels $f_{ij}^{(I)} = \frac{I_{i,j+1}}{I_{i,j}}$ aux ratios $(\frac{I}{P})_{ij}, i = 1, \dots, n-j+1$.

Pour tenter de résoudre le problème de la non convergence des $\frac{P}{I}$, dans la Chain Ladder, la méthode Munich Chain Ladder propose donc de substituer à l'utilisation dans le triangle des paiements des

facteurs de développement (Chain Ladder) communs à toutes les années d'origine, celle de facteurs dépendants de ces mêmes facteurs et de la situation des $\frac{I}{P}$ par rapport à la moyenne.

2.2.1 Triangle des montants payés

On s'intéresse dans un premier temps au triangle des montants payés (2.1). Nous voulons lier, pour $j = 1, \dots, n-1$, les facteurs de développement individuels $f_{ij}^{(P)} = \frac{P_{i,j+1}}{P_{ij}}$ aux ratios $(\frac{P}{I})_{ij}$ pour i allant de 1 à $n-j+1$.

On utilisera les notations suivantes :

$$P_i(j) = (P_{ik})_{1 \leq k \leq j}, \quad I_i(j) = (I_{ik})_{1 \leq k \leq j}, \quad B_i(j) = \{P_i(j), I_i(j)\}$$

et

$$Q_i = (Q_{ij})_{1 \leq j \leq n}, \quad \text{pour } Q_{ij} = \frac{P_{ij}}{I_{ij}}$$

Les hypothèses :

La méthode Munich Chain Ladder commence par reprendre des hypothèses similaires à celles du modèle récursif de Mack mais appliquées à notre triangle d'intérêt, ici le triangle des paiements cumulés. Les montants P_{ij} sont considérés comme des variables aléatoires. La première hypothèse se voit aussi prolongée afin de prendre en compte le fait que l'on travaille sur deux triangles : celui des paiements cumulés et des charges sinistres.

Les hypothèses sont :

H1^(P) : Les variables aléatoires $(P_{1j}, I_{1j})_{1 \leq j \leq n}, \dots, (P_{nj}, I_{nj})_{1 \leq j \leq n}$ sont indépendantes.

H2^(P) : Pour $j = 1, \dots, n-1$, il existe un paramètre $f_j^{(P)}$ réel tel que conditionnellement :

$$E\left(\frac{P_{i,j+1}}{P_{ij}} | P_i(j)\right) = f_j^{(P)},$$

H3^(P) : Pour $j = 1, \dots, n-1$, il existe un paramètre $\sigma_j^{(P)} > 0$ tel que conditionnellement :

$$\text{Var}\left(\frac{P_{i,j+1}}{P_{ij}} | P_i(j)\right) = \frac{(\sigma_j^{(P)})^2}{P_{ij}},$$

Enfin nous avons une hypothèse additionnelle qui porte sur la corrélation entre les facteurs et les ratios $\frac{I}{P}$. Afin d'établir ce lien, il faut d'abord standardiser ces variables aléatoires. On définit le résidu conditionnel standardisé pour $\frac{P_{i,j+1}}{P_{ij}}$ par

$$\text{Res}\left[\frac{P_{i,j+1}}{P_{ij}} | P_i(j)\right] = \frac{\frac{P_{i,j+1}}{P_{ij}} - E\left(\frac{P_{i,j+1}}{P_{ij}} | P_i(j)\right)}{\sigma\left(\frac{P_{i,j+1}}{P_{ij}} | P_i(j)\right)} = \frac{\frac{P_{i,j+1}}{P_{ij}} - f_j^{(P)}}{\sigma_j^{(P)}} \sqrt{P_{ij}},$$

en utilisant l'hypothèse **H2^(P)** pour la formule de l'espérance conditionnelle, et l'hypothèse **H3^(P)** pour l'écart type conditionnel; et celui des ratios $\frac{I}{P}$ par

$$\text{Res}\left[Q_{ij}^{-1} | P_i(j)\right] = \frac{Q_{ij}^{-1} - E\left[Q_{ij}^{-1} | P_i(j)\right]}{\sigma\left[Q_{ij}^{-1} | P_i(j)\right]},$$

Ainsi, la méthode Munich Chain Ladder fait l'hypothèse d'une corrélation *linéaire* entre les facteurs standardisés $\text{Res} \left[\frac{P_{i,j+1}}{P_{ij}} | P_i(j) \right]$ et les ratios $\frac{I}{P}$ standardisés.

H4^(P) : Pour $i = 1, \dots, n$ et $j = 1, \dots, n-1$, il existe un paramètre réel $\lambda^{(P)}$ tel que conditionnellement :

$$E \left\{ \text{Res} \left[\frac{P_{i,j+1}}{P_{ij}} | P_i(j) \right] | B_i(j) \right\} = \lambda^{(P)} \text{Res} \left[Q_{ij}^{-1} | P_i(j) \right],$$

Le paramètre $\lambda^{(P)}$, pente de la droite de régression des facteurs contre les ratios $\frac{I}{P}$ est le coefficient de corrélation linéaire (conditionnelle) entre les facteurs $\left(\frac{P_{i,j+1}}{P_{ij}} \right)$ et les ratios $\left(\frac{I_{i,j}}{P_{ij}} \right)$. Notons aussi que $\lambda^{(P)}$ est supposé indépendant de l'année de développement j .

Estimation des paramètres :

De manière analogue au modèle récursif de Mack, pour $j = 1, \dots, n-1$, les facteurs de la méthode Chain Ladder $f_j^{(P)}$, ainsi que les paramètres de volatilité $\sigma_j^{(P)}$, sont estimés à l'aide du triangle des paiements cumulés.

Estimation de $E \left[Q_{ij}^{-1} | P_i(j) \right]$:

On pose

$$E \left[Q_{ij}^{-1} | P_i(j) \right] \text{ pour estimer } \widehat{q}_j^{-1} = \frac{\sum_{k=1}^{n-j+1} I_{kj}}{\sum_{k=1}^{n-j+1} P_{kj}},$$

Estimation de $\sigma \left[Q_{ij}^{-1} | P_i(j) \right]$:

On pose

$$\widehat{\rho_j^{(P)}}^2 = \frac{1}{n-j} \sum_{k=1}^{n-j+1} P_{kj} \left(Q_{kj}^{-1} - \widehat{q}_j^{-1} \right)^2,$$

puis

$$\frac{\widehat{\rho_j^{(P)}}}{\sqrt{P_{ij}}} \text{ pour estimer } \sigma \left[Q_{ij}^{-1} | P_i(j) \right],$$

Estimations des résidus standardisés :

Les facteurs standardisés et les ratios $\frac{I}{P}$ standardisés sont estimés, pour $j = 1, \dots, n-1$, par

$$\text{Res} \left(\widehat{\frac{P_{i,j+1}}{P_{ij}}} | P_i(j) \right) = \frac{\frac{P_{i,j+1}}{P_{ij}} - \widehat{f}_j^{(P)}}{\widehat{\sigma_j^{(P)}}} \sqrt{P_{ij}},$$

et

$$\text{Res} \left(\widehat{Q_{ij}^{-1}} | P_i(j) \right) = \frac{Q_{ij}^{-1} - \widehat{q}_j^{-1}}{\widehat{\rho_j^{(P)}}} \sqrt{P_{ij}}.$$

Estimation du coefficient de corrélation $\lambda^{(P)}$:

$\lambda^{(P)}$ est estimé par l'estimateur des moindres carrés de la régression linéaire de $\text{Res} \left[\frac{P_{i,j+1}}{\widehat{P}_{ij}} | P_i(j) \right]$ et $\text{Res} \left[Q_{ij}^{-1} | P_i(j) \right]$:

$$\widehat{\lambda^{(P)}} = \frac{\sum_{i,j} \text{Res} \left(\frac{P_{i,j+1}}{\widehat{P}_{ij}} | P_i(j) \right) \cdot \text{Res} \left(Q_{ij}^{-1} | P_i(j) \right)}{\sum_{i,j} [\text{Res} \left(Q_{ij}^{-1} | P_i(j) \right)]^2},$$

Enfin, pour une année d'origine i allant de 1 à n , on déduit de l'hypothèse **H4**^(P) la formule récursive pour compléter le rectangle des paiements :

$$\widehat{P_{i,j+1}} = \widehat{P_{i,j}} \left[\widehat{f_j^{(P)}} + \widehat{\lambda^{(P)}} \frac{\widehat{\sigma_j^{(P)}}}{\widehat{\rho_j^{(P)}}} \left(\widehat{Q_{ij}^{-1}} - \widehat{q_j^{-1}} \right) \right] \quad j \geq n - i + 1,$$

avec $\widehat{Q_{ij}^{-1}} = \frac{\widehat{I_{ij}}}{\widehat{P_{ij}}}$. On remarque que l'on retrouve le facteur de la méthode Chain Ladder mais avec une quantité additionnelle qui dépend des triangles des montants payés et du triangle des charges.

2.2.2 Triangle des charges

On refait ensuite le travail de la section précédente en prenant comme triangle d'intérêt le triangle des charges (2.2).

De manière symétrique, on se place sous les hypothèses :

H1^(I) : Les variables aléatoires $(P_{1j}, I_{1j})_{1 \leq j \leq n}, \dots, (P_{nj}, I_{nj})_{1 \leq j \leq n}$ sont indépendantes.

H2^(I) : Pour $j = 1, \dots, n - 1$, il existe un paramètre $f_j^{(I)}$ réel tel que conditionnellement :

$$E \left(\frac{I_{i,j+1}}{I_{ij}} | I_i(j) \right) = f_j^{(I)},$$

H3^(I) : Pour $j = 1, \dots, n - 1$, il existe un paramètre $\sigma_j^{(I)} > 0$ tel que conditionnellement :

$$\text{Var} \left(\frac{I_{i,j+1}}{I_{ij}} | I_i(j) \right) = \frac{\left(\sigma_j^{(I)} \right)^2}{I_{ij}},$$

Cette fois-ci, le lien est effectué entre les facteurs de développement du triangle des charges sinistres et les ratios $\frac{P}{I}$.

H4^(I) : Pour $i = 1, \dots, n$ et $j = 1, \dots, n - 1$, il existe un paramètre réel $\lambda^{(I)}$ tel que conditionnellement :

$$E \left\{ \text{Res} \left[\frac{I_{i,j+1}}{I_{ij}} | I_i(j) \right] | B_i(j) \right\} = \lambda^{(I)} \text{Res} [Q_{ij} | I_i(j)],$$

Estimation des paramètres :

On estime ensuite les paramètres pour notre triangle d'intérêt :

Estimation de $E [Q_{ij}^{-1} | I_i(j)]$:

$$E [Q_{ij} | I_i(j)] \text{ pour estimer } \hat{q}_j = \frac{\sum_{k=1}^{n-j+1} P_{kj}}{\sum_{k=1}^{n-j+1} I_{kj}}$$

Estimation de $\sigma [Q_{ij}^{-1} | I_i(j)]$:

On pose

$$\widehat{\rho_j^{(I)}}^2 = \frac{1}{n-j} \sum_{k=1}^{n-j+1} I_{kj} (Q_{kj} - \hat{q}_j)^2,$$

puis

$$\frac{\widehat{\rho_j^{(I)}}}{\sqrt{I_{ij}}} \text{ pour estimer } \sigma [Q_{ij} | I_i(j)],$$

Estimation des résidus standardisés

Les facteurs standardisés et les ratios $\frac{P}{I}$ standardisés sont estimés, pour $j = 1, \dots, n-1$, par

$$\text{Res} \left(\frac{\widehat{I_{i,j+1}}}{I_{ij}} | I_i(j) \right) = \frac{\frac{I_{i,j+1}}{I_{ij}} - \widehat{f_j^{(I)}}}{\widehat{\sigma_j^{(I)}}} \sqrt{I_{ij}},$$

et

$$\text{Res} (\widehat{Q_{ij}} | I_i(j)) = \frac{Q_{ij} - \hat{q}_j}{\widehat{\rho_j^{(I)}}} \sqrt{I_{ij}},$$

Estimation du coefficient de corrélation $\lambda^{(I)}$:

$\lambda^{(I)}$ est estimé par :

$$\widehat{\lambda^{(I)}} = \frac{\sum_{i,j} \text{Res} \left(\frac{\widehat{I_{i,j+1}}}{I_{ij}} | I_i(j) \right) \cdot \text{Res} (\widehat{Q_{ij}} | I_i(j))}{\sum_{i,j} [\text{Res} (\widehat{Q_{ij}} | I_i(j))]^2},$$

Enfin, pour une année d'origine i allant de 1 à n , on a la formule récursive pour compléter le rectangle des charges sinistres :

$$\widehat{I_{i,j+1}} = \widehat{I_{i,j}} \left[\widehat{f_j^{(I)}} + \widehat{\lambda^{(I)}} \frac{\widehat{\sigma_j^{(I)}}}{\widehat{\rho_j^{(I)}}} (\widehat{Q_{ij}} - \hat{q}_j) \right], \quad j \geq n-i+1$$

avec $\widehat{Q_{ij}} = \frac{\widehat{P_{ij}}}{\widehat{I_{ij}}}$.

