CEA | CENTRE D'ÉTUDES ACTUARIELLES

Mémoire présenté devant le Centre d'Etudes Actuarielles pour l'obtention du diplôme du Centre d'Etudes Actuarielles et l'admission à l'Institut des Actuaires
le : _____

Par : _____Hélène Gibello et Benoît Lebrun _
Titre : _____**CREDIBILISATION DES METHODES DE PROVISIONNEMENT NON VIE**
_____

Confidentialité : ☒ NON ☐ OUI (Durée : ☐ 1an ☐ 2 ans)
Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus

Membre présent du jury de l'Institut des Actuaires : signature

_____

Membres présents du jury du Centre d'Etudes Actuarielles :

Thomas BEHAR

Vincent DAMAS

Gérard CROSET

Arnaud COHEN

Jean-Pierre DIAZ

Brigitte DUBUS

Paul ESMEIN

Michel FROMENTEAU

Benoît HUGONIN

Christophe IZART

Pierre PETAUTON

Florence PICARD

Christian-Yann ROBERT

Secrétariat :

Bibliothèque :

Entreprise : AXA Liabilities Managers

Nom : _Cédric de Linares_____
Signature :

Directeur de mémoire en entreprise :

Nom : _François-Xavier NEGRI
_____
Signature :

Invité :

Nom : _____
Signature :

Autorisation de publication et de mise en ligne sur un site de diffusion de documents actuariels
(après expiration de l'éventuel délai de confidentialité)

Signature du responsable entreprise

Signature du candidat

4 rue Chauveau Lagarde 75008 Paris Tél. : 01 44 51 72 72 / Fax : 01 44 51 72 73

REMERCIEMENTS

Nous tenons à remercier François Xavier NEGRI, Chief Risk Officer zone Europe Centrale et de l'Est chez AXA GIE, pour l'aide qu'il nous a apportée tout au long de nos travaux. Son expertise et ses conseils judicieux ont largement contribué à enrichir et approfondir le contenu de ce mémoire.

Nous souhaitons également remercier M. Christian ROBERT pour la gentillesse et la compréhension dont il a fait preuve lors de nos différents échanges.

- SOMMAIRE -

REMERCIEMENTS **2**

INTRODUCTION **6**

PARTIE 1 ..... **8**

LES METHODES CLASSIQUES DE PROVISIONNEMENT **8**

1 NOTATIONS USUELLES **8**

2 PRESENTATION DES DONNEES **10**

2.1 TRIANGLES D'ASSURANCE DIRECTE 10
2.2 TRIANGLES DE REASSURANCE 13

3 PRESENTATION DES METHODES **16**

3.1 LA METHODE DE CHAIN LADDER 18
3.2 LES METHODES DE COUT MOYEN 22
3.2.1 COUT MOYEN avec triangulation *22*
3.2.2 COUT MOYEN sans triangulation *22*
3.3 LES METHODES D'AJUSTEMENT DE COURBES 23
3.3.1 La méthode de Sherman *23*
3.3.2 La méthode de Craighead *24*
3.4 LA METHODE DU LOSS RATIO 26
3.5 LA METHODE D'HERTIG 27
3.6 LA METHODE DE BORNHUETTER-FERGUSON 28
3.7 LA METHODE DE CAPE COD 31
3.8 LA METHODE DE BENKTANDER 34
3.9 LA MÉTHODE GLM (GENERALIZED LINEAR MODEL) 36

4 RESUME DES AVANTAGES ET INCONVENIENTS DES METHODES PRESENTEES **38**

PARTIE 2 ..... **40**

LES PREMIERES APPROCHES DE CREDIBILITE LINEAIRE EN PROVISIONNEMENT **40**

1 PRESENTATION DE LA CREDIBILITE LINEAIRE **40**

1.1 LES ORIGINES DE LA THEORIE DE LA CREDIBILITE 40
1.2 PRINCIPE GENERAL DE LA THEORIE DE LA CREDIBILITE 41
1.3 REGRESSION LINEAIRE ET THEORIE DE LA CREDIBILITE 43
1.3.1 Notations du modèle de BÜHLMANN et interprétations *43*
1.3.2 Rappel du modèle de régression linéaire *44*
1.3.3 Application du modèle de régression aux hypothèses de BÜHLMANN *45*

2 LA CREDIBILITE LINEAIRE DANS DES METHODES DITES CLASSIQUES : BORNHUETTER-FERGUSON / CAPE COD ET BENKTANDER **48**

2.1 LA CREDIBILITE DANS LES METHODES BORNHUETTER-FERGUSON / CAPE COD 48
2.2 LA CREDIBILITE DANS LA METHODE BENKTANDER 53

3 UNE AUTRE FORME DE CREDIBILISATION DE L'INFORMATION DISPONIBLE : LE MUNICH CHAIN LADDER **57**

3.1 NOTATIONS 57
3.2 PRESENTATION SYNTHETIQUE DU PRINCIPE DE LA METHODE 58
3.3 ETUDE PRELIMINAIRE : EVOLUTION DES SINISTRES SURVENUS, PAYES ET EN SUSPENS AINSI QUE LEURS INFLUENCES SUR LES RATIOS $R_{i,j}^{MCL}$ ET $\left(R_{i,j}^{MCL}\right)^{-1}$ 59
3.4 LA CORRELATION ENTRE LA DIAGONALE ET LA SUR DIAGONALE D'UN TRIANGLE 63
3.5 LA METHODE DU MUNICH CHAIN LADDER 66

3.6 LA METHODE DU MUNICH CHAIN LADDER UNE CREDIBILISATION LINEAIRE 68

4 CONCLUSIONS ET POSTULATS DE CETTE PARTIE 69

PARTIE 3 70

ETUDE D'UNE NOUVELLE METHODE DE PROVISIONNEMENT UTILISANT LA THEORIE DE LA CREDIBILITE LINEAIRE VIA UNE INFORMATION EXOGENE 70

1 PRESENTATION DES PRINCIPES D'UNE NOUVELLE APPROCHE ET FORMULATIONS .. 70

1.1 LES BASES DE REFLEXION 70
1.2 FORMULATION DE L'APPROCHE PROPOSEE 72
1.2.1 Définition des ratios sélectionnés et justification économique 72
1.2.2 Formulation générale 72
1.2.3 Formulation des différents paramètres 75
1.2.4 Les résidus 76
1.2.5 Les Corrélations 76

2 ANALYSE ACTUARIELLE DE L'APPROCHE PROPOSEE 78

2.1 DONNEES UTILISEES 78
2.1.1 Les triangles de réassurance 78
2.1.2 Les triangles d'assurance 79
2.2 ETUDE DES CORRELATIONS 80
2.2.1 Analyse statique des coefficients de corrélations 80
2.2.2 Analyse graphique des corrélations 82

3 APPLICATION ET QUALITE DE L'ESTIMATION 85

3.1 ANALYSE GRAPHIQUE DES PROJECTIONS A L'ULTIME 85
3.2 ANALYSE EN « BACKTESTING » 93
3.3 ADAPTATION DE LA METHODE 95

CONCLUSION 106

ANNEXES PARTIE 1 108

1 PRESENTATION DES DONNEES 108

1.1 ILLUSTRATION DES TRIANGLES 108

2 PRESENTATION DES METHODES 109

2.1 LES DIFFERENTS ESTIMATEURS DE CHAIN LADDER 109
2.2 LES LOSS RATIO ULTIMES 110
2.2.1 Les loss ratios ultimes du triangle d'assurance estimés en N 110
2.2.2 Les loss ratios ultimes du triangle d'assurance estimés en N-1 111
2.2.3 Les loss ratios ultimes du triangle de réassurance estimés en N 112
2.2.4 Les loss ratios ultimes du triangle de réassurance estimés en N-1 113
2.3 TEST DE STABILITE 114
2.3.1 Test de stabilité des méthodes sur le triangle d'assurance 114
2.3.2 Test de stabilité des méthodes sur le triangle de réassurance 115
2.4 LA METHODE DE SHERMAN : LES PARAMETRES DE LISSAGE 116
2.5 LA METHODE DE CRAIGHEAD : LES PARAMETRES DE LISSAGE 116
2.6 LA METHODE DU LOSS RATIO : LES S/P RETENUS 117
2.6.1 S/P retenus sur le triangle d'assurance 117
2.6.2 S/P retenus sur le triangle de réassurance 117
2.7 LA METHODE GLM : LES COEFFICIENTS ESTIMES 118
2.7.1 Coefficients sur le triangle d'assurance 118
2.7.2 Coefficients sur le triangle de réassurance 118
2.8 COMPARAISON DES LOSS RATIOS ULTIMES EN FONCTION DE LA METHODE DE CADENCEMENT UTILISEE 119

ANNEXES PARTIE 2 120

3 LA CREDIBILITE DANS LES METHODES DITES CLASSIQUES 120

3.1 DEMONSTRATION DE RESULTATS INTERMEDIAIRES DANS LE BORNHUETTER-FERGUSON 120

3.2 DEMONSTRATION DE RESULTATS INTERMEDIAIRES DANS LE BENKTANDER 120
4 LE MUNICH CHAIN LADDER 122
4.1 REPRESENTATION GRAPHIQUE DE L'EVOLUTION DES $R_{i,j}^{MCL}$ SUR LE TRIANGLE D'ASSURANCE 122
4.2 DEMONSTRATION DE M MERZ ET M V WÜTHRICH SUR LA CREDIBILITE DU MCL 123
ANNEXES PARTIE 3 127
1 GRAPHIQUES DE CORRELATION DES AUTRES TRIANGLES ET PROJECTIONS ULTIMES 127
2 QUALITE DES ESTIMATIONS DES DIFFERENTES METHODES SUR LES AUTRES TRIANGLES 132
BIBLIOGRAPHIE 134

INTRODUCTION

L'évaluation des provisions pour sinistres à payer (PSAP) est un sujet qui au fil des ans fait l'objet de nombreux travaux actuariels, que ce soit au travers d'ouvrages, d'articles ou de mémoires. Cet intérêt s'explique en majeure partie par le poids essentiel que représentent les provisions de sinistres dans le bilan d'un assureur. De plus, la réforme en cours du cadre prudentiel va conduire les assureurs européens à devoir estimer leur valeur actuelle probable, ce qui contribue à en faire un sujet d'actualité.

Selon les derniers chiffres disponibles de la Fédération Française des Sociétés d'Assurances (FFSA), les provisions de sinistres représentaient en 2008 la moitié du bilan du marché français de l'assurance de dommages aux biens et responsabilité et plus de deux fois ses capitaux propres. L'importance du volume de ce poste comptable justifie à lui seul tous les efforts menés pour améliorer la qualité de l'évaluation de ces provisions. En effet, schématiquement, une erreur de 5% dans l'estimation de ces provisions aurait toutes choses égales par ailleurs et hors prise en compte de la fiscalité, une incidence supérieure à 10% sur les fonds propres. Les professionnels, spécialistes des travaux d'inventaire, savent que pour certains types de risques, ce niveau d'erreur peut facilement être dépassé.

L'assurance et désormais la réassurance sont des activités réglementées. Les autorités en charge de leur contrôle ont pour mission de protéger les intérêts des assurés. Les compagnies sont alors tenues d'évaluer leurs provisions techniques selon le principe de prudence. Le but étant d'éviter tout dérapage de la solvabilité lié à une sous estimation des engagements de la société.

La mise en place du nouveau régime « Solvabilité 2 », impose désormais d'évaluer les provisions techniques en valeur de marché, autrement dit à leur valeur actuelle probable. Le principe de prudence ne disparaît pas mais devra être explicitement mesuré. Les « excès » de provisions pourront alors être, sous certaines contraintes, éligibles en fonds propres prudentiels. Au vu des montants sous-jacents, l'enjeu pour un (ré-) assureur est alors d'estimer au plus juste ses engagements. Il doit être en mesure de justifier au régulateur de la justesse de leurs prévisions. C'est pourquoi il est nécessaire d'avoir recours à des méthodes robustes répondant à des standards partagés par le plus grand nombre.

Le présent mémoire s'inscrit dans ce contexte et cherche, par une démarche exploratoire et à partir des méthodes de provisionnement les plus répandues, à examiner la possibilité de combiner au mieux celles-ci, sous une forme linéaire ; l'objectif visé est de tenter d'obtenir une estimation de meilleure qualité. Nos travaux ont pour but de répondre à la question suivante : peut-on construire un méta-indicateur meilleur prédicteur en moyenne ou ayant une variance moindre que chacun des estimateurs premiers ?

L'idée de départ se fonde sur le constat que les méthodes traditionnelles, dites « classiques », essentiellement basées sur des techniques de triangulation, n'utilisent pas toutes les informations à disposition. Ce travail est réalisé dans le cadre conceptuel de la théorie de la crédibilité linéaire. Cette théorie est largement répandue dans les problématiques de tarification et est à l'origine de l'instauration du système bonus malus en assurance

automobile. Celle-ci a pour principe d'associer par combinaison linéaire des informations basées sur l'expérience spécifique et des informations d'origine plus générale.

Le mémoire se décompose ainsi en trois parties :

En première partie, les méthodes de provisionnement dites classiques sont exposées. Des indicateurs sont créés afin de pouvoir comparer chacune des méthodes entre elles. Cela permet de constater leurs qualités et leurs défauts.

En seconde partie, la théorie de la crédibilité est succinctement présentée. Puis, il est montré que certaines méthodes vues en Partie I sont des crédibilisations de plusieurs méthodes classiques, fonction des qualités de chacune d'elles. Enfin, la méthode du Munich Chain Ladder utilisant la crédibilité est exposée. Celle-ci ne crédibilise pas plusieurs méthodes, mais les résultats issus d'une même méthode appliqués sur différentes variables.

Enfin, la troisième partie présente une méthode alternative, basée sur l'intégration simultanée de toute l'information contenue dans les triangles de sinistres, les sinistres payés et les sinistres survenus, mais aussi les nombres de sinistres. Cette méthode est basée sur les principes de crédibilité. Elle reprend la théorie de la méthode du Munich Chain Ladder, tout en utilisant d'autres paramètres. Nous analyserons, à l'aide d'exemples concrets, les apports et défauts de cette approche et expliquerons en quoi l'utilisation d'informations supplémentaires peut dans certains cas améliorer la qualité des estimations de charge ultime.

PARTIE 1

LES METHODES CLASSIQUES DE PROVISIONNEMENT

1 Notations usuelles

Soient :

N	correspond à l'année de développement où se clôt un exercice
i	représente l'année d'origine. Il peut s'agir de l'année de souscription des contrats ou de l'année de survenance des sinistres. $i = \{1,...,I\}$ avec I l'année d'origine la plus récente connue
j	représente le délai de développement de la mesure étudiée (paiements, charges, primes). $j = \{0,1,...,N\}$. Un exercice (ou une année d'origine) est clos(e) au bout de N+1 années
$N - i + 1$	correspond à l'année courante ou l'année d'inventaire
$i + N + 1$	correspond à l'année de « clôture » de l'année d'origine i
$X_{i,j}$	représente la mesure étudiée rattachée à l'année d'origine i vue au bout de $j+1$ années
$C_{i,j} = \sum_{j=0}^{N-i+1} X_{i,j}$	représente le montant cumulé de la mesure étudiée rattaché à l'année d'origine i et vu au bout de $j+1$ années
R_i	représente les provisions rattachées à l'année d'origine i
$R = \sum_{i=1}^{I} R_i$	représente le total des provisions toutes années d'origine confondues
$f_{i,j}$	est le facteur de développement, pour une année d'origine i, du montant cumulé de la mesure étudiée de l'année $i+j$ à l'année $i+j+1$

f_j est le facteur de développement, quelle que soit l'année d'origine i, du montant cumulé de la mesure étudiée de l'année $i+j$ à l'année $i+j+1$

$F_{i,j}$ est le facteur de développement, pour une année d'origine i, du montant cumulé de la mesure étudiée de l'année $i+j$ à l'année $i+N+1$, soit jusqu'à la clôture de l'exercice

$$F_{i,j} = \prod_{k=0}^{j-1} f_{i,k}$$

F_j est le facteur de développement, quelle que soit l'année d'origine i, du montant cumulé de la mesure étudiée de l'année $i+j$ à l'année $i+N+1$, soit jusqu'à la clôture de l'exercice.

$$F_j = \prod_{k=0}^{j-1} f_k$$

A_i est l'aliment ou l'exposition pour l'année d'origine i (primes émises, primes acquises ou primes ultimes)

$Q_{i,j}$ est le ratio émis de sinistres sur primes (S/P) de l'année de d'origine i en fin d'année de développement j

Deux triangles sont fournis en ANNEXES PARTIE 1 [1.1] afin d'illustrer les notations précédentes.
Par la suite, nous utiliserons les $C_{i,j}$ et $X_{i,j}$ comme charges de sinistres ou paiements.

2 Présentation des données

Les méthodes de provisionnement présentées dans la première partie de ce mémoire sont à chaque fois illustrées par des applications numériques. Les triangles sélectionnés sont de nature et de structure volontairement différentes. L'un est issu des comptes d'un assureur direct sur une branche Multirisques Habitations à la fois courte et homogène. L'autre est extrait des comptes d'un réassureur sur la branche Automobile proportionnelle, volatile et à déroulé de paiements assez long (une quinzaine d'années environ). Ce choix de branches vise à tester l'efficacité des méthodes selon le type de branche étudiée.

2.1 Triangles d'assurance directe

Le portefeuille d'assurance est tiré d'un acteur français présent sur l'ensemble du territoire depuis plusieurs décennies avec une part de marché relativement stable sur les années étudiées.

Les données initiales sont en milliers d'euros courants. Deux traitements sur ces données ont été effectués :

- Le premier, lié à la confidentialité des données, nous oblige à transformer les données en indice. La base 1000 de référence correspond aux paiements liés aux sinistres survenus en 2004 et payés la même année.

On a donc : $I_{2004,0} = 1000$ l'indice de référence en base 1000

$$I_{2003,0} = \frac{X_{2003,0}}{X_{2004,0}} \times I_{2004,0}$$

Et par exemple

- Le second vise à raisonner sur des euros constants. Si le triangle en euros constants de 2004 ne s'applique pas pour toutes les méthodes (coût moyen entre autres), il peut contribuer à satisfaire en partie l'indépendance entre les années d'origine, hypothèse nécessaire dans les méthodes présentées par la suite (ex : Chain Ladder, Bornhuetter Ferguson,…)

L'indice FFB (indice de coût à la construction) a été utilisé, celui-ci servant à l'indexation des polices d'assurance multirisques habitations. Par hypothèse, les paiements sont réalisés en milieu d'année. Ainsi, nous prenons la moyenne de l'indice FFB sur l'année considérée en tant que déflateur de ces montants. Nous ramenons les données en euros constants de l'année 2004 de la manière suivante :

$$I^{2004}_{2003,0} = I_{2003,0} \times \left(1 + \frac{\overline{FFB_{2004}}}{\overline{FFB_{2003}}}\right) \quad \text{ou encore} \quad I^{2004}_{2001,2} = I_{2001,2} \times \left(1 + \frac{\overline{FFB_{2004}}}{\overline{FFB_{2003}}}\right)$$

Avec $\overline{FFB_{2004}}$ la moyenne de l'indice FFB sur l'année 2004

$I_{2001,2}^{2004}$ l'indice des paiements effectués en 2003 au titre de sinistres survenus en 2001 revalorisés en euros de 2004.

Par ailleurs, les années d'origine sont regroupées par année de survenance. L'historique étudié est de 9 ans.

Enfin, ce triangle est composé de sinistres survenus entre 1995 et 2004. Ainsi, les tempêtes de l'année 1999 (Lothar et Martin) apparaissent clairement. Pour ne pas « polluer » l'analyse, l'année 1999 est retirée du triangle pour l'application des méthodes.

- Vecteur de primes :

années d'origine	primes ultimes
1995	1 900
1996	2 035
1997	2 142
1998	2 178
2000	2 229
2001	2 366
2002	2 531
2003	2 749
2004	2 965

Tab. 1 : Vecteur des Primes d'assurance

- Trapèzes des sinistres survenus / Payés / Sinistres en suspens / Nombre de sinistres :

Les sinistres survenus sont égaux à la somme des sinistres payés et des sinistres en suspens.

	i\j	0	1	2	3	4
1995	1	1 763	1 664	1 591	1 575	1 547
1996	2	1 717	1 581	1 530	1 524	1 530
1997	3	1 912	1 791	1 734	1 734	1 728
1998	4	1 889	1 689	1 658	1 645	1 622
2000	5	1 789	1 714	1 572	1 548	1 542
2001	6	2 350	2 417	2 291	2 155	
2002	7	2 159	1 997	1 798		
2003	8	2 601	2 259			
2004	9	2 147				

Tab. 2 : Trapèze des sinistres survenus d'assurance

	i\j	0	1	2	3	4
1995	1	955	240	92	47	0
1996	2	806	110	19	8	0
1997	3	933	188	33	19	0
1998	4	899	181	83	37	0
2000	5	1 045	283	63	15	0
2001	6	1 226	374	162	- 17	
2002	7	1 207	365	122		
2003	8	1 419	318			
2004	9	1 147				

Tab. 3 : Trapèze des sinistres en suspens d'assurance

	i\j	0	1	2	3	4
1995	1	808	1 424	1 499	1 527	1 547
1996	2	910	1 470	1 512	1 516	1 530
1997	3	979	1 603	1 701	1 715	1 728
1998	4	991	1 508	1 575	1 608	1 622
2000	5	744	1 431	1 509	1 533	1 542
2001	6	1 125	2 043	2 128	2 172	
2002	7	952	1 633	1 675		
2003	8	1 182	1 941			
2004	9	1 000				

Tab. 4 : Trapèze des sinistres payés d'assurance

	i\j	0	1	2	3	4
1995	1	1 165	1 168	1 171	1 168	1 169
1996	2	1 253	1 251	1 248	1 247	1 248
1997	3	1 397	1 395	1 397	1 395	1 396
1998	4	1 349	1 317	1 317	1 315	1 315
2000	5	1 397	1 385	1 376	1 373	1 373
2001	6	1 405	1 421	1 412	1 412	
2002	7	1 386	1 428	1 430		
2003	8	1 373	1 447			
2004	9	1 365				

Tab. 5 : Trapèze des nombres de sinistres d'assurance

Au bout de trois ans, le nombre de sinistres ne varie quasiment plus. Les montants de sinistres survenus deviennent stables. Le développement sur ce type de branche est rapide.

2.2 Triangles de réassurance

La branche étudiée en réassurance est de l'automobile en quote-part.

Les années d'origine sont regroupées par année de souscription et non de survenance. En effet, les données par survenance ne sont pas nécessairement fournies par les cédantes. La plupart du temps, en particulier pour des traités en quote-part, la cédante déclare ses informations par exercices de souscription. La transformation des triangles d'exercices de souscription, en exercice de survenance dans certain cas n'est pas possible.

Généralement, aux années d'origine correspondent les politiques de souscription de la période : cycles de marché, critère de zone géographique, type de contrats, préférence pour telle ou telle cédante. L'étude en exercice de souscription donne donc une vision plus économique des branches en réassurance.

L'historique étudié est de 14 ans. Le développement sur ce type de branche est relativement long et contraste ainsi avec le triangle d'assurance précédemment décrit.

Le triangle du nombre de sinistres n'est pas disponible car en réassurance proportionnelle, la cédante ne fournit pas de détails sur le nombre de sinistres déclarés.

- Vecteur de primes :

UWY	Primes Ultimes
1992	559
1993	16 396
1994	16 317
1995	19 342
1996	58 541
1997	39 467
1998	29 365
1999	41 046
2000	145 860
2001	246 044
2002	121 922
2003	12 181
2004	58 844
2005	62 000

Tab. 6 : Vecteur des Primes de réassurance

Les années 1999 à 2001 se caractérisent par une bourse haute et des taux de Rate on Line[1] bas. Il y avait de gros afflux de capitaux sur le marché. Le marché était « soft », c'est-à-dire que les cédantes fixaient leurs prix. La société de réassurance a cherché à souscrire de gros volumes, la branche Automobile US était une branche intéressante car elle avait une faible volatilité comparé à l'Europe (nombreuses commissions à échelle). A partir de 2000, 90% des contrats sont souscrits sur les Etats-Unis.

Sur les années antérieures à 1997, les parts souscrites étaient assez faibles (inférieures à 10%) puis les parts ont sensiblement augmenté avec une moyenne de 35%.

- Triangle de sinistres survenus / Payés / Sinistres en suspens :

UWY	i\j	0	1	2	3	4	5	6	7	8	9	10	11	12	13
1992	1	280	607	519	544	544	604	615	601	599	630	624	624	624	624
1993	2	5 727	9 469	12 275	12 763	12 854	12 881	12 845	12 762	13 074	13 113	13 279	13 369	13 447	
1994	3	5 864	11 345	12 332	12 348	12 598	12 630	12 532	12 685	12 807	12 858	12 949	13 075		
1995	4	8 748	16 443	17 052	17 471	17 453	17 365	17 543	17 716	17 794	17 873	18 023			
1996	5	15 473	44 950	48 632	46 781	45 604	45 334	45 036	45 007	45 141	44 991				
1997	6	5 921	33 524	31 451	32 745	33 101	33 325	33 974	34 444	34 283					
1998	7	3 131	13 866	17 452	20 284	22 544	24 666	25 586	25 910						
1999	8	1 524	15 300	26 396	32 304	37 043	38 999	39 786							
2000	9	16 849	84 929	101 883	113 254	115 967	118 342								
2001	10	36 377	124 706	157 860	171 023	173 343									
2002	11	32 016	63 134	67 082	67 915										
2003	12	1 464	3 840	5 743											
2004	13	12 713	30 684												
2005	14	695													

Tab. 7 : Triangle des sinistres survenus de réassurance

UWY	i\j	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1992	1	150	145	0	0	0	41	49	15	8	8	0	0	0	0
1993	2	1 547	1 209	1 299	783	751	664	477	480	601	429	512	582	640	
1994	3	1 539	1 122	937	203	370	315	349	468	553	568	633	737		
1995	4	3 851	1 990	938	509	331	405	535	651	678	794	922			
1996	5	3 055	7 443	6 808	4 143	2 596	2 102	1 623	1 460	1 432	1 256				
1997	6	105	12 186	6 307	5 447	3 563	2 076	1 613	1 286	938					
1998	7	1 570	4 358	5 465	3 384	2 387	1 965	1 320	1 394						
1999	8	735	11 135	8 241	5 615	5 036	4 042	4 132							
2000	9	7 715	23 184	11 522	9 376	5 291	4 184								
2001	10	18 176	25 772	15 696	8 329	6 784									
2002	11	13 429	9 008	5 220	3 593										
2003	12	488	737	945											
2004	13	4 769	7 444												
2005	14	419													

Tab. 8 : Triangle des sinistres en suspens de réassurance

[1] Dans les cas de traités non proportionnels, rapports existants entre la prime et la portée de la couverture (c'est l'inverse du Pay-back). Ex : pour un traité XS dont la prime est de 0.2M et l'engagement de 5M, le Rate on line est de 0.2M/5M soit 0.04 ou 4%. Source : Glossaire d'Effisoft

UWY	i\j	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1992	1	130	462	519	544	544	563	566	586	591	622	624	624	624	624
1993	2	4 180	8 260	10 976	11 980	12 103	12 216	12 367	12 282	12 473	12 684	12 767	12 788	12 806	
1994	3	4 325	10 223	11 395	12 145	12 229	12 315	12 183	12 217	12 254	12 290	12 317	12 338		
1995	4	4 898	14 454	16 113	16 962	17 122	16 960	17 008	17 065	17 117	17 079	17 100			
1996	5	12 418	37 507	41 824	42 638	43 008	43 233	43 413	43 547	43 709	43 735				
1997	6	5 816	21 338	25 144	27 298	29 538	31 249	32 361	33 159	33 346					
1998	7	1 561	9 508	11 987	16 900	20 157	22 701	24 266	24 516						
1999	8	789	4 165	18 156	26 689	32 007	34 957	35 654							
2000	9	9 134	61 745	90 361	103 878	110 677	114 158								
2001	10	18 201	98 935	142 164	162 694	166 559									
2002	11	18 588	54 126	61 861	64 322										
2003	12	976	3 104	4 798											
2004	13	7 945	23 240												
2005	14	276													

Tab. 9 : Triangle des sinistres payés de réassurance

Ces trois triangles illustrent le fait que les triangles de réassurance peuvent être beaucoup plus volatils que ceux en assurance.

3 Présentation des méthodes

Cette première partie décrit les méthodes de provisionnement dites classiques ou parmi les plus connues, et par conséquent certainement les plus utilisées. Elles sont ensuite appliquées aux deux types de triangles afin d'observer si les charges de sinistres ultimes estimées sont similaires quelle que soit la méthode.

Pour chaque méthode, les charges de sinistres ultimes estimées sont indiquées dans l'ANNEXE PARTIE 1 [2.2] dans un tableau récapitulatif sous forme de ratio sinistres à primes (« loss ratio »).

Un test de « stabilité » ou de « sensibilité » a été mis en place afin d'appréhender systématiquement la variabilité de chacune des méthodes au niveau d'information disponible. Les résultats se trouvent en ANNEXES PARTIE 1 [2.3].
Celui-ci se déroule en deux temps : il s'agit, dans un premier temps, d'estimer les charges de sinistres ultimes à partir du triangle existant puis, dans un second temps, de refaire ce même exercice en se plaçant une année auparavant en retirant la dernière diagonale du triangle.
La sensibilité est alors évaluée par l'écart entre les deux estimations obtenues pour chacune des années d'origine disponibles et exprimées en pourcentage de la prime.

Soient :
U_i^t le montant estimé de la charge ultime de sinistres pour l'année d'origine i calculé à la date actuelle t.
U_i^{t-1} le montant estimé de la charge ultime de sinistres pour l'année d'origine i calculé une année auparavant en t-1.

Le test se définit alors selon la formule suivante :

$$\Delta_{i,t} = \left(\frac{U_i^t - U_i^{t-1}}{P_i} \right)$$

Ce critère de stabilité n'est pertinent que si, par ailleurs, l'environnement de souscription et de gestion des sinistres est aussi resté « stable ». En réassurance ce n'est pas souvent le cas du fait des cycles de marché, mais généralement les équipes de souscriptions et les tranches souscrites restent similaires. Nous admettons ce postulat afin d'avoir des critères comparatifs.