2.3 Mise en oeuvre de la méthode sur notre jeu de données

Mettons à présent en oeuvre la méthode Munich Chain Ladder sur le premier jeu de données qui nous a été proposé. Ici, nous devons utiliser deux triangles, le triangle des paiements et le triangle des charges, que nous noterons respectivement, pour plus de simplicité, P et I.

Nous avons pris la liberté de compléter les 6 valeurs manquantes de chacun des tableaux en utilisant selon un processus inversé, les données "futures" que nous avons du triangle, ainsi que les coefficients de développement de la méthode Chain Ladder. Ces valeurs ne sont cependant pas indispensables pour appliquer la méthode Munich Chain Ladder.

	0	1	2	3	4	5	6	7	8	9
1996	118989	167063	179750	186908	190860	194217	198879	199726	200403	201022
1997	123767	173770	186967	195488	201015	204518	206394	207930	209766	
1998	121147	170092	183403	189427	194477	198577	202297	206513		
1999	112074	167919	179126	186632	191689	195838	201103			
2000	111317	155731	166886	173593	182013	186220				
2001	112656	160725	176065	183339	188226					
2002	107608	147991	159641	165503						
2003	100534	137948	146703							
2004	101198	135817								
2005	100197									

TABLE 2.1: Tableau des montants payés (P)

	0	1	2	3	4	5	6	7	8	9
1996	130172	182764	196644	204474	204475	206653	205188	204714	203354	204173
1997	143403	201340	216630	217271	217123	214912	214647	215138	215388	
1998	153992	216206	211180	212559	212269	212221	212738	215036		
1999	201502	219920	213742	213227	213554	213097	213358			
2000	191830	199280	192297	192367	194919	195855				
2001	198037	203493	200067	200141	200804					
2002	176676	184838	179912	179606						
2003	172686	174617	172605							
2004	159220	165520								
2005	162815									

TABLE 2.2: Tableau des charges (I)

On commence par calculer pour chacun des triangles, 3.1 et 3.2 les facteurs de développement standards de la méthode Chain Ladder, ainsi qu'un estimateur de σ_j :

j	0	1	2	3	4	5	6	7	8
$\hat{f}_j^{(P)}$	1.40	1.08	1.04	1.03	1.02	1.02	1.01	1.01	1.00
$\hat{\sigma}_j^{(P)}$	14.30	3.81	1.72	4.00	1.11	3.49	3.95	1.74	0.76

TABLE 2.3: Tableau de statistiques pour P

j	0	1	2	3	4	5	6	7	8
$\hat{f}_j^{(I)}$	1.14	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.00
$\hat{\sigma}_j^{(I)}$	71.36	20.57	6.64	2.42	3.57	1.93	3.05	2.53	2.10

TABLE 2.4: Tableau de statistiques pour I

On peut donc désormais standardiser (moyenne nulle et variance unitaire) le triangle des résidus des facteurs individuels des paiements et des charges :

	0	1	2	3	4	5	6	7	8
0	-0.00	0.00	-0.00	-0.90	-1.00	0.56	-0.75	-0.71	0.00
1	-0.00	0.00	1.44	-0.13	-1.09	-1.35	-0.39	0.70	
2	-0.00	0.25	-1.73	-0.31	0.38	-0.11	1.14		
3	2.21	-0.99	0.51	-0.26	0.60	0.93			
4	-0.12	-0.45	0.09	1.98	1.15				
5	0.53	2.05	0.36	-0.30					
6	-0.66	0.28	-0.72						
7	-0.71	-1.22							
8	-1.38								

TABLE 2.5: Triangle des résidus des facteurs individuels

On notera que le triangle des trois premières valeurs est égal à zéro. Ce resultat s'explique par le fait que l'on ait remplacé les valeurs manquantes en utilisant les facteurs de developpement de la méthode Chain Ladder. On ne prendra cependant pas en compte ces valeurs pour le calcul de $\lambda^{(P)}$. Calculons désormais le ratio des charges par rapport aux paiements, et déterminons grâce à la méthode Chain Ladder son facteur de developpement. On notera alors ce triangle de ratio $Q^{-1} = \frac{I}{P}$ et son facteur de developpement \hat{q}_j^{-1} .

j	0	1	2	3	4	5	6	7	8	9
\hat{q}_j^{-1}	1.52	1.23	1.15	1.11	1.08	1.06	1.05	1.03	1.02	1.02

TABLE 2.6: Facteurs de développement de $\frac{I}{P}$

Nous pouvons ensuite calculer $\hat{\rho}_j^{(P)}$ et $\hat{\rho}_j^{(I)}$ grâce à ces deux quantités ainsi qu'à P et I.

j	0	1	2	3	4	5	6	7	8
$\hat{\rho}_j^{(P)}$	89.32	27.73	12.72	8.50	7.78	6.73	5.75	3.69	3.87
$\hat{\rho}_j^{(I)}$	51.34	20.88	10.38	7.25	6.85	6.11	5.37	3.51	3.75

TABLE 2.7: Valeurs par colonne des paramètres $\hat{\rho}_j^{(P)}$ et $\hat{\rho}_j^{(I)}$

Ainsi, nous pouvons en déduire le triangle des résidus des $Q_{i,j}^{-1}$. Calculons désormais le paramètre $\lambda^{(P)}$ pente des droites de régression des facteurs contre les ratios $\frac{I}{P}$. C'est le coefficient de corrélation linéaire entre le résidu de P et le résidu de Q :

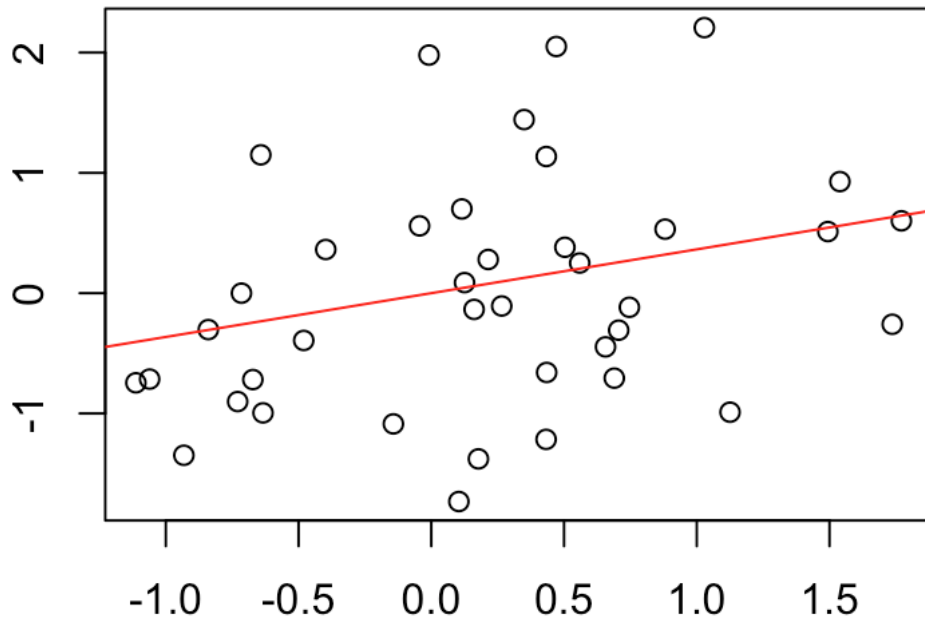


FIGURE 2.5: Représentation des facteurs en fonction des résidus

De par la méthode des moindres carrés on trouve $\lambda^{(P)}=0.364$

De la même manière, on détermine $\lambda^{(I)}$ qui vaut 0.474.

Nous disposons désormais de tous les outils nécessaires pour compléter le triangle des montants payés, ainsi que le triangle des charges. Pour compléter les deux triangles, on utilisera de manière alternée les deux formules récursives suivantes :

$$\widehat{P_{i,j+1}} = \widehat{P_{i,j}} \left[\widehat{f_j^{(P)}} + \widehat{\lambda^{(P)}} \frac{\widehat{\sigma_j^{(P)}}}{\widehat{\rho_j^{(P)}}} \left(\widehat{Q_{i,j}^{-1}} - \widehat{q_j^{-1}} \right) \right], \quad \text{et} \quad \widehat{I_{i,j+1}} = \widehat{I_{i,j}} \left[\widehat{f_j^{(I)}} + \widehat{\lambda^{(I)}} \frac{\widehat{\sigma_j^{(I)}}}{\widehat{\rho_j^{(I)}}} \left(\widehat{Q_{i,j}} - \widehat{q_j} \right) \right], \quad (2.1)$$

Nous pouvons donc compléter P et I et ainsi observer que le ratio $\frac{P}{I}$ converge bien vers 1 :

	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Ult P	201 021	210 507	208 864	206 802	192 391	197 691	177 638	168 671	160 387	164 394
Ult I	204 173	216 270	215 380	214 560	196 560	201 542	180 650	174 098	162 555	166 771
Ratio $\frac{P}{I}$	0.98	0.97	0.97	0.96	0.98	0.98	0.98	0.97	0.99	0.99

TABLE 2.8: Tableau de developpement de P et I

On s'intéresse désormais aux facteurs de developpement de P et on les compare aux facteurs standards de la méthode Chain Ladder :

$f_{i,j}$	0	1	2	3	4	5	6	7	8
1									1.003
2								1.009	1.004
3							1.021	1.008	1.004
4						1.027	1.015	1.009	1.004
5					1.023	1.017	1.007	1.005	1.003
6				1.027	1.019	1.016	1.006	1.005	1.003
7			1.037	1.025	1.019	1.015	1.006	1.005	1.003
8		1.063	1.041	1.034	1.021	1.022	1.012	1.008	1.004
9	1.342	1.068	1.036	1.024	1.018	1.013	1.004	1.003	1.002
f_{CL}	1.404	1.076	1.040	1.029	1.020	1.019	1.010	1.006	1.003

TABLE 2.9: Tableau comparatif des facteurs de développement

On calcule les provisions du rectangle des paiements P ainsi que celui du rectangle des charges I. On rappelle que les provisions de P sont obtenues en retranchant à la colonne ultime de P, l'anti-diagonale de P, tandis que celles de I seront obtenues en retranchant à la colonne ultime de I, l'anti-diagonale de P.