La plupart des méthodes utilisent un mécanisme de développement ou une vision ex ante du résultat à l'ultime (ratio sinistre / prime).

Pour les méthodes s'appuyant sur des hypothèses de développement, d'autres indicateurs supplémentaires ont été mis en place afin d'observer la sensibilité de la méthode à l'hypothèse de développement choisi :

- « M » : moyenne des loss ratios ultimes calculés suivant les différentes méthodes de développement.

- « E » : écart type des loss ratios ultimes calculés suivant les différentes méthodes de développement. Il représente l'écart d'estimation de sinistre par rapport à M en fonction des méthodes utilisées.

3.1 La méthode de CHAIN LADDER

La méthode de Chain Ladder est certainement la plus connue et la plus employée des méthodes de provisionnement utilisant le développement d'un triangle. Son succès tient sans doute en partie à sa simplicité de mise en pratique. Elle peut s'appliquer sur différents types de données (paiements cumulés, charge de sinistres ou encore le nombre de sinistres déclarés ou survenus). Dans la suite du mémoire, la méthode de Chain Ladder classique présentée ici, sera appelée SCL, pour Standard Chain Ladder.

Le principe de cette méthode est d'utiliser des facteurs de proportionnalité ou de passage entre le niveau de sinistralité vu au bout d'un certain délai et ce même niveau vu un an auparavant. Ces facteurs sont ensuite appliqués aux années d'origine les plus récentes afin d'estimer leur charge de sinistres future.

Par hypothèse, le rapport de proportionnalité est identique quelle que soit l'année d'origine.

$$\frac{C_{1,j}}{C_{1,j-1}} = \frac{C_{2,j}}{C_{2,j-1}} = \mathbf{L} = \frac{C_{i,j}}{C_{i,j-1}} = \mathbf{L} = \frac{C_{N-j,j}}{C_{N-j,j-1}} \qquad \text{(H1)}$$

Celui-ci plus communément appelé facteur de développement ne dépend ainsi que du délai de développement du triangle, soit du nombre d'années écoulées depuis l'année d'origine.

Nous utilisons la formulation très générale de Mack du facteur de développement (Bibliographie [8]) :

$$\hat{f}_j = \frac{\sum_{i=1}^{I-j-1} w_{i,j}.C_{i,j}^{a}.f_{i,j}}{\sum_{i=1}^{I-j-1} w_{i,j}.C_{i,j}^{a}} \qquad \text{avec } a \in \{0,1,2\} \qquad (1)$$

Dans la pratique, les $w_{i,j}$ peuvent correspondre à des pondérations attribuées arbitrairement par l'actuaire aux facteurs de développement individuels $f_{i,j}$ des différentes années d'origine. Ces ajustements opérationnels peuvent se justifier pour tenir compte d'informations exogènes, Une modification des règles d'indemnisation par exemple peut avoir un impact sur la cadence des paiements futurs. L'observation de pics, de points « aberrants » ou d'irrégularités dans les données remontant des systèmes d'informations peut également amener à les exclure du calcul des facteurs de développement.

Dans la suite de ce mémoire, $w_{i,j}$ sera toujours égal à 1. La signification des estimateurs en fonction de leur a est en ANNEXES PARTIE 1 [2.1].

Mack (1993 et 1999 cf. Bibliographie [8] et [9]) a développé une version stochastique de la méthode Chain Ladder. Dans ce mémoire, nous nous limitons aux méthodes déterministes mais indiquons toutefois les hypothèses du modèle :

(1) il existe un facteur de développement f_j indépendant des années d'origine tel que :

$$E\left(f_{i,j}\middle|C_{i,1},..C_{i,j}\right)= f_j \qquad \text{(H2)}$$

(2) il y a indépendance entre les différentes années d'origine

$$\left(C_{i,1},C_{i,2},..C_{i,n}\right)\perp\left(C_{k,1},C_{k,2},..C_{k,n}\right) \qquad \forall i \neq k \qquad \text{(H3)}$$

(3) il existe un paramètre s_j^2 indépendant des années d'origine tel que :

$$V\left(f_{i,j}\middle|C_{i,1},..C_{i,j}\right)= \frac{s_j^2}{w_{i,j}.C_{i,j}^a} \qquad \text{(H4)}$$

En d'autres termes, indépendamment des années d'origine, il existe des paramètres de position f_j et de dispersion s_j^2 proportionnels à la mesure cumulée de l'année précédente.

L'illustration des avantages et des inconvénients de cette méthode découle de son application numérique sur les deux triangles précédemment décrits.

Celle-ci a été faite en deux temps : les résultats obtenus suivant les différents a de Chain Ladder ont été comparés entre eux et par triangle. Puis la sensibilité de la méthode a été éprouvée à l'aide du test de stabilité.

Pour les différentes applications numériques issues des méthodes SCL les triangles ont été conservés tels quels, sans effectuer de retraitement manuel (année ou développement aberrant[2]), afin de montrer les avantages et les écueils de cette méthode.
Sur le triangle d'assurance, l'étude a été faite sur les sinistres payés (***Tab 4)***. Le triangle de réassurance étudié est celui des sinistres survenus ***(Tab 7).***

Nous considérons que les 3 hypothèses de Mack citées précédemment sont vérifiées.

[2] Mise à part l'année 1999 du triangle d'assurance.

Les coefficients de développement issus de la méthode SCL sont obtenus à partir du triangle d'assurance *Tab 4* et de réassurance *Tab 7*, pour respectivement les cas α=0, 1, 2 :

	f_0	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8
α =0	1,704	1,044	1,015	1,009	1,000	1,000	1,000	1,000	1,000
α =1	1,697	1,044	1,015	1,009	1,000	1,000	1,000	1,000	1,000
α =2	1,692	1,043	1,015	1,009	1,000	1,000	1,000	1,000	1,000
Moyenne	1,698	1,044	1,015	1,009	1,000	1,000	1,000	1,000	1,000
Variabilité	0,006	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000

Coefficients de développement du triangle d'assurance

	f_0	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}	f_{13}
α =0	3,550	1,192	1,065	1,031	1,031	1,011	1,003	1,006	1,012	1,005	1,006	1,003	1,000	1,000
α =1	3,099	1,181	1,070	1,025	1,022	1,011	1,007	1,004	1,001	1,009	1,008	1,006	1,000	1,000
α =2	2,979	1,201	1,077	1,018	1,020	1,011	1,006	1,002	0,998	1,009	1,008	1,006	1,000	1,000
Moyenne	3,209	1,191	1,071	1,025	1,024	1,011	1,005	1,004	1,004	1,008	1,007	1,005	1,000	1,000
Variabilité	0,301	0,010	0,006	0,006	0,006	0,000	0,002	0,002	0,007	0,003	0,002	0,002	0,000	0,000

Coefficients de développement du triangle de réassurance

Le premier constat est que sur un triangle à développement rapide et faiblement volatil, quel que soit l'estimateur de Chain Ladder utilisé, les coefficients de développement sont identiques hormis celui de la première année. A contrario, sur le triangle de réassurance, les coefficients sont différents jusqu'à la douzième année.

Les deux tableaux suivants fournissent les Loss Ratio Ultimes obtenus avec la méthode Chain Ladder sur les triangles respectivement d'assurance et de réassurance estimés avec un triangle vu en N et un autre vu en N-1 :

loss ratios ultimes		**1995**	**1996**	**1997**	**1998**	**2000**	**2001**	**2002**	**2003**	**2004**
Estimés en N	α =0	81,5%	75,2%	80,7%	74,5%	69,2%	92,6%	67,7%	75,5%	61,4%
	α =1	81,5%	75,2%	80,7%	74,5%	69,2%	92,6%	67,8%	75,5%	61,2%
	α =2	81,5%	75,2%	80,7%	74,5%	69,2%	92,6%	67,8%	75,5%	61,0%
Estimés en N-1	α =0	81,5%	75,2%	80,7%	74,5%	69,5%	92,0%	69,1%	78,9%	
	α =1	81,5%	75,2%	80,7%	74,5%	69,5%	92,0%	69,1%	78,6%	
	α =2	81,5%	75,2%	80,7%	74,5%	69,4%	92,0%	69,0%	78,4%	

Loss Ratio ultimes du triangle d'assurance

loss ratios ultimes		**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**
Estimés en N	α =0	111,5%	82,0%	80,4%	94,0%	77,9%	89,0%	91,0%	100,2%	84,8%	75,9%	61,9%	55,7%	73,5%	5,6%
	α =1	111,5%	82,0%	80,6%	94,4%	78,6%	88,9%	90,6%	100,3%	84,9%	75,3%	61,0%	55,3%	72,2%	4,8%
	α =2	111,5%	82,0%	80,6%	94,5%	78,6%	88,8%	90,3%	99,8%	84,4%	74,8%	60,2%	54,9%	73,0%	4,7%
Estimés en N-1	α =0	111,5%	81,5%	79,6%	93,0%	78,8%	89,9%	89,9%	99,0%	85,5%	77,2%	65,3%	43,6%	108,9%	
	α =1	111,5%	81,5%	79,9%	93,9%	78,7%	89,7%	90,1%	99,1%	84,8%	76,5%	65,3%	44,1%	95,7%	
	α =2	111,5%	81,5%	79,9%	94,0%	78,7%	89,5%	89,8%	98,7%	84,2%	75,5%	64,8%	44,6%	92,0%	

Loss Ratio ultimes du triangle de réassurance

Ce qui est observé sur les coefficients de développement se confirme sur les loss ratios ultimes Les différentes méthodes SCL semblent être plus stables sur des années de souscription développées : dés la fin de première année pour le triangle d'assurance et après quatre années pour le triangle de réassurance. Notons qu'un faible écart sur un coefficient de développement en réassurance peut induire un gros écart en montant de sinistre compte tenu de la nature à développement long de la branche.

Les résultats obtenus sur le test de stabilité (cf. ANNEXES PARTIE 1 [2.3]) montrent que les écarts sur les années développées depuis plus de trois ans, sont nuls en assurance et de l'ordre de 1% en réassurance, ce qui est très satisfaisant.
Sur les années faiblement développées, les écarts de résultats sont compris entre 1% et 3% en assurance, ce qui reste assez stable.
En revanche, en réassurance, en particulier pour 2004, il y a une différence de 35% de loss ratio. De plus, il y a une disparité des écarts en fonction du α utilisé, ce qui est moins marqué sur les autres années (cf. ANNEXES PARTIE 1 [2.8]).
Ainsi, sur des branches à déroulements longs la méthode SCL aboutit à des résultats stables pour des années suffisamment développées. Sur les années récentes, en revanche, il y a une variabilité des résultats en fonction du α utilisé.

L'estimation de la charge ultime sur les années plus récentes doit donc être considérée avec beaucoup de prudence. La charge ultime n'est calculée qu'avec un faible recul sur des assiettes de primes parfois très grandes. En effet, l'estimation de la charge ultime est fonction des montants de sinistres déjà déclarés dans le triangle. Or pour une branche ayant des déclarations tardives, par exemple la construction, ces montants sont faibles dans les années les moins développées. L'erreur d'estimation peut donc être fortement amplifiée dans le cas de branches longues.

Cette méthode parait plus fiable sur des branches où les risques sont homogènes car elle utilise l'expérience passée des années d'origine les plus anciennes.

3.2 Les méthodes de COUT MOYEN

Ces méthodes s'appuient sur l'hypothèse que l'historique des coûts moyens d'une branche étudiée, permet d'évaluer les coûts moyens futurs de cette branche et donc les sinistres ultimes.

Il existe deux méthodes basées sur ce principe : avec ou sans triangulation.

3.2.1 COUT MOYEN avec triangulation

L'actuaire dispose du triangle des nombres de sinistres $\left(n_{i,j}\right)_{i+j\leq N}$.
La méthode consiste à créer le triangle des coûts moyens $\left(CM_{i,j}\right)$ et de les projeter jusqu'à l'année N puis de déduire les sinistres ultimes.

Pour cela, la méthode se déroule en quatre étapes :

1. A l'aide d'une méthode de triangulation choisie, le triangle des nombres de sinistres est projeté pour les années de développement N-i+2 à N.
2. Le triangle des coûts moyens est construit :

$$\left(CM_{i,j}\right)=\frac{C_{i,j}}{n_{i,j}} \qquad (2)$$

3. A l'aide d'une méthode de triangulation choisie, le triangle des coûts moyens est projeté pour les années de développement N-i+2 à N.
4. Les sinistres ultimes sont déduits :

$$C_{i,N}=CM_{i,N}\cdot n_{i,N} \qquad (3)$$

3.2.2 COUT MOYEN sans triangulation

Sans triangulation, l'actuaire estime un coût moyen « fixe » auquel il applique un taux exogène et fonction des années d'origine (taux d'inflation du marché, évolution des salaires médicaux…).

Pour cela il suppose le coût moyen $\overline{C_1}$ de la première année d'origine connu et que celle-ci est totalement développée.

$$\overline{C_1}=\frac{C_{1,N}}{n_1}$$

Soit I(i) le taux exogène, alors :

$$\overline{C_i}=\overline{C_1}\cdot I(i) \text{ et } C_{i,N}=\overline{C_1}\cdot I(i)\cdot n_i$$

avec n_i le nombre de sinistre de l'année d'origine i à la date du calcul.

Afin d'illustrer les avantages et les inconvénients de cette méthode, son application ne sera effectuée que sur le triangle d'assurance. En effet, le réassureur en particulier sur des branches en quote-part, n'a généralement pas connaissance du nombre de sinistres déclarés par la cédante. Cette méthode n'est donc pas très répandue en réassurance.

Le tableau suivant fournit les Loss Ratio Ultimes obtenus avec la méthode du Cout Moyen sur le triangle d'assurance estimés avec un triangle vu en N et un autre vu en N-1 :

loss ratios ultimes	1995	1996	1997	1998	2000	2001	2002	2003	2004
Estimés en N	81,5%	75,2%	80,7%	74,5%	69,2%	90,4%	69,2%	75,9%	62,4%
Estimés en N-1	81,5%	75,2%	80,7%	74,5%	68,9%	95,3%	74,3%	84,2%	

Loss Ratio ultimes Cout Moyen du triangle d'assurance

Cette méthode nécessite une parfaite fiabilité du dénombrement des sinistres. Elle dépend aussi de la méthode de triangulation retenue pour projeter le nombre de sinistres et les coûts moyens. Le test de sensibilité montre la forte variabilité de la méthode sur les années d'origine les plus récentes, de 5% à 8%.

3.3 Les méthodes d'ajustement de courbes

Le principe des méthodes d'ajustement de courbes est d'ajuster ou lisser par une fonction spécifique, soit les facteurs de développement, soit les ratios S/P. Elles proposent parfois aussi de projeter les facteurs de développement ou les ratios S/P pour obtenir l'ultime. L'ajustement de courbes empiriques par une fonction permet aussi éventuellement d'introduire, sur des triangles sans exercice clos, un facteur de queue de distribution (« Tail factor »).

3.3.1 La méthode de Sherman

La fonction d'ajustement utilisée par cette méthode est l'« inverse power » :

$$f_j = 1 + a(t+c)^{-b} \qquad \text{avec } t = j+1$$

Les propriétés de cette fonction sont telles qu'elle s'adapte facilement aux valeurs empiriques et bien mieux qu'un certain nombre d'autres lois statistiques.
Trois paramètres sont à estimer : a et b se déterminent à l'aide d'une régression linéaire simple tandis que c s'obtient par balayage successif en recherchant le R^2 maximum de l'équation ci-dessus.

L'estimation des paramètres se fait à partir de l'équation suivante :

$$\ln(f_j - 1) = -b.\ln(a(t+c))$$

Le paramètre a est un estimateur du niveau des facteurs de développement. Le niveau de tous les facteurs de développement varie dans le même sens que l'évolution du paramètre. Il s'agit

d'un effet de translation. Lorsque la liquidation est rapide, son impact est plus significatif sur les premières années de développement. Sur une liquidation longue, l'impact tend à être semblable sur toutes les années de développement.

Le paramètre *b* est un indicateur de la vitesse de liquidation de la charge. Lorsque b diminue, la courbe s'aplatit avec rééquilibrage entre les premières années de développement et les dernières. Il a un rôle de pivot.

Le paramètre *c* est également un indicateur de la vitesse de liquidation de la charge mais, à la différence du paramètre *b*, son impact ne se situe significativement que sur les premières années de développement. Il accentue davantage la courbure de la fonction.

L'application numérique faite sur les triangles N et N-1 illustre la forte variabilité de la méthode. Le R^2 est sensiblement meilleur avec les données portant sur le triangle ayant une diagonale en moins. La flexibilité d'ajustement de la fonction s'explique par la présence de 3 paramètres. Un bon R^2 ne renseigne pas sur la robustesse de l'estimation. Une attention toute particulière doit être portée lors de l'utilisation de cette méthode sur un triangle volatil. En effet, l'ajout d'une année supplémentaire dans le triangle de réassurance dégrade significativement la qualité de l'ajustement (voir ANNEXES PARTIE 1 [2.4], le R^2 passe de 0,96 à 0,77).

3.3.2 La méthode de Craighead

La fonction d'ajustement utilisée par cette méthode est une loi exponentielle négative (ou loi de Weibull) :

$$f_j = A\left(1 - e^{\left(\frac{-j}{B}\right)^C}\right) \text{ ou } Q_{i,j} = A\left(1 - e^{\left(\frac{-j}{B}\right)^C}\right)$$

A, *B* et *C* sont les paramètres de la loi exponentielle négative.
Cette méthode est souvent employée sous sa seconde forme c'est-à-dire appliquée aux ratios de sinistres sur primes $Q_{i,j}$.

Cette loi a trois paramètres, mais deux seulement sont à estimer. *A* se définit comme le ratio S/P ultime ou encore la charge ultime. Les paramètres *B* et *C* décrivent quant à eux la longueur de la liquidation ainsi que la courbure de la fonction. Mais leur effet respectif est contraire : *B* augmente avec l'accroissement de longueur de la liquidation alors que *C* diminue.

Ainsi, en se basant sur les exercices déjà liquidés, il est possible d'estimer B et C par régression linéaire à partir de l'équation suivante :

$$\ln\left(-\ln\left(1 - \frac{Q_{i,j}}{A}\right)\right) = C\left(\ln(j) - \ln(B)\right)$$

Une fois B et C estimés, $Q_{i,N}$ peut être déterminé pour chacune des années d'origine en utilisant l'équation suivante :

$$A = \frac{Q_{i,j}}{\left(1 - e^{\left(\frac{-j}{B}\right)^C}\right)}$$

La stabilité de la méthode a été testée différemment des autres méthodes. En effet, en raison des informations nécessaires à l'estimation, la fonction de lissage a été ajustée à partir des années d'origine 1992–1995 puis 1992-1994 pour avoir une idée de l'impact d'une année d'origine en moins sur la courbe.

Les tableaux suivants fournissent les Loss Ratio Ultimes obtenus avec la méthode de Craighead sur le triangle d'assurance et de réassurance estimés avec un triangle vu en N et un autre vu en N-1 :

loss ratios ultimes	**1995**	**1996**	**1997**	**1998**	**2000**	**2001**	**2002**	**2003**	**2004**
Estimés en N	81,5%	75,2%	80,7%	74,5%	69,2%	92,5%	68,1%	77,7%	48,3%
Estimés en N-1	81,5%	75,2%	80,7%	74,5%	69,4%	92,6%	68,1%	77,8%	

Loss Ratio ultimes Craighead du triangle d'assurance

loss ratios ultimes	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**
Estimés en N	111,5%	82,0%	80,2%	93,3%	77,0%	87,1%	88,7%	98,0%	82,8%	73,3%	60,2%	55,0%	71,7%	2,3%
Estimés en N-1	111,5%	82,0%	80,2%	93,3%	77,1%	87,3%	89,0%	98,5%	83,5%	74,3%	61,4%	56,7%	75,1%	

Loss Ratio ultimes Craighead du triangle de réassurance

Les résultats de cette méthode appliquée à nos deux triangles sont sensiblement les mêmes, y compris dans le cas de la réassurance.
En revanche, l'estimation de la charge de sinistres sur les années d'origine récentes ne dépend que de la forme de la courbe de liquidation des années d'origine les plus anciennes. En cas de modifications des règles d'indemnisation, cela peut entraîner des erreurs d'estimation.

3.4 La méthode du LOSS RATIO

Cette méthode est basée sur l'hypothèse que le ratio de sinistres sur primes ultime, appelé aussi Loss Ratio, est connu a priori. Ce dernier est noté L_i^{LR} et correspond à un indicateur d'exposition au risque. Il est déterminé de différentes façons : selon l'expérience passée, à partir d'une évaluation liée à la tarification initiale des contrats ou sur une bonne connaissance du marché.

La charge ultime de sinistres s'écrit ainsi :

$$E(C_{i,N}) = L_i^{LR} \times A_i$$

Le niveau de provisions est alors directement déduit par différence entre la charge ultime estimée et les sinistres déjà réglés après j années :

$$R_i^{LR} = E(C_{i,N}) - C_{i,j}$$

Cette méthode parait très simple, elle n'utilise pas les déroulés des triangles, elle ne dépend que d'un paramètre exogène.
Elle présuppose une très bonne connaissance du marché. Il faut connaître par exemple les changements tarifaires d'une année sur l'autre. En réassurance, le souscripteur peut donner une bonne estimation du Loss Ratio attendu de son portefeuille.

Sur les deux triangles, les triangles des $Q_{i,j}$ ont été observés et les S/P ultimes ont été déduits en fonction de la connaissance des marchés respectifs ou encore de l'évolution des S/P des années antérieures (cf. ANNEXES PARTIE 1 [2.6]).

Le test de stabilité Δ ne peut être appliqué à cette méthode. En effet, le choix de ratio de sinistre sur prime est très subjectif. La variabilité de cette méthode réside dans ce choix.

Cette méthode n'est donc efficace que sur des branches stables car elle n'utilise pas l'expérience passée des sinistres dans l'évaluation de la charge finale, même si cette expérience est inclue dans le choix du loss ratio. Elle est utile pour une première estimation des sinistres ultimes sur les années de souscription récentes qui n'ont pas ou peu d'historique.

3.5 La méthode d'Hertig

A l'instar de la méthode Chain Ladder, la méthode développée par Joakim Hertig (cf. Bibliographie [4]) est basée sur l'analyse des facteurs de développement. Plus précisément, ceux-ci sont calculés à partir des S/P du triangle.

Cette méthode repose sur deux hypothèses de base :

- les ratios $Q_{i,j}$ suivent une loi lognormale
- le développement de j à j+1 des S/P pour une année d'origine i suit une croissance logarithmique $dq_{i,j}$

$$dq_{i,j} = \ln\left(\frac{Q_{i,j+1}}{Q_{i,j}}\right) \qquad \text{avec } j = 1,\ldots,N$$

Si l'examen des données suggère de retenir une distribution normale des $dq_{i,j}$,

$$dq_{i,j} \sim N\left(m_j, s_j^2\right)$$

Les paramètres m_j et s_j^2 sont alors estimés par :

$$\hat{m}_j = \frac{\sum_i dq_{i,j}}{N_j} \sim N\left(m_j, \frac{s_j^2}{N_j}\right)$$

$$\hat{s}_j^2 = \frac{\sum_i \left(dq_{i,j} - \hat{m}_j\right)^2}{N_j - 1} \sim s_j^2 \frac{c_{(f_j)}^2}{f_j} \sim \frac{s_j^2}{N_j - 1} c_{N_j - 1}^2$$

avec N_j le nombre de taux de croissance observé du délai j au délai $j+1$

L'estimation centrale du S/P Q_i conditionnellement à $Q_{i,p}$ avec $p = I = i+1$ s'obtient alors à l'aide de la formule suivante :

$$\hat{Q}_{i|p} = Q_{i,p} e^{\left(\hat{m}_p + \ldots + \hat{m}_N\right)} e^{\frac{s_{(p)}^2}{2}}$$

avec $s_{(p)}^2 = \hat{s}_p^2 \frac{N_p + 1}{N_p} + \ldots + \hat{s}_J^2 \frac{N_N + 1}{N_N}$ car les années de développement sont indépendantes par hypothèse.

La variance de l'estimateur se détermine alors ainsi :

$$V\left[\hat{Q}_{i|p}\right] = \hat{Q}_{i|p}\sqrt{e^{s^2(p)} - 1}$$

L'application numérique fait apparaître une forte variabilité des estimations entre deux dates sur les années récentes du triangle de réassurance. Fondée sur les mêmes principes que Chain Ladder (utilisation de la seule expérience passée), les estimations issues de cette méthode pâtissent de la même manière du manque de recul nécessaire, rendant leurs qualités très incertaines.

3.6 La méthode de BORNHUETTER-FERGUSON

Bien que plus sophistiquée, la méthode de Bornhuetter-Ferguson est basée sur une combinaison des principes des méthodes du Loss Ratio et Chain Ladder. En effet, elle utilise une cadence et un paramètre exogène pour estimer la charge ultime de sinistres d'une année d'origine donnée.

Elle repose sur trois hypothèses principales :

(1) Une estimation ex ante des sinistres est déterminée comme dans la méthode du Loss Ratio :

$$\hat{C}_{i,N} = E\left(C_{i,N}\right) = L_i^{BF} \times A_i \qquad \text{(H5)}$$

(2) Les sinistres non connus (ou non réglés) sont indépendants des sinistres déjà connus (ou réglés)

$$C_{i,j} \quad \perp \quad C_{i,N} - C_{i,j} \qquad \text{(H6)}$$

(3) Les facteurs de développement $F_{i,j}$ sont tels que :

$$F_{i,j} = \frac{E\left(C_{i,N}\right)}{E\left(C_{i,j}\right)} \qquad \text{(H7)}$$

La détermination des $F_{i,j}$ est généralement issue des cadences de développement de la méthode de Chain Ladder.

Une estimation de la charge ultime de sinistres est obtenue avec la formule suivante :

$$\hat{C}_{i,N}^{BF} = C_{i,j} + \left(1 - \frac{1}{F_{i,j}}\right) \times \hat{C}_{i,N}$$

La provision se déduit alors :

$$R_i^{BF} = \hat{C}_{i,N}^{BF} - C_{i,j} = \left(1 - \frac{1}{F_{i,j}}\right) \times \hat{C}_{i,N}$$

L'ultime du Bornhuetter-Ferguson peut encore s'écrire :

$$\hat{C}_{i,N}^{BF} = \frac{1}{F_{i,j}} \times C_{i,j} \times F_{i,j} + \left(1 - \frac{1}{F_{i,j}}\right) \times \hat{C}_{i,N}$$

$$\hat{C}_{i,N}^{BF} = w_{i,j} \times C_{i,j} \times F_{i,j} + \left(1 - w_{i,j}\right) \times \hat{C}_{i,N}$$

avec $w_{i,j} = \frac{1}{F_{i,j}}$

Ainsi, l'estimation de la charge ultime selon la méthode de Bornhuetter Ferguson correspond à une pondération entre une estimation de la méthode de cadencement de référence par exemple de type Chain Ladder et la charge de sinistres ultimes supposée connue via un S/P ultime défini a priori.

Soit encore :

$$\hat{C}_{i,N}^{BF} = w_{i,j} \times \hat{C}_{i,N}^{CL} + \left(1 - w_{i,j}\right) \times \hat{C}_{i,N}^{LR} \quad (4)$$

Un désavantage ou difficulté de cette méthode est qu'elle utilise un loss ratio attendu qui n'est pas toujours connu. Les résultats dépendent fortement de ce ratio ainsi que de l'aliment choisi.

En revanche, et contrairement à la méthode du « loss ratio », cette méthode tient compte des sinistres déjà connus dans le calcul de la charge ultime de sinistres à l'aide d'une cadence de développement.

Les applications numériques sur les deux triangles ont été faites à l'aide des différentes méthodes de Chain Ladder vues précédemment afin d'observer l'impact du choix de la méthode de cadencement. Les S/P ultimes conservés, sont ceux utilisés dans la méthode du Loss Ratio. Les ultimes ainsi obtenus sont ci-dessous :

loss ratios ultimes	1995	1996	1997	1998	2000	2001	2002	2003	2004
Estimés en N	81,5%	75,2%	80,7%	74,5%	69,2%	92,6%	67,9%	75,4%	65,1%
Estimés en N-1	81,5%	75,2%	80,7%	74,5%	69,5%	92,1%	69,5%	77,0%	

Loss Ratio ultimes Bornhuetter-Ferguson du triangle d'assurance

loss ratios ultimes	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Estimés en N	111,5%	82,0%	80,6%	94,5%	78,6%	88,9%	90,7%	100,2%	85,1%	76,2%	62,7%	58,9%	74,4%	62,5%
Estimés en N-1	111,5%	81,5%	79,9%	93,9%	78,7%	89,7%	90,1%	99,1%	85,1%	77,7%	67,6%	54,4%	83,5%	

Loss Ratio ultimes Bornhuetter-Ferguson du triangle de réassurance

Les résultats moyens et les écarts entre les ultimes obtenus en fonction de la méthode de développement sont en ANNEXES PARTIE 1 [2.8].

Si l'on compare les écarts types relatifs entre les résultats des méthodes Chain Ladder et le Bornhuetter-Ferguson, la variabilité parait similaire sur les années antérieures à 2005. Sur l'année 2005, la méthode Bornhuetter-Ferguson obtient des résultats beaucoup plus stables. Ceci est observé sur l'application des deux triangles, même si les résultats obtenus sur le triangle d'assurance ne sont que faiblement significatifs.

Le test de stabilité Δ sur la méthode de Bornhuetter-Ferguson est fait avec la méthode de cadencement du Chain Ladder a =1, soit le Chain Ladder standard (cf. ANNEXES PARTIE 1 [2.3]).

Sur le triangle de réassurance, les écarts sur les années antérieures à 2001 sont très faibles et similaires à la méthode de cadencement seule utilisée. Sur les années 2003 et 2004, le test de sensibilité donne des résultats nettement meilleurs que la méthode Chain Ladder. Sur ces années, c'est le poids du S/P choisi qui est le plus fort.

Les deux méthodes précédentes (Loss Ratio et Bornhuetter Ferguson) sont fortement dépendantes de l'estimation de la charge ultime a priori que peut faire l'actuaire. Ainsi, elles ne garantissent pas une stabilité du résultat année après année.