	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Prov P	0	741	2 351	5 699	6 171	9 465	12 135	21 968	24 570	64 197
Prov I	3 151	6 504	8 867	13 457	10 340	13 316	15 147	27 395	26 738	66 574

TABLE 2.10: Tableau de provisions des triangles P et I

On répertorie ci-dessous pour chaque année d'exercice ,la charge ultime des sinistres, le developpement , ainsi que les provisions de P :

	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Ult P	201 021	210 507	208 864	206 802	192 391	197 691	177 638	168 671	160 387	164 394
Dev Total	100.00 %	99.65 %	98.87 %	97,24 %	96.79 %	95.21 %	93.17 %	86.98 %	84.68 %	60.95 %
Prov	0	741	2 351	5 699	6 171	9 465	12 135	21 968	24 570	64 197

TABLE 2.11: Tableau récapitulatif de la méthode Munich Chain Ladder

On notera que ce sont les montants de provisions de P issus de la méthode Munich Chain Ladder que l'on confrontera à ceux de la méthode Chain Ladder pour comparer les deux méthodes .

	Munich Chain Ladder	Chain Ladder
Dev Total	92,1 %	91,4 %
Prov	147 299	162 795

TABLE 2.12: Tableau comparatif des méthodes Munich Chain Ladder et Chain Ladder

On remarque entre les deux méthodes une différence de provisions de 15496, soit une différence de près de 10 % . L'écart du taux de développement total est quant à lui de 0,7%.

Notre jeu de données ne révèle donc pas une différence significative entre les deux méthodes.

Cependant , il est à noter que la provision réalisée par la méthode Munich Chain Ladder appartient à l'intervalle de prédiction de T.Mack [147 170,178 380], en particulier, dans les estimations basses de l'incertitude de Mack.

De plus, on constate que les projections des charges ultimes issues du triangle des paiements et de

celui des charges sinistres témoignent d'une convergence très améliorée du ratio $\frac{P}{I}$ vers 1.

D'ailleurs, les tests de cette méthode effectués dans une importante mutuelle d'assurance sur des triangles moyens et longs de branches variées ont pu montrer une convergence systématique des charges ultimes à condition que les triangles traités soient sensiblement complets .

Si cette méthode était confirmée sur de nombreux autres exemples à liquidation moyenne ou longue puis par des résultats théoriques, cette méthode serait ainsi d'une grande utilité pratique .

Chapitre 3

Paid-incurred chain claims reserving method

La Munich Chain Ladder est l'une des méthodes de provisionnement dont le succès repose sur sa capacité à combiner de manière relativement efficace les informations sur les charges sinistres et celles sur les paiements cumulés. Son objectif était, comme nous l'avons détaillé dans la partie précédente, de réduire l'écart entre les prévisions de la méthode Chain Ladder basées sur les données de P (tableau des paiements cumulés) et de I (tableau de charges sinistres) que nous disposions. Cependant, cette méthode s'avérait complexe au sens où la précision des estimateurs qu'elle introduisait était difficile à quantifier dans le cadre du modèle stochastique. Dans cette partie, nous allons désormais nous intéresser à la "Paid-incurred chain claims reserving method", dite méthode PIC. Cette méthode fut développée par Michael Merz et Mario V. Wuthrich en 2010. Tout comme la méthode Munich Chain Ladder, elle utilise les triangles des montants payés et encourus, mais a pour particularité de les introduire dans le cadre d'un modèle lognormale avec estimation bayésienne.

3.1 Présentation du modèle lognormale

Le modèle lognormale est abordable sous deux approches : l'une avec estimation fréquentiste et l'autre avec estimation bayésienne. Pour mieux comprendre leur principe et les caractéristiques qui leur sont propres, nous allons les présenter en les appliquant sur le triangle P des paiements.

3.1.1 Modèle lognormale avec estimation fréquentiste

Une variable aléatoire X dit suivre une loi lognormale de paramètres μ et σ^2 si la variable $Y = \log(X)$ suit une loi normale d'espérance μ et de variance σ^2 . Cette loi est définie sur R_+ et sa fonction de densité est :

$$f_Y(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right) = \frac{1}{x} f_X(\log(x), \mu, \sigma), \quad (3.1)$$

On a $E(X) = \exp(\mu + \frac{\sigma^2}{2})$ et $\text{Var}(X) = (\exp(\sigma^2) - 1) \exp(2\mu + \sigma^2)$.

On considère ici un modèle paramétrique sur les facteurs de développement. Déjà évoqués dans la méthode Chain Ladder, on reprendra les mêmes notations.

On rappelle que

$$f_{i,j+1} = \frac{P_{i,j+1}}{P_{i,j}}$$

On définit $\eta_{i,j} = \log(f_{i,j}) = \log(\frac{P_{i,j}}{P_{i,j-1}})$ avec $P_{i,-1} = 1$.

On suppose que pour tout i, j , $\eta_{i,j}$ sont des variables aléatoires indépendantes et pour tout $j = 0, \dots, J$

$$\forall i, \eta_{i,j} | P_{i,j-1} \sim \mathcal{N}(g_j, s_j^2).$$

Autrement dit $P_{i,j} | P_{i,j-1}$ suit une loi lognormale

$$T_{i,j} = \frac{P_{i,j}}{P_{i,j-1}} | P_{i,j-1} \sim \mathcal{LN}(g_j, s_j^2).$$

On en conclut donc, par indépendance des $\eta_{i,j}$, et par les moments de la loi normale :

$$E[P_{i,j} | P_{i,j-1}] = E[P_{i,j-1} \frac{P_{i,j}}{P_{i,j-1}} | P_{i,j-1}] = P_{i,j-1} E[T_{i,j}] = P_{i,j-1} \exp(g_j + \frac{s_j^2}{2}),$$

De la même façon on obtient :

$$Var[P_{i,j} | P_{i,j-1}] = P_{i,j-1} \exp(g_j + \frac{s_j^2}{2}) \exp(s_j^2 - 1).$$

On retrouve une structure avec des facteurs de développement similaires à ceux de Mack avec

$$f_{j-1} = \exp(g_j + \frac{s_j^2}{2}) \text{ et } \sigma_j^2 = P_{i,j-1} \exp(2g_j + \frac{s_j^2}{2}) \exp(s_j^2 - 1).$$

On cherche désormais les estimateurs du maximum de vraisemblance :

$$\hat{g}_j = \frac{1}{J-j+1} \sum_{i=0}^{J-j} \log(f_{i,j-1}), \quad \hat{s}_j^2 = \frac{1}{J-j} \sum_{i=0}^{J-j} (\log(f_{i,j-1}) - \hat{g}_j)^2.$$

Comme l'échantillon $\eta_{i,j}$ est gaussien, on connaît aussi la loi exacte des estimateurs

$$\hat{g}_j \sim \mathcal{N}(g_j, \frac{s_j^2}{J-j+1}), \quad \frac{J-j}{s_j^2} \hat{\sigma}_j^2 \sim \chi_{J-j}^2, \quad \hat{g}_j \perp \hat{\sigma}_j^2.$$

L'estimateur sans biais du modèle lognormale, si s_j^2 est connu, est

$$\hat{P}_{i,J}^{LN} = P_{i,J-i} \exp(\sum_{j=J-i+1}^J \hat{g}_j + \sum_{j=J-i+1}^J \frac{s_j^2}{2} (1 - \frac{1}{J-j+1})).$$

On peut donc compléter le triangle P grâce à la formule ci-dessus.

3.1.2 Modèle lognormale avec estimation bayésienne

On considère ici le modèle où l'on définit pour tout i, j , $\eta_{i,j}$ comme variables aléatoires indépendantes où $j = 0, \dots, J$

$$\forall i, \eta_{i,j} | P_{i,j-1}, G_j = g_j \sim \mathcal{N}(g_j, s_j^2)$$

Autrement dit, la densité de $\frac{P_{i,j}}{P_{i,j-1}}$ est donnée par :

$$f_{\frac{P_{i,j}}{P_{i,j-1}}}(y) = \frac{1}{y s_j \sqrt{2\pi}} \exp(-\frac{(\log(y) - g_j)^2}{2s_j^2}), y > 0.$$

Pour tout $j = 0, \dots, J$, la distribution a priori est :

$$G_j \sim \mathcal{N}(\phi_j, \sigma_j^2).$$

On note $D_I = (P_{i,j})_{i+j \leq I}$. La loi normale étant conjuguée avec elle-même, la loi a posteriori $G_j | D_I$ restera donc de loi normale.

3.1.3 Le cadre fréquentiste

On définit tout d'abord :

- $\mathcal{B}_j^P = (P_{i,k} : 0 \leq i \leq J, 0 \leq k \leq j)$,
- $\mathcal{B}_j^I = (I_{i,k} : 0 \leq i \leq J, 0 \leq k \leq j)$,
- $\mathcal{B}_j = \mathcal{B}_j^I \cup \mathcal{B}_j^P$,
- $\Theta = (\Phi_0, \dots, \Phi_J, \Psi_0, \dots, \Psi_J, \sigma_0, \dots, \sigma_J, \tau_0, \dots, \tau_J)$ un vecteur de paramètres où $\sigma_j > 0$ pour tous j .
Il est à noter que dans le cadre fréquentiste, Θ est un vecteur inconnu mais déterministe.
- Le vecteur aléatoire

$$(\epsilon_{0,0}, \dots, \epsilon_{J,J}, \zeta_{0,0}, \dots, \zeta_{J,J}),$$

suit une loi gaussienne multidimensionnelle, où les composantes sont indépendantes.

- On a $\epsilon_{i,j} \sim \mathcal{N}(\Phi_j, \sigma_j)$ pour $i=0, \dots, J$ et $j=0, \dots, J$.
- On a $\zeta_{i,j} \sim \mathcal{N}(\Psi_j, \tau_j)$ pour $i=0, \dots, J$ et $j=0, \dots, J$.
- Le montant payé $P_{i,j}$ est donné par récurrence par $P_{i,j} = P_{i,j-1} \exp(\epsilon_{i,j})$, et $P_{i,0} = \exp(\epsilon_{i,0})$,
- Le montant encouru $I_{i,j}$ est donné par récurrence par $I_{i,j} = I_{i,j-1} \exp(\zeta_{i,j})$, et $I_{i,0} = \exp(\zeta_{i,0})$.

On rappelle qu'une loi a priori gaussienne a été mise sur chacune des variables aléatoires $\zeta_{i,j}$ et $\epsilon_{i,j}$. Les paramètres de leur loi normale respectivement (Φ_j, σ_j) et (Ψ_j, τ_j) sont des composantes inconnues mais déterministes du vecteur Θ que l'on cherchera également à estimer .

On a pour tout $j \geq 0$:

$$\log \frac{P_{i,j}}{P_{i,j-1}} \Big|_{\{\mathcal{B}_{j-1}^P\} \sim \mathcal{N}(\Phi_j, \sigma_j^2)}, \quad (3.2)$$

Cela nous donne la formule suivante :

$$E[P_{i,j} | \mathcal{B}_{j-1}^P] = P_{i,j-1} \exp(\Phi_j + \frac{\sigma_j^2}{2}), \quad (3.3)$$

Enfin, grâce aux propriétés de l'espérance conditionnelle on obtient :

$$E[P_{i,J} | \mathcal{B}_{j-1}^P] = P_{i,j} \exp(\sum_{k=j+1}^J \Phi_k + \frac{\sigma_k^2}{2}), \quad (3.4)$$

Pour la suite, nous allons utiliser un lemme sur les propriétés des lois gaussiennes multidimensionnelles :

Lemme 3.1.1 : Supposons que (X_1, \dots, X_n) suit une loi Gaussienne multidimensionnelle, de moyenne (m_1, \dots, m_n) , et avec une matrice de covariance définie positive Σ .

Alors on a :

$$X_1 | X_2, \dots, X_n \sim \mathcal{N}(m_1 + \Sigma_{1,2} \Sigma_{2,2}^{-1} (X^{(2)} - m^{(2)}), \Sigma_{1,1} - \Sigma_{1,2} \Sigma_{2,2}^{-1} \Sigma_{2,1}), \quad (3.5)$$

où $X^{(2)} = (X_2, \dots, X_n)$ suit une loi Gaussienne multidimensionnelle, de moyenne (m_1, \dots, m_n) , et de matrice de covariance définie positive $\Sigma_{2,2}$.