Le test de stabilité confirme ce qui est observé sur les écarts relatifs.
Pour les deux triangles, sur les années fortement développées, les résultats du test de stabilité sont très faibles et similaires à la méthode de cadencement seule utilisée. Sur les années en cours de développement le test de sensibilité donne des résultats nettement meilleurs que la méthode Chain Ladder (surtout sur le triangle de réassurance). Sur ces années, c'est le S/P choisi qui a le poids le plus fort.

Les deux méthodes précédentes (Loss Ratio et Bornhuetter Ferguson) sont fortement dépendantes de l'estimation de la charge ultime a priori que peut faire l'actuaire. Ainsi, elles ne garantissent pas une stabilité du résultat année après année.

3.7 La méthode de CAPE COD

Cette méthode est en quelque sorte une extension de la méthode de Bornhuetter Ferguson. Elle tente de s'affranchir du caractère subjectif de la détermination a priori de la charge ultime de sinistres.

Cette méthode estime et applique un Loss Ratio $L_{G_S}^{CC}$ sur un groupe G_s d'années d'origine a priori semblables. Il est déterminé par le rapport entre la somme des sinistres comptabilisés à la date du calcul et la somme des aliments ultimes rapportés à la date actuelle estimés par une méthode de cadencement :

$$L_{G_S}^{CC} = \frac{\sum_{k \in G_S} C_{k,N-k}}{\sum_{k \in G_S} \frac{A_k}{F_{k,N-k}}}$$

Avec G_s le s^{ème} groupe d'années d'origine comparables, $s = \{1,2,...,S\}$ et S le nombre de groupe d'années d'origine comparables.

Reprenons les hypothèses du modèle de Bornhuetter-Ferguson pour poser celles du modèle Cape Cod :

(1) Une estimation ex ante des sinistres est déterminée comme dans la méthode du Loss Ratio pour chaque groupe G_s :

$$\hat{C}_{i,N}^{G_S} = E\left(C_{i,N}^{G_S}\right) = L_{G_S}^{CC} \times A_i \qquad \text{(H8)}$$

(2) Les sinistres non connus (ou non réglés) sont indépendants des sinistres déjà connus (ou réglés)

$$C_{i,j} \quad \perp \quad C_{i,N} - C_{i,j} \qquad \text{(H9)}$$

(3) Les facteurs de développement $F_{i,j}$ sont tels que :

$$F_{i,j} = \frac{E\left(C_{i,N}\right)}{E\left(C_{i,j}\right)} \qquad \text{(H10)}$$

L'estimation de la charge ultime devient donc :

$$\begin{aligned} \hat{C}_{i,N}^{CC} &= C_{i,j} + \left(1 - \frac{1}{F_{i,j}}\right) \times E\left(\hat{C}_{i,N}^{G_S}\right) \\ &= C_{i,j} + \left(1 - \frac{1}{F_{i,j}}\right) \times L_{G_s}^{CC} \times A_i \end{aligned} \qquad \textit{pour tout } i \in G_s$$

De plus, sous hypothèse d'indépendance entre les années d'origine au sein d'un même groupe G_s , on pose :

$$F_{i,j} = F_j$$

La somme des provisions pour le s[ème] groupe peut alors s'écrire :

$$\sum_{i \in G_s} R_i^{CC} = L_{G_s}^{CC} \times \sum_{i \in G_s} \left(1 - \frac{1}{F_{N-i+1}}\right) \times A_i$$

et la somme totale des provisions :

$$R^{CC} = \sum_{s=1}^{S} \sum_{i \in G_s} R_i^{CC}$$

De même que pour les méthodes Loss Ratio et Bornhuetter Ferguson, les provisions correspondent à une part non absorbée de l'aliment multipliée par un niveau de risque.

L'estimation du Loss ratio dans la méthode Cape Cod découle de l'estimation de l'ultime de cette méthode :

$$\sum_{i=1}^{I} C_{i,N-i+1} + \sum_{i=1}^{I} R_i = \sum_{i=1}^{I} \left(L_i \times A_i\right) \text{ avec } L_i = L_{G_S} \qquad \textit{pour} \quad i \in G_S$$

Dés lors pour le s[ème] groupe :

$$\sum_{i \in G_s} \left(L_i \times A_i\right) = L_{G_s}^{CC} \sum_{i \in G_s} A_i$$

Ainsi :

$$\sum_{i\in G_s} C_{i,N-i+1} + L_{G_s}^{CC} \times \sum_{i\in G_s}\left(1-\frac{1}{F_{N-i+1}}\right)\times A_i = L_{G_s}^{CC}\sum_{i\in G_s} A_i$$

$$\sum_{i\in G_s} C_{i,N-i+1} = L_{G_s}^{CC} \times \sum_{i\in G_s}\left(\frac{A_i}{F_{N-i+1}}\right)$$

$$L_{G_s}^{CC} = \frac{\sum_{i\in G_s} C_{i,N-i+1}}{\sum_{i\in G_s}\left(\frac{A_i}{F_{N-i+1}}\right)}$$

Le niveau de charge ultime se détermine comme le rapport entre le montant cumulé des charges et le montant de l'aliment déjà absorbé. Le taux d'absorption de l'aliment étant calculé à partir des cadences de règlement (issues par exemple de la méthode SCL).

Cette méthode nécessite elle aussi une bonne connaissance de son portefeuille afin de pouvoir effectuer les segmentations adéquates des années de souscription semblables. Elle est efficace sur des branches ayant des politiques de souscription homogènes.

Les ultimes ainsi obtenus sont ci-dessous :

loss ratios ultimes	**1995**	**1996**	**1997**	**1998**	**2000**	**2001**	**2002**	**2003**	**2004**
Estimés en N	81,5%	75,2%	80,7%	74,5%	69,2%	92,6%	67,9%	75,3%	66,6%
Estimés en N-1	81,5%	75,2%	80,7%	74,5%	69,5%	92,0%	69,5%	77,1%	

Loss Ratio ultimes Cape Cod du triangle d'assurance

loss ratios ultimes	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**
Estimés en N	111,5%	82,0%	80,6%	94,4%	78,6%	89,0%	90,7%	100,0%	84,3%	75,1%	62,0%	57,8%	72,2%	56,6%
Estimés en N-1	111,5%	81,5%	79,9%	93,9%	78,7%	89,8%	90,2%	98,9%	84,3%	76,5%	67,1%	53,3%	80,8%	

Loss Ratio ultimes Cape Cod du triangle de réassurance

Les résultats moyens et les écarts entre les ultimes obtenus sont en ANNEXES PARTIE 1 [2.8].

En comparant les écarts types relatifs (en pourcentage du sinistre moyen obtenu) entre les résultats des méthodes SCL et Cape Cod, la variabilité est similaire sur les années ayant plus d'une année de développement. Sur l'année n'ayant qu'une année de développement, la méthode Cape Cod comme la méthode de Bornhuetter-Ferguson obtient des résultats beaucoup plus stables.

Le test de stabilité Δ sur la méthode de Cape Cod est fait avec la méthode de cadencement du Chain Ladder a =1, soit le Chain Ladder standard (cf. ANNEXES PARTIE 1 [2.3]).

Les résultats du test de stabilité sont comparables à ceux obtenus sur la méthode de Bornhuetter-Ferguson.

3.8 La méthode de BENKTANDER

En 1976, Benktander publia un article dans lequel il propose une nouvelle méthode de provisionnement. Celle-ci est une combinaison linéaire entre les estimations de provisions des méthodes Chain Ladder et Bornhuetter Ferguson.

La provision s'écrit alors :

$$R_i^{BK} = w.R_i^{CL} + (1-w).R_i^{BF}$$

La pondération w proposée dans la méthode correspond à la part déjà connue des sinistres.

$$w = \frac{1}{F_{i,N-i+1}}$$

avec $F_{i,N-i+1}$ estimé à l'aide de la méthode Chain Ladder.

L'idée sous jacente est que plus l'expérience augmente, plus l'estimation réalisée par la méthode Chain Ladder est fiable.

Ainsi, la provision s'écrit :

$$R_i^{BK} = \frac{1}{F_{i,N-i+1}}.R_i^{CL} + (1-\frac{1}{F_{i,N-i+1}}).R_i^{BF} \qquad (5)$$

Cette équation peut aussi s'écrire en fonction des charges ultimes :

$$\begin{aligned} R_i^{BK} &= \frac{1}{F_{i,N-i+1}}.\left(\hat{C}_{i,N}^{CL} - C_{i,N-i+1}\right) + (1-\frac{1}{F_{i,N-i+1}}).\left(\hat{C}_{i,N}^{BF} - C_{i,N-i+1}\right) \\ &= \frac{1}{F_{i,N-i+1}}.\hat{C}_{i,N}^{CL} + (1-\frac{1}{F_{i,N-i+1}}).\hat{C}_{i,N}^{BF}. - C_{i,N-i+1} \\ &= \hat{C}_{i,N}^{BK} - C_{i,N-i+1} \end{aligned}$$

L'estimation de la charge ultime selon la méthode de Benktander, s'écrit alors :

$$\hat{C}_{i,N}^{BK} = \frac{1}{F_{i,N-i+1}} \times \hat{C}_{i,N}^{CL} + \left(1-\frac{1}{F_{i,N-i+1}}\right) \times \hat{C}_{i,N}^{BF}$$

Ce résultat est à rapprocher de la formule du modèle de Bornhuetter-Ferguson (4), à ceci prés que :

1. la cadence de règlement retenue dans l'estimation du $\hat{C}_{i,N}^{BF}$ est celle de la méthode de Chain Ladder (SCL)
2. la charge de sinistre attendue utilisée dans le second terme de l'équation (4) n'est plus celle établie a priori mais celle estimée in fine dans la méthode de Bornhuetter-Ferguson. A ce titre cette méthode est aussi appelée la méthode de Bornhuetter-Ferguson itérative.

Les loss ratios ultimes issus de la méthode Benktander sont ci-dessous :

loss ratios ultimes	**1995**	**1996**	**1997**	**1998**	**2000**	**2001**	**2002**	**2003**	**2004**
Estimés en N	81,5%	75,2%	80,7%	74,5%	69,2%	92,6%	67,8%	75,5%	62,9%
Estimés en N-1	81,5%	75,2%	80,7%	74,5%	69,5%	92,0%	69,1%	77,9%	

Loss Ratio ultimes Benktander du triangle d'assurance

loss ratios ultimes	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**
Estimés en N	111,5%	82,0%	80,6%	94,4%	78,6%	88,9%	90,6%	100,3%	84,9%	75,3%	61,2%	55,8%	72,8%	49,0%
Estimés en N-1	111,5%	81,5%	79,9%	93,9%	78,7%	89,7%	90,1%	99,1%	84,8%	76,6%	65,7%	47,0%	86,3%	

Loss Ratio ultimes Benktander du triangle de réassurance

Les tests de stabilité sont en ANNEXES PARTIE 1 [2.3]. Sur les années faiblement développées, en particulier l'année la moins développée, le test de stabilité est un peu moins bon que pour la méthode de Bornhuetter-Ferguson, mais il est nettement meilleur que celui de la méthode de SCL.

3.9 La méthode GLM (Generalized Linear Model)

Cette méthode a été décrite en 1972 par NELDER et WEDDERBURN. Il s'agit d'un modèle qui s'applique à la famille des lois exponentielles. Il s'agit tout d'abord de faire l'hypothèse que l'espérance de $X_{i,j}$, notée $m_{i,j}$, est une fonction linéaire de *p* prédicteurs avec *p=I+J-1*.

Le modèle à estimer s'écrit sous la forme suivante :

$$h_{i,j} = g(m_{i,j}) = c_0 + a_i + b_j$$

avec $m_{i,j} = E(X_{i,j})$

a_i le paramètre associé à l'année d'origine *i*

b_j le paramètre associé au délai de développement *j*

Sous sa forme matricielle, le modèle s'écrit également comme suit :

$$h_{i,j} = M'.b$$

avec M' la matrice identité

b le vecteur des paramètres a_i et b_j

$h_{i,j}$ est appelé le prédicteur linéaire. Il correspond à la transformation de la variable $m_{i,j}$ définie selon le choix de la fonction *g*, appelée fonction de lien.
Le choix de *g* dépend de l'hypothèse faite sur la distribution de la variable à expliquer $X_{i,j}$ (cf. quelques exemples de correspondance dans le tableau ci-dessous).

Forme de g	Distribution de $X_{i,j}$
Identité	Normale
logarithme	Poisson Gamma
Logit	Binomiale

L'application numérique porte pour le triangle d'assurance sur les paiements non cumulés et sur les sinistres survenus dans le cas du triangle de réassurance. La distribution Poisson est retenue dans les deux cas.

Les loss ratios ultimes estimés à partir de la méthode GLM sont ci-dessous :

loss ratios ultimes	1995	1996	1997	1998	2000	2001	2002	2003	2004
Estimés en N	81,5%	75,2%	80,7%	74,5%	69,2%	92,6%	67,8%	75,5%	61,2%
Estimés en N-1	81,5%	75,2%	80,7%	74,5%	69,5%	92,1%	69,1%	78,6%	

Loss Ratio ultimes GLM du triangle d'assurance

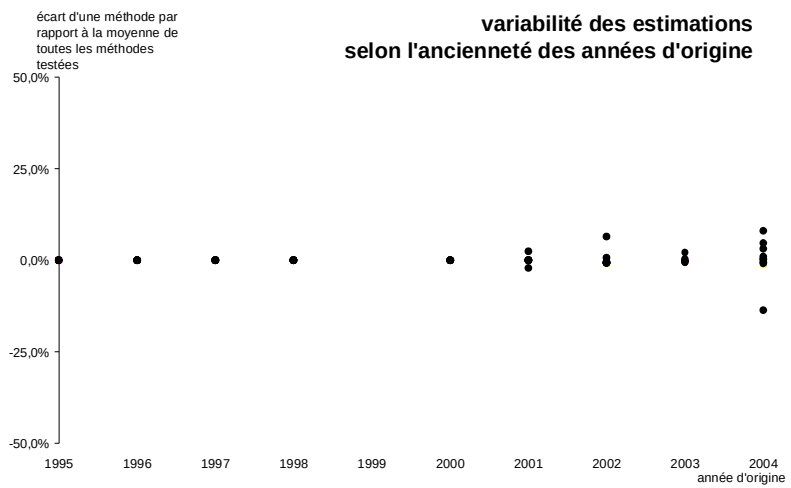
loss ratios ultimes	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Estimés en N	111,6%	81,0%	81,2%	96,1%	82,0%	87,1%	74,9%	78,0%	76,4%	68,9%	65,2%	47,2%	72,8%	4,5%
Estimés en N-1	111,6%	80,9%	81,3%	96,2%	82,4%	86,8%	72,3%	73,7%	74,6%	67,6%	68,7%	42,6%	88,1%	

Loss Ratio ultimes GLM du triangle de réassurance

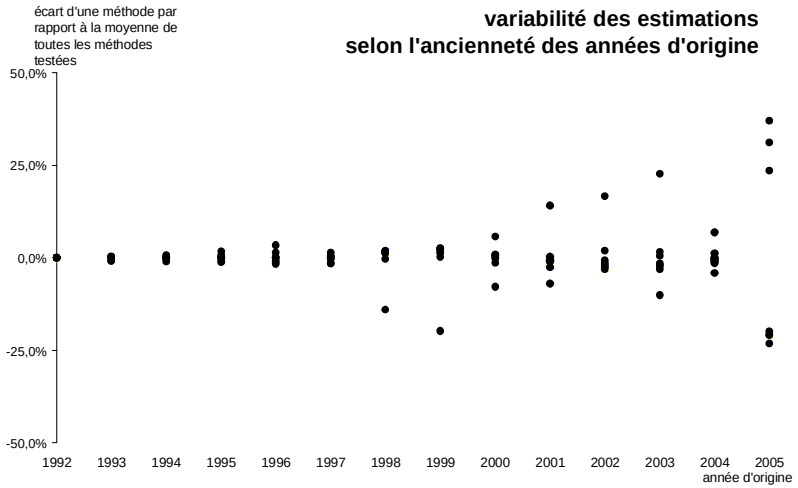
Les tests de stabilité sont en ANNEXES PARTIE 1 [2.3].
Cette méthode, appliquée sur le triangle d'assurance donne des résultats parfaitement identiques à ceux de CHAIN LADDER avec alpha=1.
Cette particularité ne se retrouve pas sur le triangle de réassurance où les résultats diffèrent significativement des autres méthodes utilisées. Cette méthode apparaît également plus stable que les autres, le retrait d'une année d'information n'ayant visiblement que peu d'impact sur la valeur des paramètres estimés.

4 Résumé des avantages et inconvénients des méthodes présentées

Triangle d’assurance



Triangle de réassurance



Les deux graphiques ci-dessus illustrent en synthèse les commentaires que nous avons formulés tout au long de cette première partie et pour chacune des méthodes décrites.

Le triangle d’assurance, souligne les caractéristiques des branches courtes à développement rapide. Le triangle de réassurance représente les branches longues à développement long.

L’incertitude des différentes estimations se concentre sur les premières années de liquidation (ou faiblement développées).
Un triangle homogène conduit à une rapide convergence des estimations des différentes méthodes. A l’inverse, un triangle hétérogène conduit à une variabilité plus forte des résultats selon la méthode utilisée, celle-ci pouvant même se prolonger jusqu’à un stade avancé de la liquidation.

Les chapitres précédents ont permis de présenter les principales méthodes de provisionnement déterministes existantes. Chacune présente des avantages et des inconvénients. Il n'existe pas une bonne ou une mauvaise méthode. C'est plutôt la nature de la branche d'étude des provisions comme le montrent les deux graphiques précédents qui guide le choix d'une méthode par rapport à une.

Voici un tableau récapitulatif des « plus et moins » de chaque méthode suivant l'étude effectuée :

Branches	**courtes - développement rapide**			**Longues - développemet lent**		
typologie des années	**fortement développées**	**intermédiaires**	**faiblement développées**	**fortement développées**	**intermédiaires**	**faiblement développées**
méthodes						
Chain-Ladder	+++	++	+	+++	+	- -
loss ratio	- -	-	++	- -	+	+++
Coût Moyen	+++	+	-	NA	NA	NA
Sherman	+++	++	+	+++	-	- -
Craighead	+++	+	-	+++	-	- -
Hertig	+++	++	+	+++	-	- -
Bornhuetter Ferguson	++	++	++	++	+++	+
Cape Cod	++	++	++	++	+++	+
Benktander	++	++	++	++	+++	+
GLM	+++	++	+	+++	+	- -

On remarque que quelle que soit la méthode utilisée sur les branches courtes, les estimations semblent bonnes ou assez bonnes. Il y a une convergence des méthodes. A ce stade de l'analyse, il est difficile de conclure sur quelles sont les méthodes à crédibiliser compte tenu de leurs avantages.

La seconde conclusion de ce tableau synthétique est que quelle que soit le type de branche (courte ou longue), sur les années fortement développées, la plupart des méthodes fournissent une estimation bonne ou assez bonne. L'exception étant sur la méthode du loss ratio. En effet, la méthode du loss ratio fournit une estimation satisfaisante lorsque les années sont faiblement développées, cette caractéristique est utilisée dans certaines des méthodes utilisant la crédibilité qui seront présentées dans la Partie 2.

Sur les branches longues pour les années faiblement développées ou intermédiaires, on constate que peu de méthodes fonctionnent. Nous verrons dans la Partie 2 que celles qui fonctionnent sont des crédibilisations d'autres méthodes ; c'est le cas des méthodes Bornhuetter-Ferguson, Cape Cod et Benktander.

L'autre conclusion de cette Partie est que les méthodes n'ont pas les mêmes caractéristiques suivant la nature du développement du triangle (long ou rapide). Ce constat est utilisé dans la crédibilisation faite par la méthode du Munich Chain Ladder qui sera analysée à la fin de la Partie 2. Une branche courte peut s'apparenter à un triangle de sinistres survenus et une branche longue à celui des sinistres payés. C'est pourquoi cette méthode crédibilise la méthode du SCL appliquée aux sinistres payés d'une part et survenus d'autre part.

PARTIE 2
LES PREMIERES APPROCHES DE CREDIBILITE LINEAIRE EN PROVISIONNEMENT

Dans cette partie, la crédibilité linéaire est présentée afin d'appréhender la signification de cette notion. Aussi, le lien entre un programme d'optimisation linéaire ou une régression linéaire et la théorie de la crédibilité est exposé.
Certaines des méthodes décrites dans la Partie I sont ensuite reprises afin de constater qu'elles utilisent déjà la théorie de la crédibilité. Enfin une méthode innovante dans le domaine du provisionnement, tant dans la méthode que dans son approche de crédibilité est étudiée : la méthode du Munich Chain Ladder.

1 Présentation de la crédibilité linéaire

1.1 Les origines de la théorie de la crédibilité

Arthur MOWBRAY (1914) a été l'un des premiers actuaires à avoir traité de la théorie de la crédibilité dans la branche d'assurance liée aux accidents du travail. Les préoccupations de l'époque étaient de savoir à partir de quelle taille de portefeuille d'un client, il était possible d'utiliser les seules données individuelles de ce client pour la tarification des garanties.
L'objectif était ainsi de déterminer un seuil d'admissibilité (garantissant la stabilité des estimations) au-delà duquel il était possible d'utiliser l'expérience individuelle comme seule donnée pour la tarification. On parle alors de crédibilité de stabilité ou des fluctuations limitées.
Ces travaux mènent à des décisions binaires. Selon la taille du portefeuille du client, la tarification se faisait soit selon le risque individuel, soit selon le risque global.

Albert W. WHITNEY (1918) introduit le concept de crédibilité partielle. Il mentionne « la nécessité, par souci d'équité pour l'assuré, de pondérer d'un coté l'expérience collective et de l'autre l'expérience individuelle ».

Par la suite, Arthur L. BAILEY (1945, 1950) démontre que la minimisation de l'erreur quadratique dans un contexte bayesien permet d'obtenir une estimation de la prime de crédibilité par une fonction linéaire des observations.

Mais ce sont surtout les travaux de Hans BÜHLMANN (1967, 1969) qui vont populariser la théorie de la crédibilité avec un modèle plus formalisé. A partir de ceux-ci, de nombreux travaux de recherche ont été menés et ont permis de développer et généraliser cette théorie. Parmi les principaux, mentionnons ceux de BÜHLMANN-STRAUB (1970) qui intègrent un facteur de poids aux données dans les formules, ceux de JEWELL (1975) qui approfondissent la généralisation de la formule initiale avec son modèle hiérarchique et ceux de HACHEMEISTER qui, la même année, incorporent la régression linéaire à la théorie de la crédibilité pour traiter d'éventuelles tendances dans les données.

1.2 Principe général de la théorie de la crédibilité

La problématique de la crédibilité est de connaître la répartition des pondérations à donner à plusieurs estimateurs afin de créer un estimateur meilleur que les estimateurs pris individuellement. Montrons via un exemple simple que cette recherche de la pondération s'effectue via un programme d'optimisation linéaire.

Soient $(X_1, \mathbf{L}, X_n)$ un échantillon de taille n de la loi parente P_X, les X_i sont indépendantes et identiquement distribuées (iid) de même loi qu'une variable aléatoire X.

Le modèle statistique est : $(c^n, \mathrm{B}, \mathrm{P})$

L'espérance mathématique de X dans cet univers $E(X)$ peut être déterminée à l'aide de différents estimateurs $\hat{X}$. Ici nous posons deux estimateurs différents.

Soit $\hat{X}_1$ et $\hat{X}_2$ deux estimateurs de X,

on pose par construction que $E(\hat{X}_1) = E(\hat{X}_2) = E(X)$

Ces estimateurs sont ainsi sans biais.

Leur variance est supposée égale respectivement à $V(\hat{X}_1)$ et $V(\hat{X}_2)$.

A présent, analysons la qualité de ces deux estimateurs à partir de leur résidu $e = X - \hat{X}$:

On aura :

$$E(e) = E(X - \hat{X}_i) = 0 \text{ dans les deux cas}$$

et

$$V(e) = V(X - \hat{X}) = E\left(\left(X - \hat{X}_i - E(X - \hat{X}_i) \right)^2 \right)$$

On cherche à présent à définir un troisième estimateur $\hat{X}_3$ (« méta estimateur ») comme une combinaison de $\hat{X}_1$ et $\hat{X}_2$ et dont l'objectif est d'améliorer l'estimation de X par rapport à celles fournies indépendamment par $\hat{X}_1$ et $\hat{X}_2$.

Posons $\hat{X}_3 = z \cdot \hat{X}_1 + (1-z) \cdot \hat{X}_2$ *avec* $z \in [0,1]$ et $e_3 = X - \hat{X}_3$

$$E(\hat{X}_3) = E(z \cdot \hat{X}_1 + (1-z) \cdot \hat{X}_2) = z \cdot E(\hat{X}_1) + (1-z) \cdot E(\hat{X}_2) = z \cdot E(X) + (1-z) \cdot E(X) = E(X)$$

Cet estimateur est encore sans biais.

$$V(\hat{X}_3) = V(z \cdot \hat{X}_1 + (1-z) \cdot \hat{X}_2) = z^2 \cdot V(\hat{X}_1) + (1-z)^2 \cdot V(\hat{X}_2) + 2 \cdot z \cdot (1-z) \cdot COV(\hat{X}_1; \hat{X}_2)$$

Etude du résidu ε_3 :

On aura toujours $E(e_3) = E(X - \hat{X}_3) = 0$

$$\begin{aligned} V(e_3) &= V\left(X - \left(z \cdot \hat{X}_1 + (1-z) \cdot \hat{X}_2\right)\right) \\ &= z^2 \cdot V(X - \hat{X}_1) + (1-z)^2 \cdot V(X - \hat{X}_2) + 2 \cdot z \cdot (1-z) \cdot COV(X - \hat{X}_1; X - \hat{X}_2) \\ &= z^2 \cdot V(e_1) + (1-z)^2 \cdot V(e_2) + 2 \cdot z \cdot (1-z) \cdot COV(e_1; e_2) \\ &= z^2 \cdot (V(e_1) + V(e_2) - 2 \cdot COV(e_1; e_2)) + 2 \cdot z \cdot (COV(e_1; e_2) - V(e_2)) + V(e_2) \end{aligned}$$

Posons $\Phi(z)$ telle que :

$$\Phi(z) = V(e_3)$$

$$\frac{d\Phi(z)}{dz} = 2 \cdot (z \cdot (V(e_1) + V(e_2) - 2 \cdot COV(e_1; e_2)) + 2 \cdot COV(e_1; e_2) - V(e_2))$$

$$\frac{d\Phi(z)}{dz} = 0 \Leftrightarrow z = \frac{V(e_2) - COV(e_2; e_1)}{V(e_1) + V(e_2) - 2 \cdot COV(e_2; e_1)}$$

$$\frac{d\Phi(z)}{dz} = 0 \Leftrightarrow z^* = \frac{V(e_2) - COV(e_2; e_1)}{V(e_1 - e_2)}$$

Etude de la dérivée seconde de $\Phi(z)$

$$\frac{\partial^2 \Phi(z)}{\partial z^2} = z^2 \cdot [V(e_1) + V(e_2) - 2.COV(e_2; e_1)] = V(e_1 - e_2) > 0 \qquad (a)$$

L'étude de la variance du résidu ε_3 permet de vérifier que la fonction est bien convexe (fonction quadratique) et qu'elle admet bien un minimum en $z^* \in [0;1]$.

Il est ainsi possible de trouver un « méta estimateur » $\hat{X}_3 = z^* \cdot \hat{X}_1 + (1-z^*) \cdot \hat{X}_2$ au moins aussi bon que les deux autres estimateurs individuels car son résidu est toujours inférieur ou égal à celui des deux autres.

Le fait de trouver un « méta estimateur » meilleur que les deux estimateurs le constituant est une forme de « crédibilisation » des deux estimations initiales. Celle-ci se fait en leur attribuant une pondération z*, fonction de leur qualité respective. Cette pondération a pu être trouvée grâce à un programme de minimisation de la fonction de son résidu.

1.3 Régression linéaire et théorie de la crédibilité

Cette partie vise à faire le lien entre le modèle de régression et la théorie de la crédibilité linéaire. Plus précisément, il s'agit de montrer que la crédibilité linéaire est un cas particulier du modèle de régression, du fait d'hypothèses supplémentaires. Ce résultat sera par la suite utilisé, en particulier dans la sous-partie traitant de la méthode du Munich Chain Ladder.

1.3.1 Notations du modèle de BÜHLMANN et interprétations

Il ne s'agit pas ici de décrire précisément le modèle de crédibilité de Bühlmann mais de présenter les hypothèses et les principaux résultats afin de mettre en évidence sa relation avec le modèle de régression linéaire :

Posons $X_{j,s}$, la variable « expérience de sinistres »
(par exemple, le ratio sinistre sur primes).
j l'indice du contrat
s l'indice de la période

Soit $q_{j,s}$ le paramètre de risque du j^{ème} contrat à la s^{ème} période.
avec j=1,…,k et s=1,…,t

Dans la plupart des modèles, le paramètre de risque pour un contrat particulier ne change pas sur la période d'estimation. Les paramètres se réécrivent alors ainsi : $q_1, q_2, \ldots q_k$.

Le j^{ème} contrat est décrit par le vecteur $(q_j, X_{j,1}, \ldots X_{j,t}) = (q_j, X_j)$.

Outre l'hypothèse d'intégrabilité des variables $X_{j,s}$, les hypothèses du modèle sont :

B1 : les contrats (q_j, X_j) pour $j = 1, \ldots k$ sont indépendants et identiquement distribués.
B2 : pour chaque contrat j et pour $q_{j,s} = q_j$, les variables conditionnelles $X_{j,1}/q_j, \ldots X_{j,t}/q_j$ sont indépendantes et identiquement distribuées.
B3 : il existe une fonction de distribution $U(q)$ connue et identique à tous les paramètres de risque appelée aussi « fonction de structure ».

B1 représente l'indépendance entre les contrats alors que B2 représente l'indépendance à l'intérieur d'un contrat (ou autrement dit, l'homogénéité dans le temps).

D'après B2, le paramètre de risque pour un contrat particulier ne change pas sur la période d'estimation. Les paramètres de risque pour chaque contrat j se réécrivent ainsi : $q_1, q_2, \ldots q_k$.

B3 implique que les contrats sont similaires dans leur fonction de structure mais se distinguent car leurs réalisations q_j diffèrent selon les contrats.