$\Sigma_{1,1}$ est la variance de X_1

$\Sigma_{1,2} = \Sigma_{2,1}$ est la matrice de covariance entre X_1 et $X^{(2)}$.

On rappellera que pour le modèle PIC, on a $I_{i,J} = P_{i,J}$ où $I_{i,j}$ suit une loi lognormale . On peut donc appliquer le lemme précédent au vecteur $(\log I_{i,j+k}, \log I_{i,j}, \log I_{i,j-1}, \dots, \log I_{i,0})$ qui a une distribution gaussienne. On obtient que pour tout $0 \leq j \leq k+j \leq J$:

$$\log I_{i,j+k} | \mathcal{B}_j^I \sim \mathcal{N}(\mu_{j+k} + \frac{v_{j+k}^2}{v_j^2}(\log I_{i,j} - \mu_j), v_{j+k}^2(1 - \frac{v_{j+k}^2}{v_j^2})), \quad (3.6)$$

Avec $\mu_j = \sum_{m=0}^J \Phi_m - \sum_{n=j}^{J-1} \Psi_n$ et $v_j = \sum_{m=0}^J \sigma_m^2 - \sum_{n=j}^{J-1} \tau_n^2$.

Ainsi en utilisant la propriété de Markov nous pouvons prédire l'espérance de $I_{i,J}$, sachant \mathcal{B}_{j-1}^I :

$$\begin{aligned} E[I_{i,J} | \mathcal{B}_j^I] &= I_{i,j}^{1-\alpha_j} \exp \left\{ (1-\alpha_j) \sum_{l=j}^{J-1} \Psi_l + \alpha_j \left(\mu_j + \frac{v_j^2}{2} \right) \right\} \\ &= I_{i,j} \exp \left\{ \sum_{k=j}^{J-1} \Psi_k + \frac{\tau_k^2}{2} \right\} \\ &\quad \times \exp \left\{ \alpha_j \left(\mu_j - \log I_{i,j} - \sum_{k=j}^{J-1} \frac{\tau_k^2}{2} \right) \right\}, \end{aligned} \quad (3.7)$$

avec $\alpha_j = 1 - \frac{v_j^2}{v_j^2} = \frac{1}{v_j^2} \sum_{k=j}^{J-1} \tau_k^2$.

Attardons-nous sur les équations (3.4) et (3.7) qui permettent de déterminer respectivement les espérances de $P_{i,j}$ et $I_{i,j}$ sachant \mathcal{B}_{j-1}^I .

On observe que $P_{i,j}$ respecte la deuxième hypothèse de Mack et de la Munich Chain Ladder en ayant pris $\exp(\sum_{k=j+1}^J \Phi_k + \frac{\sigma_k^2}{2})$ comme paramètre réel .

Ce n'est cependant pas le cas pour l'espérance de $I_{i,j}$ sachant \mathcal{B}_{j-1}^I , pour laquelle on retrouve une partie similaire à l'espérance de $P_{i,j}$ sachant \mathcal{B}_{j-1}^I , c'est à dire le terme

$$\exp\left(\sum_{k=j}^{J-1} \Psi_k + \frac{\tau_k^2}{2}\right),$$

mais avec en plus un "facteur correcteur"

$$\exp(\alpha_j(\mu_j - \log I_{i,j} - \sum_{k=j}^{J-1} \tau_k^2)).$$

On observe que plus j est faible, et plus le "facteur correcteur" s'avère important, de même qu' α_j est décroissant avec j.

En utilisant les résultats précédents et toujours sous les hypothèses de départ du modèle PIC, on obtient pour $P_{i,J} = I_{i,J}$, sachant \mathcal{B}_j , $0 \leq j < J$:

$$\log P_{i,J} | \mathcal{B}_j \sim \mathcal{N}(\mu_J + (1-\beta_j)(\log P_{i,j} - \eta_j) + \beta_j(\log I_{i,j} - \mu_j), (1-\beta_j)(v_j^2 - w_j^2)), \quad (3.8)$$

avec

$$\beta_j = \frac{\sum_{m=j+1}^J \sigma_m^2}{\sum_{m=j+1}^J \sigma_m^2 + \sum_{n=j}^{J-1} \tau_n^2} = \frac{v_j^2 - w_j^2}{v_j^2 - w_j^2} \geq 0,$$

$$\mu_j = \sum_{m=0}^J \Phi_m - \sum_{n=j}^{J-1} \Psi_n, \eta_j = \sum_{m=0}^j \Phi_m \text{ et } \eta_J = \mu_J,$$

$$v_j = \sum_{m=0}^J \sigma_m^2 - \sum_{n=j}^{J-1} \tau_n^2, w_j = \sum_{m=0}^j \sigma_m^2 \text{ et } v_J = w_J.$$

L'équation (3.8) nous montre que pour le modèle lognormale PIC, nous pouvons calculer de manière analytique la loi explicite des paiements ultimes, sachant \mathcal{B}_j . On peut donc calculer avec (3.8) l'espérance conditionnelle de $P_{i,j}$, sachant \mathcal{B}_j :

$$\begin{aligned} E[P_{i,j}|\mathcal{B}_j] &= P_{i,j} \exp \left\{ \sum_{l=j+1}^J \Phi_l + \frac{\sigma_l^2}{2} \right\} \\ &\times \exp \left\{ \beta_j \left(\log \frac{I_{i,j}}{P_{i,j}} - (\mu_j - \eta_j) - \sum_{l=j+1}^J \frac{\sigma_l^2}{2} \right) \right\} \\ &= I_{i,j} \exp \left\{ \sum_{l=j}^{j-1} \Psi_l \right\} \\ &\times \exp \left\{ (1 - \beta_j) \left(\log \frac{P_{i,j}}{I_{i,j}} - (\eta_j - \mu_j) + \sum_{l=j+1}^J \frac{\sigma_l^2}{2} \right) \right\} \\ &= \exp \left\{ (1 - \beta_j) \left(\log P_{i,j} + \sum_{l=j+1}^J \Phi_l \right) + \beta_j \left(\log I_{i,j} + \sum_{l=j}^{J-1} \Psi_l \right) \right\} \\ &\times \exp \left\{ (1 - \beta_j) \left(\frac{v_j^2 - w_j^2}{2} \right) \right\}. \end{aligned} \tag{3.9}$$

Ces trois formules montrent que l'on peut comprendre cette espérance conditionnelle de différentes manières : d'une part, on peut la voir comme un facteur multiplicatif de $P_{i,j}$ ou de $I_{i,j}$, proche de la méthode Chain Ladder, mais la dernière formule peut aussi nous la faire voir comme une moyenne pondérée entre le triangle des paiements et celui des charges.

Ainsi, si on considère de manière simultanée les paiements cumulés ainsi que les charges sinistres, on obtient un "facteur de correction" :

$$\exp(\beta_j(\log \frac{I_{i,j}}{P_{i,j}} - (\mu_j - \eta_j) - \sum_{l=j+1}^J \frac{\sigma_l^2}{2})), \tag{3.10}$$

Ce facteur apparaît donc comme une correction à l'estimateur $E[P_{i,j}|\mathcal{B}_j^P]$ du modèle classique Chain Ladder. On observe qu'un ratio $\frac{I}{P}$ élevé donne un "facteur de correction" élevé à l'estimateur de la méthode Chain Ladder. On peut ainsi faire un parallèle avec la méthode Munich Chain Ladder qui propose également une correction à l'estimateur de Mack basé sur le ratio $\frac{I}{P}$.

Nous connaissons donc la loi explicite des paiements ultimes, il nous reste cependant à estimer les paramètres de cette loi. Dans le cadre d'un modèle fréquentiste, on cherche à maximiser la vraisemblance. Comme on travaille avec des lois normales, on utilisera donc la moyenne empirique et la variance empirique qui tiendront lieu d'estimateurs de maximum de vraisemblance.

On obtient alors :

$$\widehat{\Phi}_j = \frac{1}{J-j+1} \sum_{i=1}^{J-j} \log\left(\frac{P_{i,j}}{P_{i,j-1}}\right),$$

et

$$\hat{\sigma}_j = \frac{1}{J-j} \sum_{i=1}^{J-j} \left(\log\left(\frac{P_{i,j}}{P_{i,j-1}}\right) - \widehat{\Phi}_j \right)^2.$$

Les expressions de $\widehat{\Psi}_j$ et $\widehat{\tau}_j$ sont respectivement similaires à celles de $\widehat{\Phi}_j$ et $\widehat{\sigma}_j$. Il suffit simplement de remplacer les valeurs de P par les valeurs de I .

Maintenant que nous avons étudié le cas fréquentiste, intéressons nous au cas bayésien sur lequel repose le modèle PIC.

3.1.4 La méthode PIC

On définit tout d'abord :

- $D_j^P = \{P_{i,j} : i+j \leq J\}$
- $D_j^I = \{I_{i,j} : i+j \leq J\}$
- $D_j = D_j^I \cup D_j^P$

Sous les caractéristiques de ce modèle on a

- $\sigma_0, \dots, \sigma_J$ déterministes
- $\tau_0, \dots, \tau_{J-1}$ déterministes
- $\Phi_m \sim \mathcal{N}(\phi_m, s_m^2)$ pour $m \in \{0, \dots, J\}$
- $\Psi_n \sim \mathcal{N}(\psi_n, t_n^2)$ pour $n \in \{0, \dots, J-1\}$

Dans le modèle fréquentiste de la section précédente, nous avons raisonné dans un cadre d'utilisation de paramètres inconnus mais déterministes de Θ que nous avons estimé.

Dans cette section, les paramètres de Θ sont des variables aléatoires sur lesquelles on pose une loi a priori gaussienne pour Φ_m et pour Ψ_n , et nous supposons que σ_m et τ_n sont déterministes et connus. On peut penser à plusieurs méthodes pour obtenir σ_m et τ_n , comme par exemple adopter une approche empirique. Nous pourrions alors analytiquement déterminer la loi a posteriori, nécessaire aux calculs et interprétations du modèle explicite.

Nous présenterons tout d'abord dans une première partie le résultat permettant de compléter le triangle inférieur, puis, dans une seconde partie, nous en détaillerons les calculs.

La valeur estimative de chacun des montants ultimes du triangle des paiements est donnée par :

$$\begin{aligned} E[P_{i,J} | \mathcal{D}_j] &= P_{i,J-i}^{1-\beta_{J-i}} I_{i,J-i}^{\beta_{J-i}} \exp \left\{ (1-\beta_{J-i}) \sum_{l=J-i+1}^J \phi_l^{\text{post}} + \beta_{J-i} \sum_{l=J-i}^{J-1} \psi_l^{\text{post}} \right\} \\ &\times \exp \left\{ (1-\beta_{J-i}) \frac{v_J^2 - w_{J-i}^2}{2} + \frac{(s_i^{\text{post}})^2}{2} \right\}, \end{aligned} \quad (3.11)$$

$$\text{où } v_j = \sum_{m=0}^J \sigma_m^2 - \sum_{n=j}^{J-1} \tau_n^2,$$

$$w_j = \sum_{m=0}^j \sigma_m^2,$$

$$\beta_j = \frac{\sum_{m=j+1}^J \sigma_m^2}{\sum_{m=j+1}^J \sigma_m^2 + \sum_{n=j}^{J-1} \tau_n^2} \geq 0.$$

De plus on a, $\left(s_i^{\text{post}}\right)^2 = (e_i)' \Sigma(D_j) e_i$,
 où $e_i = (0, \dots, 0, 1 - \beta_{J-i}, \dots, 1 - \beta_{J-i}, 0, \dots, 0, \beta_{J-i})' \in R^{2J+1}$.

ϕ_l^{post} et ψ_l^{post} seront explicités dans la partie qui suit.

Il est important de noter que nous venons de définir un modèle stochastique, du nom de PIC qui considèrerait simultanément les données sur les paiements cumulés ainsi que celles sur les charges sinistres pour prédire les provisions.