Pour faciliter les écritures qui suivront, posons :

$$m(q_j) = E(X_{js} / q)$$
$$m = E(X_{js})$$
$$a = V(E(X_{js} / q_j))$$
$$s^2 = E(V(X_{js} / q_j))$$

Ces écritures permettront de montrer dans la Partie 2 1.3.3 comment la formule de crédibilité peut s'énoncer en partant d'un modèle de régression linéaire simple.

1.3.2 Rappel du modèle de régression linéaire

Soit $(X_1, \mathbf{L}, X_n)$ un échantillon aléatoire et Y une variable aléatoire dont le 2nd moment existe.

Soit le modèle suivant :

$$Y_i = a + bX_i + e_i \qquad \text{(RL)}$$

avec les X_i non corrélés

e_i est le terme d'erreur qui respecte les hypothèses suivantes :

1. les e_i sont identiquement distribuées avec une variance constante s^2 (homoscédasticité des erreurs)
2. les e_i sont indépendamment distribuées pour *i=1,…,n* (non autocorrélation des erreurs)
3. e_i suit une distribution normale

Au final, on a $e_i \sim iid\ N(0\ ;\ s^2)$

L'estimation des paramètres a et b s'obtient par la minimisation de la fonction quadratique φ suivante, appelée aussi la méthode des moindres carrés ordinaires - MCO :

$$j = E\left[(Y - a - bX)^2\right]$$

Pour minimiser cette fonction selon des paramètres a et b, il faut que leur dérivée première s'annule et que φ soit convexe :

$$\frac{dj}{da} = E\left[-2\left(Y - a - bX\right)\right] = 0 \qquad \text{et} \qquad \frac{dj}{db} = E\left[-2X\left(Y - a - bX\right)\right] = 0$$

La résolution de ce système de deux équations permet d'obtenir les estimations $\hat{a}$ et $\hat{b}$ des paramètres respectifs a et b pour l'équation (RL) :

$$\begin{aligned} \hat{a} &= E(Y) - \hat{b}.E(X) \\ \hat{b} &= \frac{COV(X,Y)}{V(X)} \end{aligned} \qquad \text{(RLO)}$$

1.3.3 Application du modèle de régression aux hypothèses de BÜHLMANN

Pour simplifier, raisonnons avec un seul contrat. Reprenons l'hypothèse B2 qui, pour un contrat, définit $(X_1, \mathbf{L}, X_n)$ comme des variables aléatoires conditionnellement indépendantes sachant q et en reprenant les notations de la Partie 2 paragraphe 1.3.1, étudions le modèle suivant :

$$m(q) = a + \sum_{i=1}^{n} b_i X_i \qquad \text{où } m(q) = E(X_{n+1}/q)$$

L'estimation des paramètres a et b s'obtient par la minimisation de la fonction quadratique φ suivante :

$$E\left[\left\{E[m(q)] - a - \sum_{i=1}^{n} b_i X_i\right\}^2\right] = j$$

Nous cherchons à minimiser cette fonction :

$$\frac{dj}{da} = -2.E\left[E[m(q)] - a - \sum_{i=1}^{n} b_i X_i\right] = 0 \qquad \text{et}$$

$$\frac{dj}{db_s} = -2.E\left[X_s.\left(E[m(q)] - a - \sum_{i=1}^{n} b_i X_i\right)\right] = 0$$

Pour un seul contrat, les paramètres structuraux posons ainsi :

$$\begin{aligned} m(q) &= E(X_{n+1}/q) \\ m &= E(X_i) \\ a &= V(E(X_i/q)) \\ s^2 &= E(V(X_i/q)) \end{aligned}$$

On a alors le système suivant :

$$\begin{cases} a = m\left(1-\sum_{i=1}^{n} b_i\right) \\ E(X_s.E[m(q)]) - a.E(X_s) - \sum_{i=1}^{n} b_i E(X_s.X_i) = 0 \end{cases}$$

$$\begin{cases} a = m\left(1-\sum_{i=1}^{n} b_i\right) \\ COV(X_s.E[m(q)]) + E(X_s.E[m(q)]) - a.E(X_s) - \sum_{i=1}^{n} b_i \left(COV(X_s, X_i) + E(X_i).E(X_s)\right) = 0 \end{cases}$$

$$\begin{cases} a = m\left(1-\sum_{i=1}^{n} b_i\right) \\ COV(X_s, E[m(q)]) + m^2 - a.m - \sum_{i=1}^{n} b_i \left(COV(X_s, X_i) + m^2\right) = 0 \end{cases}$$

$$\begin{cases} a = m\left(1-\sum_{i=1}^{n} b_i\right) \\ COV(X_s, E[m(q)]) - \sum_{i=1}^{n} b_i \cdot COV(X_s, X_i)) = 0 \end{cases}$$

Tout d'abord, développons le terme $COV(X_s, X_i)$:

$$\begin{aligned} COV(X_s, X_i) &= E(COV(X_i, X_s/q)) + COV(E(X_i/q), E(X_s/q)) \\ &= d_{s,t} \cdot E(V(X_i/q)) + V(E(X_i/q)) \\ &= d_{s,t} \cdot s^2 + a \end{aligned}$$

$$\text{Avec } d_{s,t} \begin{cases} 1 & si \quad s = t \\ 0 & \quad\quad s \neq t \end{cases}$$

Par ailleurs :

$$\begin{aligned} COV(X_t, m(q)) &= E(COV(X_t, m(q)/q)) + COV(E(X_t/q), E(m(q)/q)) \\ &= V(E(X_t/q)) \\ &= a \end{aligned}$$

En remplaçant on obtient :

$$a = \frac{s^2}{s^2 + na} \cdot m = (1 - b) \cdot m$$
$$b = \frac{n.a}{n.a + s^2}$$

$$m(q) = \frac{s^2}{s^2 + na} \cdot m + (1 - \frac{s^2}{s^2 + na}) \cdot \overline{X}$$

Nous obtenons bien ainsi les résultats de BÜHLMANN.

Pour faire le lien avec le paragraphe précédent Partie 2 1.2, posons :

$\hat{b} = \hat{z}$
$\overline{X} = \hat{X}_1$ l'estimateur basé sur l'information individuelle (exogène)
$m = \hat{X}_2$ l'estimateur basé sur l'ensemble de l'information disponible

$\hat{X}_j = \hat{z}.\hat{X}_2 + (1 - \hat{z}).\hat{X}_1$

si $s \to \infty$ $\hat{b} \to 0$ l'information du groupe sera la meilleure car la variabilité intra-classe $E(V(X_i / q))$ est élevée. Cela traduit une faible homogénéité à l'intérieur de la classe de risque considérée qui ne permet pas de la différencier d'une autre.

si $n \to \infty$ $\hat{b} \to 1$ l'information individuelle sera la meilleure. Plus la taille de la population étudiée est grande, plus il est possible de converger vers la « vraie » valeur de la classe de risque considérée et ainsi de la différencier d'une autre

si $a \to \infty$ $\hat{b} \to 1$ l'information individuelle sera la meilleure car le niveau du terme $COV(\hat{X}_1, \hat{X}_2) = a$ traduit la qualité de la relation entre la vraie valeur de la classe de risque et son estimateur individuel.

Dans ce paragraphe, nous avons démontré que le modèle de régression et la théorie de la crédibilité se fondent sur les mêmes principes. Ce résultat sera repris pour montrer que la méthode Munich Chain Ladder est une crédibilisation car son modèle repose sur un modèle de régression.

Il existe aussi d'autres références sur les techniques de crédibilité en provisionnement comme par exemple le modèle de Witting qui est un modèle sur le nombre de sinistres. Nous ne les détaillons pas dans ce mémoire.

2 La crédibilité linéaire dans des méthodes dites classiques : Bornhuetter-Ferguson / Cape Cod et Benktander

A la fin de la Partie 1, lors de la synthèse des analyses, il a été constaté que les méthodes Bornhuetter-Ferguson / Cape Cod et Benktander fournissaient des résultats beaucoup plus satisfaisants sur les années faiblement développées ou intermédiaires. Celles-ci sont reprises ci après et reformulées en reprenant les résultats obtenus dans la Partie 2. L'idée est ici de mettre en évidence qu'elles font appel implicitement aux principes de la théorie de la crédibilité.

2.1 La crédibilité dans les méthodes Bornhuetter-Ferguson / Cape Cod

Les variables aléatoires $C_{i,j}$ représentent les sinistres payés ou les sinistres survenus en année d'origine i pour l'année de développement j.

Analysons les composants des méthodes Bornhuetter-Ferguson et par extension Cape Cod :

L'espérance mathématique $E(C_{i,N})$ dans cet univers peut être déterminée à l'aide de différents estimateurs $\hat{X}^{BF}$.

Soit le premier estimateur, un estimateur obtenu de façon exogène $\hat{X}_1^{BF} = \hat{X}_1^{LR} = L^{LR} \cdot A = \hat{C}_{i,N}$ avec L^{LR} le Loss ratio et A l'aliment. Ici afin de simplifier nos notations, nous prenons comme hypothèse que les aliments de chaque année d'origine sont identiques et que le L^{LR} aussi (H5 dans la Partie I).

En première partie, il a été vu qu'il s'agissait d'une estimation ex ante des sinistres, c'est-à-dire le montant de sinistre ultime obtenu grâce à la méthode du Loss Ratio. Aussi, c'est l'une des seules méthodes permettant d'obtenir une bonne estimation du sinistre ultime sur les années faiblement développées (cf. Partie 1 4).

Si l'estimation du L^{LR} est juste, cet estimateur est sans biais. C'est une des hypothèses du modèle de Bornhuetter Ferguson.

Soit le second estimateur $\hat{X}_2^{BF} = \hat{X}_2^{CL} = C_{i,j} \cdot F_{i,j}$, avec $F_{i,j}$ le facteur de développement pour passer de l'année de développement j à N, obtenu à l'aide de la méthode Chain Ladder.

$$F_{i,j} = \frac{E(C_{i,N})}{E(C_{i,j})}$$

$\hat{X}_2^{CL}$ représente le montant de sinistre ultime obtenu grâce à la méthode Chain Ladder (SCL) Comme cela a été vu en Partie I, l'estimation du sinistre ultime sur les années intermédiaires ou fortement développées est assez bonne ou bonne en utilisant cette méthode.

Si l'on construit un méta estimateur $\hat{X}_3 = z \cdot \hat{X}_2^{CL} + (1-z) \cdot \hat{X}_1^{LR}$

En utilisant les résultats obtenus dans le paragraphe sur la crédibilité générale Partie 2 (1.2), montrons que pour un z* optimal en terme d'écart par rapport à l'ultime, ce méta estimateur correspond à l'estimation de sinistre ultime par la méthode de Bornhuetter-Ferguson et par extension de Cap-Code (cf. Partie I).

Reprenons la technique de moindres carrés ordinaires et appliquons la au « méta estimateur » $\hat{X}_3$ afin de déterminer l'estimateur optimal z*.

Pour cela, nous devons faire l'hypothèse suivante[3] :

$$V(\hat{X}_2^{CL}) = V(C_{i,j}.F_{i,j}) = F_{i,j}.V(C_{i,N}) = F_{i,j}\, n^2 \text{ avec } V(C_{i,N}) = n^2$$

En remplaçant, $\hat{X}_3$ peut encore s'écrire :

$$\hat{X}_3 = z \cdot C_{i,j} \cdot F_{i,j} + (1-z) \cdot \hat{C}_{i,N}$$

$\hat{X}_3$ est un estimateur sans biais.

Posons j tel que :

$$j = E\left[\left(C_{i,N} - \hat{X}_3\right)^2\right] = E\left[\left(C_{i,N} - \left(z \cdot C_{i,j} \cdot F_{i,j} + (1-z) \cdot \hat{C}_{i,N}\right)\right)^2\right]$$

Or, compte tenu des hypothèses faites sur le caractère sans biais des estimateurs on peut remplacer dans l'équation j :

$$V\left(z \cdot \left(C_{i,N} - C_{i,j}.F_{i,j}\right)\right) = z^2 \cdot E\left[\left(C_{i,N} - C_{i,j}.F_{i,j}\right)^2\right]$$

et

$$V\left((1-z) \cdot \left(C_{i,N} - \hat{C}_{i,N}\right)\right) = (1-z)^2 \cdot E\left[\left(C_{i,N} - \hat{C}_{i,N}\right)^2\right]$$

et

$$COV\left(\left(C_{i,N} - C_{i,j}.F_{i,j}\right), \left(C_{i,N} - \hat{C}_{i,N}\right)\right) = E\left(\left(C_{i,N} - C_{i,j}.F_{i,j}\right) \cdot \left(C_{i,N} - \hat{C}_{i,N}\right)\right)$$

[3] Voir Patrick DAHL, Introduction to reserving, corrected edition, 2003

Afin de simplifier l'équation, on remarque que :

$$\begin{aligned} COV\left(\left(C_{i,N} - C_{i,j}.F_{i,j}\right), \left(C_{i,N} - \hat{C}_{i,N}\right)\right) &= COV\left(C_{i,N}, C_{i,N}\right) - COV\left(C_{i,j} \cdot F_{i,j}, C_{i,N}\right) \\ &= V\left(C_{i,N}\right) - F_{i,j} \cdot COV\left(C_{i,j}, \left(C_{i,N} - C_{i,j}\right) + C_{i,j}\right) \\ &= n^2 - 0 - F_{i,j} \cdot V\left(C_{i,j}\right) \\ &= n^2 - n^2 = 0 \end{aligned}$$

Avec $COV\left(C_{i,j}, \left(C_{i,N} - C_{i,j}\right)\right) = 0$ car les paiements non effectués sont indépendants de ceux déjà effectués :

$$C_{i,j} \perp C_{i,N} - C_{i,j} \text{ (H6 dans la Partie I)}$$

(cf. en ANNEXES PARTIE 2 [3.1] les démonstrations intérimaires)

Ainsi,

$$j = V\left(z \cdot \left(C_{i,N} - C_{i,j}.F_{i,j}\right)\right) + V\left((1-z).\left(C_{i,N} - \hat{C}_{i,N}\right)\right)$$

On peut encore écrire j sous forme d'équation quadratique :

$$j = z^2 \cdot \left[V\left(\left(C_{i,N} - C_{i,j}.F_{i,j}\right)\right) + V\left(\left(C_{i,N} - \hat{C}_{i,N}\right)\right)\right] - 2z \cdot V\left(\left(C_{i,N} - \hat{C}_{i,N}\right)\right) + V\left(\left(C_{i,N} - \hat{C}_{i,N}\right)\right)$$

Le premier terme de l'équation étant toujours positif, j est une fonction convexe. On retrouve le résultat de l'inéquation (a). L'optimum est donc le point qui minimise la fonction.