Non seulement nous avons préservé la structure Chain Ladder en ayant utilisé "un facteur de correction" aux facteurs de développement des paiements cumulés, mais nous avons aussi dépassé le cadre déterministe en ajoutant une composante bayésienne au résultat précédent.

De plus, pour déterminer les provisions, nous avons donc utilisé simultanément deux leviers d'informations : celui des paiements cumulés ainsi que celui des charges sinistres pour aboutir à des montants de charges ultimes identiques.

3.1.5 Détails du résultat

Intéressons-nous pour commencer à la loi a posteriori de $P_{i,j} | \{D_j^P\}$:

En ce qui concerne les paiements cumulés, nous introduisons les paramètres $\Phi = (\Phi_0, \dots, \Phi_J)$,

La loi a posteriori de $\Phi | \{D_j^P\}$ est donnée (en posant $P_{i,-1} = 1$) par :

$$u(\Phi | D_j^P) \propto \prod_{j=0}^J \prod_{i=0}^{J-j} \exp \left\{ -\frac{1}{2\sigma_j^2} \left(\Phi_j - \log \frac{P_{ij}}{P_{i,j-1}} \right)^2 \right\} \\ \times \prod_{j=0}^J \exp \left\{ -\frac{1}{2s_j^2} (\Phi_j - \phi_j)^2 \right\},$$

Elle a également des composantes indépendantes avec

$$\Phi_j | \{D_j^P\} \sim \mathcal{N} \left(\phi_j^{P,\text{post}}, \left(s_j^{P,\text{post}} \right)^2 \right), \text{ où}$$

$$\phi_j^{P,\text{post}} = \gamma_j^P \frac{1}{J-j+1} \sum_{i=0}^{J-j} \log \frac{P_{ij}}{P_{i,j-1}} + (1 - \gamma_j^P) \phi_j,$$

$$\left(s_j^{P,\text{post}} \right)^2 = \left(\frac{1}{s_j^2} + \frac{J-j+1}{\sigma_j^2} \right)^{-1},$$

$$\gamma_j^P = \frac{J-j+1}{J-j+1 + \frac{\sigma_j^2}{s_j^2}}.$$

L'espérance de la loi a posteriori est donc une moyenne pondérée entre la moyenne a priori ϕ_j et la moyenne empirique $\bar{\phi}_j$ où

$$\bar{\phi}_j = \frac{1}{J-j+1} \sum_{i=0}^{J-j} \log \frac{P_{i,j}}{P_{i,j-1}}.$$

.

Une fois la loi a posteriori de $P_{i,j} | \{D_j^P\}$ complètement déterminée, les moments peuvent également être calculés, et une simulation Monte Carlo, nous fournirait alors la distribution empirique a posteriori du vecteur $(P_{1J}, \dots, P_{JJ}) | \{D_j^P\} = (I_{1J}, \dots, I_{JJ}) | \{D_j^P\}$, ainsi que notre prédiction

$$E \left[P_{i,J} | D_j^P \right] = P_{i,J-i} \prod_{l=J-i+1}^J \exp \left\{ \phi_l^{P, \text{post}} + \frac{\sigma_l^2}{2} + \frac{(s_l^{P, \text{post}})^2}{2} \right\},$$

Nous allons désormais nous concentrer sur les paiements encourus $I_{i,j}$.
Nous définissons les paramètres

$$\Psi_j = -\mu_j = -\sum_{j=0}^J \Phi_j \sim \mathcal{N} \left(\psi_j, t_j^2 \right), \text{ indépendant de } \Psi_0, \dots, \Psi_{J-1}$$

$$\psi_J = -\sum_{j=0}^J \phi_j,$$

$$t_j^2 = \sum_{j=0}^J s_j^2,$$

que nous utiliserons pour $I_{i,j}$.

La loi a posteriori de $\Psi | \{D_j^I\}$ est donnée par :

$$\begin{aligned} u(\Psi | D_j^I) &\propto \prod_{i=0}^J \exp \left\{ -\frac{1}{2v_{j-i}^2} \left(\sum_{n=J-i}^J \Psi_n + \log I_{i,J-i} \right)^2 \right\} \\ &\times \prod_{j=0}^{J-1} \prod_{i=0}^{J-j-1} \exp \left\{ -\frac{1}{2\tau_j^2} \left(\Psi_j + \log \frac{I_{i,j}}{I_{i,j+1}} \right)^2 \right\} \\ &\times \prod_{j=0}^J \exp \left\{ -\frac{1}{2t_j^2} (\Psi_j - \psi_j)^2 \right\}. \end{aligned}$$

Ainsi, la loi a posteriori de $\Psi | D_j^I$ est une loi gaussienne multidimensionnelle de moyenne $\psi^{\text{post}} (D_j^I)$ et de matrice de covariance $\Sigma (D_j^I)$ où
 $\psi^{\text{post}} (D_j^I) = (\psi_0^{I, \text{post}}, \dots, \psi_J^{I, \text{post}})'$, obtenue par $\psi^{\text{post}} (D_j^I) = \Sigma (D_j^I) (b_0^I, \dots, b_J^I)'$, où le vecteur (b_0^I, \dots, b_J^I) est donné par $b_j^I = t_j^{-2} \psi_j - \tau_j^{-2} \sum_{i=0}^{J-j-1} \log \frac{I_{ij}}{I_{i,j+1}} - \sum_{i=0}^j v_i^{-2} \log I_{j-i,i}$.

La loi a posteriori de $P_{iJ} = I_{iJ}$, sachant D_j^I , est ainsi déterminée.

On obtient alors :

$$E [P_{ij} | D_j^I] = I_{ij-i}^{1-\alpha_{j-i}} \exp \left\{ (1 - \alpha_{j-i}) \sum_{l=j-i}^{J-1} \psi_l^{I, \text{post}} + \alpha_{J-i} \left(-\psi_l^{I, \text{post}} + \frac{v_l^2}{2} \right) + \frac{(s_i^{I, \text{post}})^2}{2} \right\}, \quad (3.12)$$

$$\text{ou } (s_i^{I, \text{post}})^2 = (e_i^I)' \Sigma(D_j^I) e_i^I$$

$$\text{avec } e_i^I = (0, \dots, 0, 1 - \alpha_{j-i}, \dots, 1 - \alpha_{j-i}, 1 - \alpha_{j-i})' \in R^{J+1}.$$

On peut désormais relier les paiements cumulés aux paiements encourus :

On détermine la loi a posteriori de : $\Theta | \{D_j\}$

On observe que

$$\log I_{i,j} | \{g_j^P, \theta\} \sim \mathcal{N} \left(\mu_j - \eta_j, v_j^2 - w_j^2 \right) = \mathcal{N} \left(\sum_{m=j+1}^J \Phi_m - \sum_{n=j}^{J-1} \Psi_n, \sum_{m=j+1}^J \sigma_m^2 + \sum_{n=j}^{J-1} \tau_n^2 \right).$$

En posant $P_{i,-1} = 1$,

$$\begin{aligned}
I_{D_J}(\Theta) = & \prod_{j=0}^J \prod_{i=0}^{J-j} \frac{1}{\sqrt{2\pi}\sigma_j P_{i,j}} \exp \left\{ -\frac{1}{2\sigma_j^2} \left(\Phi_j - \log \frac{P_{ij}}{P_{i,j-1}} \right)^2 \right\} \\
& \times \prod_{i=1}^J \frac{1}{\sqrt{2\pi} \left(v_{j-i}^2 - w_{j-i}^2 \right) I_{i,J-i}} \\
& \times \exp \left\{ -\frac{1}{2 \left(v_{j-i}^2 - w_{j-i}^2 \right)} \left(\mu_{j-i} - \eta_{j-i} - \log \frac{I_{i,j-1}}{P_{i,j-1}} \right)^2 \right\} \\
& \times \prod_{j=0}^{J-1} \prod_{i=0}^{J-j-1} \frac{1}{\sqrt{2\pi}\tau_j I_{i,j}} \exp \left\{ -\frac{1}{2\tau_j^2} \left(\Psi_j + \log \frac{I_{i,j}}{I_{i,j+1}} \right)^2 \right\}.
\end{aligned} \tag{3.13}$$

La loi a posteriori de $u(\Theta|D_J)$ est donc

$$u(\Theta|D_J) \propto I_{D_J}(\Theta) \prod_{m=0}^J \exp \left\{ -\frac{1}{2s_m^2} (\Phi_m - \phi_m)^2 \right\} \times \prod_{n=0}^{J-1} \exp \left\{ -\frac{1}{2t_n^2} (\psi_n - \Psi_n)^2 \right\}. \tag{3.14}$$

On en conclut que la distribution a posteriori $u(\Theta|D_J)$ suit une loi gaussienne multidimensionnelle de moyenne à posteriori $\theta^{\text{post}}(D_J)$ et de matrice de covariance a posteriori $\Sigma(D_J)$.

L'inverse de la matrice de covariance $\Sigma(D_J)^{-1} = (a_{n,m})_{0 \leq n,m \leq 2J}$ est donnée par :

$$\begin{aligned}
a_{n,m} &= (s_n^{-2} + (J-n+1)\sigma_n^{-2}) 1_{\{n=m\}} + \sum_{i=0}^{(n-1) \wedge (m-1)} (v_i^2 - w_i^2)^{-1}, \\
a_{j+1+n, j+1+m} &= (t_n^{-2} + (J-n)\tau_n^{-2}) 1_{\{n=m\}} + \sum_{i=0}^{n+m} (v_i^2 - w_i^2)^{-1}, \\
a_{n, J+1+m} &= a_{j+1+m, n} = -\sum_{i=0}^{(n-1) \wedge m} (v_i^2 - w_i^2)^{-1},
\end{aligned} \tag{3.15}$$

avec $0 \leq n \leq J, 0 \leq m \leq J-1$

La moyenne a posteriori $\theta^{\text{post}}(D_J) = (\phi_0^{\text{post}}, \dots, \phi_J^{\text{post}}, \psi_0^{\text{post}}, \dots, \psi_{J-1}^{\text{post}})'$ est obtenue par : $\theta^{\text{post}}(D_J) = \Sigma(D_J)(c_0, \dots, c_J, b_0, \dots, b_{J-1})'$, avec le vecteur $(c_0, \dots, c_J, b_0, \dots, b_{J-1})$, donné par $c_j = s_j^{-2}\phi_j + \sigma_j^{-2} \sum_{i=0}^{J-j} \log \frac{P_{ij}}{P_{i,j-1}} + \sum_{i=J-j+1}^J \left(v_{j-i}^2 - w_{j-i}^2 \right)^{-1} \log \frac{I_{i,j-i}}{P_{i,j-i}}$,
et $b_j = t_j^{-2}\psi_j - \tau_j^{-2} \sum_{i=0}^{J-j-1} \log \frac{I_{i,j}}{I_{i,j+1}} - \sum_{i=J-j}^J \left(v_{j-i}^2 - w_{j-i}^2 \right)^{-1} \log \frac{I_{i,j-i}}{P_{i,j-i}}$,

On trouve ainsi le résultat présenté dans la partie précédente :

$$\begin{aligned}
E[P_{i,J}|D_J] &= P_{i,J-i}^{1-\beta_{J-i}} I_{i,J-i}^{\beta_{J-i}} \exp \left\{ (1-\beta_{J-i}) \sum_{l=J-i+1}^J \phi_l^{\text{post}} + \beta_{J-i} \sum_{l=J-i}^{J-1} \psi_l^{\text{post}} \right\} \\
&\times \exp \left\{ (1-\beta_{J-i}) \frac{v_{j-i}^2 - w_{j-i}^2}{2} + \frac{(s_i^{\text{post}})^2}{2} \right\},
\end{aligned} \tag{3.16}$$

où $(s_i^{\text{post}})^2 = (e_i)' \Sigma(D_J) e_i$ et

$e_i = (0, \dots, 0, 1 - \beta_{J-i}, \dots, 1 - \beta_{J-i}, 0, \dots, 0, \beta_{J-i})' \in R^{2J+1}$.

3.1.6 Détails du calcul

Dans la pratique, et dans le package R, le calcul des valeurs ultimes est légèrement simplifié. En effet, on ne dispose pas toujours d'avis d'experts ou de suffisamment d'informations sur les paramètres a priori. On choisit donc d'utiliser un modèle à valeurs a priori non informatives pour Φ_m et Ψ_n . Autrement dit, on pose $s_m^2 \rightarrow \infty$ et $t_n^2 \rightarrow \infty$. Cela implique donc que $\gamma_m^P = \gamma_m^I = 1$. Notre résultat final ne dépend donc que de \mathcal{D}_J^P et \mathcal{D}_J^I et notre formule finale ne change pas :

$$E[P_{i,J}|\mathcal{D}_J] = P_{i,J-i}^{1-\beta_{J-i}} I_{i,J-i}^{\beta_{J-i}} \exp \left\{ (1 - \beta_{J-i}) \sum_{l=J-i+1}^J \phi_l^{\text{post}} + \beta_{J-i} \sum_{l=J-i}^{J-1} \psi_l^{\text{post}} \right\} \times \exp \left\{ (1 - \beta_{J-i}) \frac{v_J^2 - w_{J-i}^2}{2} + \frac{(s_i^{\text{post}})^2}{2} \right\}, \quad (3.17)$$

Cependant les expressions de ϕ_l^{post} et ψ_l^{post} se modifient avec :

$$\phi_l^{\text{post}} = \frac{1}{J-l+1} \sum_{i=0}^{J-l} \log \frac{P_{i,l}}{P_{i,l-1}}, \quad (3.18)$$

$$\psi_l^{\text{post}} = \frac{1}{J-l+1} \sum_{i=0}^{J-l} \log \frac{I_{i,l}}{I_{i,l-1}}, \quad (3.19)$$

où nous avons simplement remplacé dans les expressions de la partie 3.1.5 les γ_l^P et γ_l^I par 1. On retrouve ainsi des paramètres similaires à ceux de la moyenne empirique.

De plus, dans le package R, les paramètres τ_j et σ_j sont calculés grâce à la formule de la variance empirique.

3.2 Mise en oeuvre de la méthode PIC

3.2.1 Mise en oeuvre sur notre 1er jeu de données

Nous allons désormais tester la méthode PIC sur notre premier jeu de données.

De même que dans la partie précédente sur la méthode Munich Chain Ladder, nous avons remplacé les valeurs manquantes de nos deux tableaux (surlignés en jaune).

On rappelle le jeu de données :

	0	1	2	3	4	5	6	7	8	9
1996	118 989	167 063	179 750	186 908	190 860	194 217	198 879	199 726	200 403	201 022
1997	123 767	173 770	186 967	195 488	201 015	204 518	206 394	207 930	209 766	
1998	121 147	170 092	183 403	189 427	194 477	198 577	202 297	206 513		
1999	112 074	167 919	179 126	186 632	191 689	195 838	201 103			
2000	111 317	155 731	166 886	173 593	182 013	186 220				
2001	112 656	160 725	176 065	183 339	188 226					
2002	107 608	147 991	159 641	165 503						
2003	100 534	137 948	146 703							
2004	101 198	135 817								
2005	100 197									

TABLE 3.1: Tableau des montants payés (P)

	0	1	2	3	4	5	6	7	8	9
1996	130 172	182 764	196 644	204 474	204 475	206 653	205 188	204 714	203 354	204 173
1997	143 403	201 340	216 630	217 271	217 123	214 912	214 647	215 138	215 388	
1998	153 992	216 206	211 180	212 559	212 269	212 221	212 738	215 036		
1999	201 502	219 920	213 742	213 227	213 554	213 097	213 358			
2000	191 830	199 280	192 297	192 367	194 919	195 855				
2001	198 037	203 493	200 067	200 141	200 804					
2002	176 676	184 838	179 912	179 606						
2003	172 686	174 617	172 605							
2004	159 220	165 520								
2005	162 815									

TABLE 3.2: Tableau des charges (I)

Sans détailler les calculs, nous obtenons les résultats suivants :

Année	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Ult	201 021	211 915	210 578	209 746	195 126	200 771	181 211	168 551	167 347	173 426
Dev	100 %	98 %	98 %	96 %	95 %	94 %	91 %	87 %	81 %	57%
Prov	0	2 149	4 065	8 642	890	12 544	15 708	21 848	31 529	73 229

TABLE 3.3: Tableau récapitulatif méthode PIC

On obtient donc pour la méthode PIC, un développement total de 90,6% ainsi qu'une provision totale de 178 625.

	CL	MCL	PIC
Dev. Totale	91.4 %	92.1 %	90.6 %
IBNR	162 775	147 299	178 625
Prov. Ult.	1 903 845	1 888 369	1 919 696

TABLE 3.4: Tableau comparatif et récapitulatif du premier jeu de données

On observe que la méthode PIC possède le développement total le plus faible parmi les trois méthodes étudiées.

3.2.2 Mise en oeuvre sur le 2 ème jeu de données

Mettons désormais en oeuvre les trois méthodes sur un deuxième jeu de données. Ce jeu de données est très similaire au premier et comporte de même un triangle des montants payés ainsi qu'un triangle des charges :

	0	1	2	3	4	5	6	7	8	9
1996	30 095	51 193	55 104	55 825	56 173	56 234	56 257	56 196	56 452	56 494
1997	28 999	49 329	53 098	53 567	53 820	53 864	54 112	54 090	54 005	
1998	18 854	32 071	33 067	33 517	33 647	33 833	34 109	34 129		
1999	18 748	73 295	85 681	88 872	90 183	90 559	90 796			
2000	19 771	34 356	37 249	37 875	37 951	38 244				
2001	22 784	37 150	40 532	40 941	41 125					
2002	23 954	41 796	46 283	47 083						
2003	28 086	45 732	48 575							
2004	29 247	51 627								
2005	29 617									

TABLE 3.5: Tableau des montants payés (P)

	0	1	2	3	4	5	6	7	8	9
1996	41 397	58 250	56 830	56 519	56 888	56 961	56 851	56 832	56 899	56 886
1997	39 940	56 200	54 829	54 560	54 559	54 731	54 494	54 177	53 792	
1998	25 644	36 084	35 646	35 293	35 294	35 175	35 444	35 348		
1999	35 873	95 253	93 380	93 508	91 537	91 745	91 804			
2000	38 041	41 477	40 119	39 898	39 459	39 691				
2001	39 163	43 436	42 208	41 304	41 968					
2002	49 976	52 538	51 976	52 365						
2003	48 008	50 822	50 765							
2004	69 947	73 105								
2005	54 745									

TABLE 3.6: Tableau des charges (I)

On observe une singularité dans notre jeu de données : en effet, l'année 1999 semble perturbée au sens où entre 1999 et 2000, les montants payés semblent avoir été multipliés par 4 (surlignés en rouge). Nous avons donc décidé de ne pas prendre en compte l'année 1999 pour compléter les 6 valeurs manquantes (surlignés en jaune). Nous reviendrons dans la sous partie suivante sur le caractère "exceptionnel" de l'année 1999. Dans cette sous partie nous comparons les 3 méthodes sur le jeu de données.

Voici ci-dessous des tableaux comparatifs des paiements ultimes, provisions globales et provisions par année d'exercice pour chacune des méthodes Chain Ladder, Munich Chain Ladder et PIC :

	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
CL	56 494	54 045	34 208	90 964	38 443	41 486	47 847	50 355	58 981	67 382
MCL	56 494	54 037	34 350	90 891	38 668	41 480	49 483	49 899	66 471	64 474
PIC	56 494	54 046	34 529	91 155	38 767	41 682	48 556	50 199	58 442	64 717

TABLE 3.7: Tableau des paiements ultimes par année

	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
CL	0	28	216	422	277	321	448	812	7 354	27 489
MCL	0	32	221	95	423	354	2 400	1 323	14 844	34 856
PIC	0	41	399	358	523	556	1 472	1 623	6 814	35 099

TABLE 3.8: Tableau des provisions par année

	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
CL	1.00	1.00	1.00	1.00	0.99	0.99	0.98	0.96	0.88	0.44
MCL	1.00	1.00	0.99	1.00	0.99	0.99	0.95	0.97	0.78	0.46
PIC	1.00	1.00	0.99	1.00	0.99	0.99	0.97	0.97	0.88	0.46

TABLE 3.9: Tableau de développement par année

	CL	MCL	PIC
Dev. Totale	91.0 %	90.0 %	91.3 %
Prov totales	48 506	54 550	46 889
Païement Ult.	540 205	546 299	538 589

TABLE 3.10: Tableau comparatif et récapitulatif du second jeu de données

La méthode de Mack nous permet de déterminer une incertitude globale. Ici, l'incertitude globale est de 27 761, ce qui représente pour la méthode Chain Ladder une incertitude de près de 57 % de la provision globale .

Cette incertitude est bien plus importante que celle du premier jeu de données qui était seulement de 9 % : cela peut s'expliquer par le caractère "exceptionnel" de l'année 1999.

L'intervalle de prédiction des provisions est [20 745, 76 268].

La méthode Munich Chain Ladder ainsi que la méthode PIC appartiennent bien toutes deux à cet intervalle de confiance.

On remarque qu'à l'inverse du 1er jeu de données, la provision obtenue par la méthode Munich Chain Ladder est la plus importante, tandis que celle obtenue par la méthode PIC est la plus faible.

La provision obtenue par la méthode Chain Ladder est quant à elle toujours située entre les deux.

Si on s'intéresse au développement total , on observe que les résultats de la méthode Chain Ladder et de la méthode PIC sont très proches : une différence de développement total de seulement 0.3 points.

La méthode Munich Chain Ladder a une différence bien plus marquée avec les autres méthodes : elle admet 1 point de moins que la méthode Chain Ladder et 1.3 point de moins que la PIC.

Concernant le tableau de développement par année 3.9, on remarque que de 1996 à 2001, la différence de développement des paiements est quasiment nulle entre les méthodes : le développement est proche de 100 % pour chacune des méthodes. Les résultats sont de même très similaires de 2002 à 2005.

La différence la plus significative repose sur l'année 2004 : les méthodes Chain Ladder et PIC y ont un développement total égal mais différent de 10 % de celui de la méthode Munich Chain Ladder.

En terme de provisions, la méthode Munich Chain Ladder prévoit une provision près de 2 fois supérieure à celle de la méthode Chain Ladder et plus de 2 fois supérieure à celle de la méthode PIC . La différence de provisions pour l'année 2004 est de 7 490 entre la méthode Munich Chain Ladder et Chain Ladder et de 8 030 entre la Munich Chain Ladder et la PIC.

Cette différence de provisions pour l'année 2006 est comparable à la différence de provision totale entre les méthodes. On peut donc légitimement s'avancer sur le fait que la différence entre la méthode

Munich Chain Ladder et les deux autres méthodes vient du développement de l'année 2006.

On pourrait penser que la différence s'évalue au niveau de f_1 .

Cependant, les provisions pour l'année 2005 sont similaires pour les 3 méthodes.

Cela signifie donc que la différence s'évalue au niveau de f_0 dont la valeur est sans doute plus faible pour la méthode Munich Chain Ladder que pour les deux autres méthodes.

3.2.3 Focus sur le deuxième jeu de données

Revenons sur le 2^{ème} jeu de données et le caractère "exceptionnel" de l'année 1999. Dans la partie précédente, nous avons directement appliqué les trois méthodes à notre jeu de données pour tester leur robustesse et surtout s'intéresser à "leur façon de réagir" face à une singularité. Cependant, il n'est pas raisonnable en pratique de traiter l'année 1999 de manière identique aux autres années.

On observe sur le tableau suivant que les valeurs des facteurs de développement $f_{1999,j}$ sont bien plus importants pour $j=0$ et $j=1$. On observe que l'année 1999 a une influence importante sur les facteurs de développement f_0 qui est inférieur 0.3 lorsqu'on ne prend pas en compte l'année 1999.

	j	0	1	2	3	4	5	6	7	8
1996					1.01	1.00	1.00	1.00	1.00	1.00
1997				1.01	1.00	1.00	1.00	1.00	1.00	
1998			1.03	1.01	1.00	1.01	1.01	1.00		
1999	3.91	1.17	1.04	1.01	1.00	1.00				
2000	1.74	1.08	1.02	1.00	1.01					
2001	1.63	1.09	1.01	1.00						
2002	1.74	1.11	1.02							
2003	1.63	1.06								
2004	1.77									
f_j Chain Ladder	1.99	1.10	1.02	1.01	1.00	1.00	1.00	1.00	1.00	1.00
f_j sans 1999	1.70	1.08	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00

TABLE 3.11: Tableau comparatif des facteurs de développement des méthodes Chain Ladder et PIC

Les montants de l'année 1999 sont colossaux : pour presque chaque colonne, le coût à l'année 1999 est plus important que la somme des coûts des années 1997 et 1998 (3.6 et 3.5).

En pratique, nous pourrions traiter ce jeu de donnée de deux façons :

- Si l'importance de des montants de l'année 1999 provient de l'apparition d'un évènement en particulier, on pourrait retrancher aux montants du tableau relatifs à 1999, ceux liés à cet évènement.
- Si c'est l'année qui est exceptionnelle dans sa globalité, il serait plus raisonnable de ne pas prendre en compte cette année et de la traiter à part.

Nous ne disposons malheureusement pas d'assez d'informations pour conclure, comme par exemple la donnée concernant la diversité des produits que notre jeu de données serait susceptible d'avoir.

Nous savons cependant que notre jeu de données est composé d'une partie "dégâts non corporels" (A.2) et d'une partie "dégâts corporels" (A.1) (à retrouver en annexe).

Intéressons-nous à ces deux sous-triangles pour mieux comprendre l'année 1999.

On observe tout d'abord que les dégâts corporels ne représentent que près de 20 % des montants payés, là où pour l'année 1999, ils en représentent près de 13% (voir table A.3 en annexe), soit le pourcentage

le plus faible parmi toutes les années du triangle.

Concernant les facteurs de développement $f_{i,j}$ des tables A.4 et A.5 (en annexe), ceux relatifs au tableau des dégâts non corporels sont, à l'année 1999, entre $j=0$ et $j=1$ de 0.46 points supérieurs à ceux des dégâts corporels. La variance des facteurs de développement $f_{i,0}$ relatifs au tableau des dégâts non corporels, est, quant à elle, près de deux fois supérieure à celle des dégâts corporels.

On peut donc supposer que le caractère exceptionnel de l'année 1999 vient principalement des dégâts non corporels.

Mettons en oeuvre la méthode Chain Ladder sans prendre en compte l'année 1999.

	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Ultime sans 1999	56 494	54 045	34 208	90 964	38 461	41 491	47 714	49 871	57 054	55 676
Ultime avec 1999	56 494	54 045	34 208	90 964	38 443	41486	47 847	50 355	58 981	67 382
Dev sans 1999	1.00	1.00	1.00	1.00	0.99	0.99	0.99	0.97	0.90	0.53
Dev avec 1999	1.00	1.00	1.00	1.00	0.99	0.99	0.98	0.96	0.88	0.44
Provision sans 1999	0	40	78	167	216	365	630	1 295	5 426	26 058
Provision avec 1999	0	28	216	422	277	321	448	812	7 354	27 489

TABLE 3.12: Tableau comparatif de la méthode Chain Ladder avec et sans l'année 1999

	Avec 1999	Sans 1999
Dev. Totale	91.0 %	93.5 %
Prov totale	48 506	34 280
Paielement Ult.	540 205	525 978

TABLE 3.13: Tableau comparatif et récapitulatif

On remarque que l'année sans 1999 a des facteurs de développement totaux supérieurs comparés à l'année avec 1999.

De plus, le développement total y est plus important (93.5 %) tandis que la provision totale y est plus faible (34 280).

Comme attendue, on trouve une valeur de provisions totale beaucoup plus faible que celle des 3 méthodes utilisées dans la partie précédente (3.10).

On observe que l'on se situe bien dans l'intervalle de confiance [20 745,76 268] calculé précédemment avec la méthode de Mack. Si on avait calculé ici l'intervalle de confiance, on en aurait trouvé un bien moins large qui aurait représenté seulement 10% des provisions globales.

Pour conclure, la singularité de l'année 1999 de notre jeu de données est très intéressante à prendre en compte dans la comparaison de nos 3 méthodes. La différence notable des résultats obtenus avec la méthode Chain Ladder appliquée aux deux cas, montre bien qu'il ne serait pas raisonnable de la tenir à l'écart.

Conclusion

Les provisions jouent un rôle important dans le monde de l'assurance. Elles décrivent la part la plus importante pour une assurance non-vie. Le rôle principal d'un actuaire est alors, grâce à des méthodes de provisionnement, d'estimer le montant de provisions nécessaires jusqu'au développement ultime des sinistres, pour que l'assureur puisse honorer ses engagements vis-à-vis de ses assurés.

Au sein de ce mémoire, nous avons étudié 3 principales méthodes de provisionnement : la méthode déterministe Chain Ladder alias modèle de Mack, les méthodes stochastiques Munich Chain Ladder et PIC.

L'étude de ces différentes méthodes de provisionnement nous a montré la complexité quant au choix de l'une ou l'autre. Aucune n'étant parfaite, nous ne pouvons pas dire que l'utilisation d'une méthode est préférable à toute autre dans tous les cas.

Non seulement le choix d'une méthode dépend avant tout de son adéquation aux données, mais il serait également important et judicieux, dans une optique de gestion des risques, de réaliser un suivi des résultats qu'elles fournissent et de les confronter à la réalité observée a posteriori ou encore les confronter aux résultats d'autres méthodes.

Il y a également lieu de vérifier les hypothèses et la validité de la mise en place du modèle. Si un modèle est inadéquat aux données par exemple, un modèle qui représenterait mal la réalité, il serait difficile d'évaluer le risque du modèle voire de quantifier l'intégralité de ceux auxquels la compagnie pourrait être soumise lors de l'évaluation des provisions.

La méthode la plus répandue pour estimer les provisions est la méthode déterministe Chain Ladder. Le principal avantage de cette méthode est sa simplicité de mise en œuvre.

Néanmoins, elle présente plusieurs inconvénients : l'hypothèse d'indépendance des années de survie n'est pas souvent vérifiée car dans la réalité le déroulement des sinistres n'est pas identique d'une survie à l'autre.

De plus, il existe une incertitude importante sur les survies puisque le calcul de la charge ultime s'appuie sur le produit des estimations des facteurs de développement.

L'incertitude est alors d'autant plus grande que le risque est long (comme c'est le cas des assurances de responsabilité civile ou les assurances de construction garantie décennale par exemple).

Enfin, la méthode Chain Ladder est une méthode déterministe dans le sens où elle ne définit pas de modèle probabiliste permettant de mesurer l'incertitude autour du montant de provisions estimé : le modèle de Mack aura donc été introduit en tant que modèle stochastique pour évaluer l'erreur d'estimation de la méthode Chain Ladder.

En outre, les prédictions de nos provisions seraient idéales si, pour les trouver, on prenait en considération simultanément, et au travers d'un modèle mathématique consistant, différentes sources d'informations telles que les montants payés, les charges sinistres, ou même le taux de sinistre antérieur. Le mieux serait même d'attribuer à chacune de ces informations un poids de crédibilité qui serait plus ou moins grand en fonction de « l'accessibilité » de l'information. Cependant, cela s'avèrerait difficile.

C'est la raison pour laquelle la plupart des méthodes de provisionnement sont basées sur une seule source d'information (par exemple les montants payés OU les charges sinistres), ou bien, peuvent également s'appliquer séparément sur les différentes sources d'informations mises à notre disposition, et ensuite fusionner les résultats pour obtenir une prédiction de nos provisions.

Par contre, une telle façon de procéder n'autorisera pas à obtenir une incertitude relative fiable, puisque la fusion des résultats aura été réalisée subjectivement, sur l'avis d'experts.

L'une des premières méthodes de provisionnement à avoir réussi à combiner les données sur les paiements cumulés et celles sur les charges sinistres est la méthode Munich Chain Ladder.

Les évaluations des provisions statistiques pouvaient être sensiblement impactées par des charges et des règlements cumulés dont les développements n'étaient pas convergents. Son objectif était donc de réduire l'écart entre les prédictions issues des Chain Ladder appliquées aux triangles des montants payés et des charges sinistres, en ajustant les facteurs de développement Chain Ladder avec le ratio $\frac{P}{I}$.

De plus, afin de garantir le caractère indépendant et identiquement distribué de ses variables, la méthode Munich Chain Ladder utilise les résidus conditionnels de ces variables.

Cependant, la particularité de cette méthode repose sur le fait qu'elle implique des estimations de paramètres, dont la précision est difficile à quantifier dans le cadre d'un modèle stochastique consistant.

Enfin, la dernière méthode évoquée fut la méthode PIC. L'idée principale est de combiner un modèle lognormale de provisionnement appliqué au triangle des montants payés, et un modèle de provisionnement bayésien appliqué au triangle des charges, pour aboutir à une charge ultime totale identique. Utilisant une approche bayésienne pour modéliser nos paramètres, la méthode PIC nous donne alors la distribution « prédictive » complète de notre variable de provision R . Non seulement le modèle PIC nous permet de calculer de manière analytique le montant des provisions, mais aussi l'erreur de prédiction quadratique moyenne associée.

La méthode PIC fournira alors un unique estimateur pour les réserves de paiements, basé sur nos 2 sources d'informations.

Cependant, il est clair que la méthode PIC s'avère terriblement fastidieuse car des plus techniques et des plus calculatoires.

Comme nous avons pu le voir à travers notre deuxième jeu de données, ces méthodes ne sont pas toujours très robustes quand il s'agit de jeux de données comportant des années à caractère exceptionnel, ou des mouvements bruts de tendance. Elles sont performantes dès lors que le passé est régulier, le présent et le futur structurellement peu différents du passé, et le domaine étudié globalement peu volatile. Deux possibilités s'offrent alors à nous :

- Soit nous décidons, au préalable, de traiter les informations que nous avons en homogénéisant les données pour optimiser la performance de ces méthodes
- Soit nous décidons de mettre à part les données "atypiques" pour les faire bénéficier d'un traitement particulier.

Toutefois, on remarquera que plus les méthodes se révèlent évoluées et complexes, plus leurs points négatifs diminuent.

Ainsi, nous avons bien évoqué 3 méthodes de provisionnement à complexité croissante avec chacune ses forces et ses faiblesses.

Il est intéressant de noter qu'à l'inverse des méthodes de provisionnement agrégé étudiées précédemment, les méthodes ligne à ligne sur données individuelles de sinistres, toutes aussi complexes, prennent, néanmoins, plus facilement en compte, les valeurs extrêmes .

Bibliographie

A.Charpentier, ed. (2014). Computational Actuarial Science with R. Chapman et Hall-CRC.

D.Hindley (2017). Claims Reserving in General Insurance. Cambridge University Press.

Kirschner, Kerley, et Isaacs, (2002). Two approaches to calculating correlated reserves indications across multiple lines of business. 2002 CAS Forum Fall.

Merz et Wuethrich (2010). PIC claims reserving method. Insurance : Mathematics and Economics 46.3, p. 568-579.

Partrat, C.Lecoeur, E.Nessi, J.-M.Nisipasu, E.Reiz, (2008).Provisionnement technique en assurance non-vie : Perspectives actuarielles modernes. Economica.

Quarg, G. et Mack, T. (2008). Munich Chain Ladder : a reserving method that reduces the gap between IBNR projections based on paid losses and IBNR projections based on incurred losses. Variance 2.2.

R Core Team(2019). R : A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. url : [https ://www.R-project.org/](https://www.R-project.org/).

Annexe A

Annexe code R

A.1 Code R Méthode Chain Ladder

Code R pour les différents CC-plots, Figure 1.1 :

```
rm(list=ls())
load("~/Desktop/Memoire actuariat/fretri1auto9605.rda")
total.paid<-fretri1auto9605[["total"]][["paid"]]
total.incur<-fretri1auto9605[["total"]][["incur"]]
###
total.paid
cc1<-total.paid[4:9,1]
plot(total.paid[3:9,2],total.paid[3:9,3],xlim=c(0,190000),ylim=c(0,190000))
title(main = "CC plot pour j=1")
plot(total.paid[2:9,3],total.paid[2:9,4],xlim=c(0,200000),ylim=c(0,200000))
title(main = "CC plot pour j=2")
plot(total.paid[2:9,4],total.paid[2:9,5],xlim=c(0,210000),ylim=c(0,210000))
title(main = "CC plot pour j=3")
```

Code R pour obtenir la table 1.2 et la Figure 1.2 :

```
a<-matrix(nrow=9,ncol=9)
for (i in (1:9)){
  for (j in (1:9)){
    if (is.na(total.paid[i,j+1])){a[i,j]<-NA}
    else{a[i,j]<-total.paid[i,j+1]/total.paid[i,j]}
  }
}
###
xdata<-c(1:9)
plot(xdata, a[,1],type="o", col="blue", ylab="Development Factor",xlab="Time",ylim=c(0.9,1.6))
points(xdata, a[,2], col="red", pch="o")
lines(xdata, a[,2], col="red")
points(xdata, a[,3], col="green", pch="o")
lines(xdata, a[,3], col="green")
points(xdata, a[,4], col="black", pch="o")
lines(xdata, a[,4], col="black")
points(xdata, a[,5], col="pink", pch="o")
lines(xdata, a[,5], col="pink")
legend("topleft",legend=c("j=0","j=1","j=2","j=3","j=4"), col=c("blue","red","green","black","pink"),lty=1,pch="o")
```

Code R pour obtenir les Figures 1.3 et 1.4 :

```
rm(list=ls())
load("~/Desktop/Memoire actuariat/fretri1auto9605.rda")
total.paid<-fretri1auto9605[["total"]][["paid"]]
###
```



```
library(ChainLadder)
mack<-MackChainLadder(total.paid)
plot(mack)
```

A.2 Code R Muncih Chain Ladder

Pour obtenir les différents graphiques et mieux comprendre la méthode Munich Chain Ladder, nous avons cherché à la coder par nous-mêmes . Voici ci-joint le code.

```
rm(list=ls())
load("~/Desktop/Memoire actuariat/fretrilauto9605.rda")
total.paid<-fretrilauto9605[["total"]][["paid"]]
total.incur<-fretrilauto9605[["total"]][["incur"]]
print(total.paid)
print(total.incur)
###
P<-total.paid
I<-total.incur
n<-10
#####
fP<-c()
for(j in 1:(n-1)){
  a<-sum(P[1:(n-j),j+1])/sum(P[1:(n-j),j])
  fP<-append(fP,a)
}
print(fP)
###
fI<-c()
for(j in 1:(n-1)){
  a<-sum(I[1:(n-j),j+1])/sum(I[1:(n-j),j])
  fI<-append(fI,a)
}
print(fI)
#####
phiP<-c()
for(j in 1:(n-2)){
  a<-sum(P[1:(n-j),j]*((P[1:(n-j),j+1]/P[1:(n-j),j])-fP[j])*((P[1:(n-j),j+1]/P[1:(n-j),j])-fP[j]))
  a<-a/(n-j-1)
  phiP<-append(phiP,a)
}
phiP<-sqrt(phiP)
b<-min((phiP[n-2]^4)/(phiP[n-3]^2),phiP[n-2]^2,phiP[n-3]^2)
phiP<-append(phiP,sqrt(b))
print(phiP)
###
phiI<-c()
for(j in 1:(n-2)){
  a<-sum(I[1:(n-j),j]*((I[1:(n-j),j+1]/I[1:(n-j),j])-fI[j])*((I[1:(n-j),j+1]/I[1:(n-j),j])-fI[j]))
  a<-a/(n-j-1)
  phiI<-append(phiI,a)
}
phiI<-sqrt(phiI)
b<-min((phiI[n-2]^4)/(phiI[n-3]^2),phiI[n-2]^2,phiI[n-3]^2)
phiI<-append(phiI,sqrt(b))
print(phiI)
#####
Q<-I/P
print(Q)
fQ<-c()
for(j in 1:(n)){
```

```

a<-sum(I[1:(n-j+1),j])/sum(P[1:(n-j+1),j])
fQ<-append(fQ,a)
}
print(fQ)
###
phiQP<-c()
for(j in 1:(n-1)){
  a<-sum(P[1:(n-j+1),j]*(Q[1:(n-j+1),j]-fQ[j])*(Q[1:(n-j+1),j]-fQ[j]))
  a<-a/(n-j)
  phiQP<-append(phiQP,a)
}
phiQP<-sqrt(phiQP)
print(phiQP)
###
resQ<-Q
print(resQ)
for (i in 1:n){
  for (j in 1:(n-i+1)){
    resQ[i,j]<-((Q[i,j]-fQ[j])*sqrt(P[i,j]))/phiQP[j]
  }
}
print(resQ)
#####
P2<-matrix(nrow=(n-1),ncol=(n-1))
for (i in 1:(n-1)){
  for (j in 1:(n-1)){
    P2[i,j]<-P[i,j+1]/P[i,j]
  }
}
print(P2)
###
resP<-matrix(nrow=(n-1),ncol=(n-1))
for (i in 1:(n-1)){
  for (j in 1:(n-1)){
    resP[i,j]<-((P[i,j+1]/P[i,j])-fP[j])*sqrt(P[i,j])/phiP[j]
  }
}
print(resP)
#####
a<-0
b<-0
for (i in 1:(n-1)){
  for (j in 1:(n-i)){
    a<-a+(resP[i,j]*resQ[i,j])
    b<-b+(resQ[i,j]*resQ[i,j])
  }
}
lambdaP<-a/b
#####
x<-seq(-2,2,0.1)
y<-lambdaP*x
plot(resQ[1:(n-1),1:(n-1)],resP)
lines(x, y, col="red")
#####
phiQI<-c()
for(j in 1:(n-1)){
  a<-sum(I[1:(n-j+1),j]*(1/Q[1:(n-j+1),j]-1/fQ[j])*(1/Q[1:(n-j+1),j]-1/fQ[j]))
  a<-a/(n-j)
  phiQI<-append(phiQI,a)
}
phiQI<-sqrt(phiQI)

```

```

print(phiQI)
#####
I2<-matrix(nrow=(n-1),ncol=(n-1))
for (i in 1:(n-1)){
  for (j in 1:(n-1)){
    I2[i,j]<-I[i,j+1]/I[i,j]
  }
}
print(I2)
###
resI<-matrix(nrow=(n-1),ncol=(n-1))
for (i in 1:(n-1)){
  for (j in 1:(n-1)){
    resI[i,j]<-(((I[i,j+1]/I[i,j])-fI[j])*sqrt(I[i,j]))/phiI[j]
  }
}
print(resI)
###
resQ2<-Q
for (i in 1:n){
  for (j in 1:(n-i+1)){
    resQ2[i,j]<-((1/Q[i,j]-1/fQ[j])*sqrt(I[i,j]))/phiQI[j]
  }
}
print(resQ2)
#####
a2<-0
b2<-0
for (i in 1:(n-1)){
  for (j in 1:(n-i)){
    a2<-a2+(resI[i,j]*resQ2[i,j])
    b2<-b2+(resQ2[i,j]*resQ2[i,j])
  }
}
lambdaI<-a2/b2
print(lambdaI)

#####
UP<-P
UI<-I
DP<-matrix(nrow=(n),ncol=(n-1))
DI<-matrix(nrow=(n),ncol=(n-1))
for (k in (n+1):(2*n-1)){
  for (i in 2:n){
    for (j in (1:(n-1))){
      if ((i+j)==k){
        DP[i,j]<-fP[j]+lambdaP*phiP[j]*((UI[i,j]/UP[i,j])-fQ[j])/phiQP[j]
        UP[i,(j+1)]<-UP[i,j]*DP[i,j]
        DI[i,j]<-fI[j]+lambdaI*phiI[j]*((UP[i,j]/UI[i,j])-1/fQ[j])/phiQI[j]
        UI[i,(j+1)]<-UI[i,j]*DI[i,j]
      }
    }
  }
}
print(DP);print(UP);print(DI);print(UI)

```

A.3 Focus sur le deuxième jeu de données

	0	1	2	3	4	5	6	7	8	9
1996				9 349	9 488	9 422	9 425	9 420	9 422	9 419
1997			15 102	15 240	15 326	15 395	15 464	15 408	15 392	
1998		6 275	6 461	6 487	6 481	6 484	6 516	6 522		
1999	2 796	9 845	11 569	11 909	11 998	12 056	12 160			
2000	47 87	8 613	9 465	9 555	9 387	9 509				
2001	3 216	5 235	6 358	6 403	6 393					
2002	3 460	9 378	11 926	12 063						
2003	3 854	6 588	6 751							
2004	5 055	10 441								
2005	4 774									

TABLE A.1: Tableau des montants payés pour les dégâts corporels

	0	1	2	3	4	5	6	7	8	9
1996				46 476	46 685	46 812	46 832	46 775	47 030	47 074
1997			37 995	38 327	38 494	38 469	38 647	38 682	38 612	
1998		25 795	26 606	27 029	27 165	27 348	27 592	27 607		
1999	15 952	63 450	74 112	76 963	78 184	78 503	78 635			
2000	14 983	25 742	27 784	28 320	28 563	28 734				
2001	19 568	31 914	34 174	34 538	34 732					
2002	20 493	32 418	34 356	35 019						
2003	24 231	39 144	41 823							
2004	24 192	41 186								
2005	24 842									

TABLE A.2: Tableau des montants payés pour les dégâts non corporels

	0	1	2	3	4	5	6	7	8	9
1996				17%	17%	17%	17%	17%	17%	17%
1997			28%	28%	28%	29%	29%	28%	29%	
1998		20%	20%	19%	19%	19%	19%	19%		
1999	15%	13%	14%	13%	13%	13%	13%			
2000	24%	25%	25%	25%	25%	25%				
2001	14%	14%	16%	16%	16%					
2002	14%	22%	26%	26%						
2003	14%	14%	14%							
2004	17%	20%								
2005	16%									

TABLE A.3: Pourcentage des montants corporels sur les montants totaux

	0	1	2	3	4	5	6	7	8
1				1.01	0.99	1.00	1.00	1.00	1.00
2			1.01	1.01	1.00	1.00	1.00	1.00	
3		1.03	1.00	1.00	1.00	1.00	1.00		
4	3.52	1.18	1.03	1.01	1.00	1.01			
5	1.80	1.10	1.01	0.98	1.01				
6	1.63	1.21	1.01	1.00					
7	2.71	1.27	1.01						
8	1.71	1.02							
9	2.07								

TABLE A.4: Facteurs de développement des montants payés pour les dégâts corporels

	0	1	2	3	4	5	6	7	8
1				1.00	1.00	1.00	1.00	1.01	1.00
2			1.01	1.00	1.00	1.00	1.00	1.00	
3		1.03	1.02	1.01	1.01	1.01	1.00		
4	3.98	1.17	1.04	1.02	1.00	1.00			
5	1.72	1.08	1.02	1.01	1.01				
6	1.63	1.07	1.01	1.01					
7	1.58	1.06	1.02						
8	1.62	1.07							
9	1.70								

TABLE A.5: Facteurs de développement des montants payés pour les dégâts non corporels