En dérivant j par rapport à z :

$$\frac{dj}{dz} = 2 \cdot \left(z.V\left(C_{i,N} - C_{i,j}.F_{i,j}\right) - (1-z).V\left(C_{i,N} - \hat{C}_{i,N}\right)\right)$$

A l'optimum, $\frac{dj}{dz} = 0$

$$\Leftrightarrow z^*.V\left(C_{i,N} - C_{i,j}.F_{i,j}\right) - (1-z^*).V\left(C_{i,N} - \hat{C}_{i,N}\right) = 0$$

Or :

$$\begin{aligned} V\left(C_{i,N} - C_{i,j}.F_{i,j}\right) &= V\left(C_{i,N}\right) + V\left(C_{i,j}.F_{i,j}\right) - 2.COV\left(C_{i,N}, C_{i,j}.F_{i,j}\right) \\ &= n^2 + F_{i,j}^2 V\left(C_{i,j}\right) - 2n^2 \\ &= n^2 + F_{i,j}^2 . \frac{n^2}{F_{i,j}} - 2n^2 \\ &= n^2 . \left(F_{i,j} - 1\right) \end{aligned}$$

de plus $V\left(C_{i,N} - \hat{C}_{i,N}\right) = V\left(C_{i,N}\right) = n^2$

On en déduit en remplaçant que :

$$\frac{dj}{dz} = 0 \Leftrightarrow z^*.V\left(C_{i,N} - C_{i,j}.F_{i,j}\right) - (1 - z^*).V\left(C_{i,N} - \hat{C}_{i,N}\right) = 0$$

$$\frac{dj}{dz} = 0 \Leftrightarrow z^* n^2 . \left(F_{i,j} - 1\right) - n^2 + z^* n^2 = 0$$

$$\frac{dj}{dz} = 0 \Leftrightarrow z^* = \frac{1}{F_{i,j}}$$

Ainsi pour un z* optimal, le méta estimateur est $\hat{X}_3 = \frac{1}{F_{i,j}} \cdot \hat{X}_2^{CL} + \left(1 - \frac{1}{F_{i,j}}\right) \cdot \hat{X}_1^{LR}$

Sous les hypothèses standards de Mack de la méthode SCL, on obtient le montant de sinistre ultime estimé selon la méthode de Bornhuetter-Ferguson et par extension de Cap-Code (cf. Partie I). Ces méthodes sont donc des crédibilisations d'autres méthodes. Plus une année d'origine est jeune, plus le poids de l'estimateur du Loss Ratio sera fort. A l'inverse, plus une année d'origine sera développée, plus le poids de l'estimateur de la méthode de Chain Ladder sera fort.

Ce qui est intéressant de noter est qu'ici le méta estimateur obtenu dépend d'un z qui est propre pour chaque année d'origine. Le méta estimateur n'applique donc pas le même poids aux deux estimateurs $\hat{X}_1^{BF}$ et $\hat{X}_2^{BF}$ suivant les années d'origine. Plus l'année d'origine est faiblement développée, plus le poids de $\hat{X}_1^{BF}$ est fort et celui de $\hat{X}_2^{BF}$ est faible et vice et versa.

Prenons un exemple illustratif :

Soient i les années d'origine pour i=1,…,5

i	z	LR	Chain Ladder	Bornhuetter-Ferguson
1	0%	100	0	100
2	25%	100	20	80
3	50%	100	40	70
4	75%	100	80	85
5	90%	100	110	109

Cet exemple montre que les z sont différents pour chaque année d'origine. Ainsi, pour les années 1 et 2 le poids affecté au sinistre ultime issu du Loss Ratio est plus fort que celui affecté au sinistre ultime obtenu à l'aide de la méthode de Chain Ladder. Pour les années 4 et 5, c'est l'inverse.

L'explication économique de cette observation est que lorsqu'une année est faiblement développée, surtout sur une branche longue, il n'y a pas beaucoup d'information disponible. Dans ce cas, la méthode de Bornhuetter-Ferguson propose d'utiliser plus fortement une estimation ex ante des sinistres. Ceci permet de pallier à un des défauts de la méthode de Chain Ladder qui sous estime les sinistres sur ces années.
Puis, plus les années d'origine se développent, plus la méthode de Bornhuetter-Ferguson permet de donner du poids au développement constaté des années d'origine, en diminuant le poids de l'estimation ex ante des sinistres.

Cette méthode utilise donc les résultats constatés à la fin de la Partie I quant aux avantages et inconvénients des méthodes SCL et du loss ratio.

Cette méthode reprend donc bien l'idée de la théorie de la crédibilité : à l'instant *t* c'est l'information ou l'estimateur le plus pertinent qui sera utilisé.

2.2 La crédibilité dans la méthode Benktander

De la même façon que dans le paragraphe précédent, analysons les composants de la méthode Benktander, et utilisons un programme d'optimisation afin de montrer qu'il s'agit d'une crédibilisation de plusieurs méthodes de provisionnement.

L'espérance mathématique de $C_{i,N}$ dans cet univers $E(C_{i,N})$ peut être déterminée à l'aide de différents estimateurs $\hat{X}^{BK}$.

Soit le premier estimateur, le second estimateur de la méthode Bornhuetter-Ferguson $\hat{X}_1^{BK} = \hat{X}_1^{BF} = C_{i,j} \cdot F_{i,j} = \hat{X}^{CL}$.

Soit le second estimateur, le méta estimateur de la méthode Bornhuetter-Ferguson $\hat{X}_2^{BK} = \hat{X}_3^{BF} = \frac{1}{F_{i,j}} \cdot \hat{X}^{CL} + (1 - \frac{1}{F_{i,j}}) \cdot \hat{X}^{LR}$.

On construit un nouveau méta estimateur :
$\hat{X}_3^{BK} = z \cdot \hat{X}_1^{BK} + (1-z) \cdot \hat{X}_2^{BK} = z \cdot \hat{X}_2^{BF} + (1-z) \cdot \hat{X}_3^{BF}$.

Montrons que pour un z*, ce méta estimateur correspond à l'estimation de sinistre ultime par la méthode de Benktander.

$$\hat{X}_4 = \hat{X}_3^{BK} = z \cdot \hat{X}_2^{CL} + (1-z) \cdot \left(\frac{1}{F_{i,j}} \hat{X}_2^{CL} + \left(1 - \frac{1}{F_{i,j}} \right) \cdot X_1^{LR} \right)$$

En remplaçant, $\hat{X}_4$ peut encore s'écrire :

$$\hat{X}_4 = z \cdot C_{i,j} \cdot F_{i,j} + (1-z) \cdot \left(\frac{1}{F_{i,j}} \cdot C_{i,j} \cdot F_{i,j} + \left(1 - \frac{1}{F_{i,j}} \right) \cdot \hat{C}_{i,N} \right) \qquad \text{avec } F_{i,j} \text{ donné}$$

Reprenons la technique des moindres carrés ordinaires et appliquons la au « méta estimateur » $\hat{X}_4$ afin de déterminer l'estimateur optimal z*.

$\hat{X}_4$ est un estimateur sans biais.

Posons j tel que :

$$j = E\left[\left(C_{i,N} - \hat{X}_4\right)^2\right] = E\left[\left(C_{i,N} - \left(z \cdot C_{i,j} \cdot F_{i,j} + (1-z) \cdot \left(\frac{1}{F_{i,j}} \cdot C_{i,j} \cdot F_{i,j} + \left(1 - \frac{1}{F_{i,j}}\right) \cdot \hat{C}_{iN}\right)\right)\right)^2\right]$$

$$j = E\left[\left(z \cdot \left(C_{i,N} - C_{i,j} \cdot F_{i,j}\right) + (1-z) \cdot \left(C_{i,N} - \left(\frac{1}{F_{i,j}} \cdot C_{i,j} \cdot F_{i,j} + \left(1 - \frac{1}{F_{i,j}}\right) \cdot \hat{C}_{i,N}\right)\right)\right)^2\right]$$

Posons $\hat{D}_{i,j} = \left(\frac{1}{F_{i,j}} \cdot C_{i,j} \cdot F_{i,j} + \left(1 - \frac{1}{F_{i,j}}\right) \cdot \hat{C}_{i,N}\right)$

Or, compte tenu des hypothèses faites sur le caractère sans biais des estimateurs on peut remplacer dans l'équation j :

$$V\left(z \cdot \left(C_{i,N} - C_{i,j}.F_{i,j}\right)\right) = z^2 \cdot E\left[\left(C_{i,N} - C_{i,j}.F_{i,j}\right)^2\right]$$

et

$$V\left((1-z) \cdot \left(C_{i,N} - \hat{D}_{i,j}\right)\right) = (1-z)^2 \cdot E\left[\left(C_{i,N} - \hat{D}_{i,j}\right)^2\right]$$

et

$$COV\left(\left(C_{i,N} - C_{i,j}.F_{i,j}\right), \left(C_{i,N} - \hat{D}_{i,j}\right)\right) = E\left(\left(C_{i,N} - C_{i,j}.F_{i,j}\right) \cdot \left(C_{i,N} - \hat{D}_{i,j}\right)\right) = 0$$

cf. la démonstration en ANNEXES PARTIE 2 [3.2]

$$j = V\left(z \cdot \left(C_{i,N} - C_{i,j}.F_{i,j}\right)\right) + V\left((1-z).\left(C_{i,N} - D_{i,j}\right)\right)$$

donc $$j = z^2 \cdot V\left(\left(C_{i,N} - C_{i,j}.F_{i,j}\right)\right) + (z-1)^2 \cdot V\left(\left(C_{i,N} - D_{i,j}\right)\right)$$

$$j = z^2 \cdot V\left(\left(C_{i,N} - C_{i,j}.F_{i,j}\right) + V\left(\left(C_{i,N} - D_{i,j}\right)\right)\right) + (1-2z) \cdot V\left(\left(C_{i,N} - D_{i,j}\right)\right)$$

Le premier terme de l'équation étant toujours positif, j est une fonction convexe. On retrouve le résultat de l'inéquation (a). L'optimum est donc le point qui minimise cette fonction.

En dérivant j par rapport à z :

$$\frac{dj}{dz} = 2 \cdot \left(z.V\left(C_{i,N} - C_{i,j}.F_{i,j}\right) + (z-1).V\left(C_{i,N} - \hat{D}_{i,j}\right)\right)$$

A l'optimum, $\frac{dj}{dz} = 0$

$$\Leftrightarrow z^*.V\left(C_{i,N} - C_{i,j}.F_{i,j}\right) + (z^* - 1).V\left(C_{i,N} - \hat{D}_{i,j}\right) = 0$$

$$\Leftrightarrow z^* = \frac{V\left(C_{i,N} - \hat{D}_{i,j}\right)}{V\left(C_{i,N} - C_{i,j}.F_{i,j}\right) + V\left(C_{i,N} - \hat{D}_{i,j}\right)}$$

$$\Leftrightarrow z^* = \frac{1}{\frac{V\left(C_{i,N} - C_{i,j}.F_{i,j}\right)}{V\left(C_{i,N} - \hat{D}_{i,j}\right)} + 1}$$

En reprenant les résultats de la démonstration de la méthode de Bornhuetter-Ferguson du paragraphe précédent :

$$V\left(C_{i,N} - C_{i,j}.F_{i,j}\right) = n^2.\left(F_{i,j} - 1\right)$$

$$V\left(C_{i,N} - \hat{D}_{i,j}\right) = \frac{n^2}{F_{i,j}}.\left(F_{i,j} - 1\right)$$ cf. la démonstration en ANNEXES PARTIE 2 [3.2]

En remplaçant :

$$\Leftrightarrow z^* = \frac{1}{\frac{n^2.\left(F_{i,j} - 1\right)}{\frac{n^2}{F_{i,j}}.\left(F_{i,j} - 1\right)} + 1}$$

$$z^* = \frac{1}{1 + F_{i,j}}$$

On remarque que le z* ne permet pas de retrouver la formule de la méthode de Benktander.

Le z de la méthode de Benktander est tel que :

$$z^{BK} = \frac{1}{F_{i,j}}$$

Or si l'on se réfère aux articles de Hülimann Werner (Bibliographie [6]) et de Mack (Bibliographie [6]), ils montrent qu'effectivement cette méthode n'utilise pas le coefficient de crédibilité optimal, mais qu'il s'en rapproche. Cette méthode est une crédibilisation de deux méthodes même si son coefficient de crédibilité n'est pas celui de l'optimum.

$$0 \le z^{BK} \le z^* \le 1$$

La fonction j étant convexe, le coefficient de crédibilité de la méthode de Benktander permet tout de même d'améliorer l'estimation du sinistre ultime par rapport aux estimations prises individuellement par les méthodes Chain Ladder ou Bornhuetter-Ferguson.

Plusieurs éléments sont à noter :

- Cette démonstration s'inscrit dans le cadre des hypothèses standards de Mack de la méthode Chain Ladder auxquelles s'ajoute une hypothèse sur $V\left(C_{i,j}.F_{i,j}\right)$. On remarque qu'il ne s'agit pas des mêmes hypothèses que dans l'article de Mack sur la méthode de Benktander, c'est pourquoi on obtient un z* sensiblement différent.
- Comme pour le méta estimateur de la méthode Bornhuetter-Ferguson, le méta estimateur obtenu dépend d'un z qui est propre pour chaque année d'origine.
- Il est tout à fait possible de créer un méta estimateur à partir d'un autre.

Ci-dessous voici un exemple illustratif de la crédibilisation linéaire obtenue par la méthode de Benktander :

i	z	LR	Chain Ladder	Bornhuetter-Ferguson	Benktander
1	0%	100	0	100	100
2	25%	100	20	80	65
3	50%	100	40	70	55
4	75%	100	80	85	81
5	90%	100	110	109	110

Les conclusions de la méthode du Bornhuetter-Ferguson peuvent être reprises pour la méthode de Benktander.

3 Une autre forme de crédibilisation de l'information disponible : le Munich Chain Ladder

Contrairement aux méthodes Bornhuetter-Ferguson / Cape Cod et Benktander, la méthode du Munich Chain Ladder (MCL) n'a pas été présentée dans la Partie I car nous ne la classons pas dans les méthodes dites « Classiques ».

Dans cette partie, la présentation de la méthode du Munich Chain Ladder ne sera pas décrite suivant le plan de l'article de Quarg G et Mack T (Bibliographie [15]) mais de façon à faire le lien entre la méthode et la crédibilité linéaire. C'est pourquoi plusieurs études préliminaires sont nécessaires afin de comprendre quels sont les paramètres utilisés par la méthode.

3.1 Notations

Soient :

$P_i = \left(P_{i,j}\right)_{j \in T}$ représente le processus des sinistres payés pour l'année d'origine i.

$\mathrm{P}_i(s) = \left\{P_{i,1}, \mathbf{L}, P_{i,s}\right\}$ est l'ensemble des montants de sinistres payés sous contrainte de connaître le processus de développement des payés jusqu'à l'année s.

$I_i = \left(I_{i,j}\right)_{j \in T}$ représente le processus des sinistres survenus pour l'année d'origine i.

$\mathrm{I}_i(s) = \left\{I_{i,1}, \mathbf{L}, I_{i,s}\right\}$ est l'ensemble des sinistres survenus sous contrainte de connaître le processus de développement des sinistres survenus jusqu'à l'année s.

a_i est l'indice de développement représentant la date actuelle pour l'année d'origine i.

$\left(P/I\right)_{i,j} = \dfrac{P_{i,j}}{I_{i,j}} = R_{i,j}^{MCL}$ représente le ratio des sinistres payés cumulés pour l'année d'origine i à la date t quelconque de développement sur les sinistres survenus cumulés pour l'année d'origine i à la date t quelconque de développement.

$(P/I)_j = \frac{\sum_{i=1}^{n} P_{i,j}}{\sum_{i=1}^{n} I_{i,j}}$ représente le ratio des sinistres payés cumulés pour toutes les années d'origine à la date j de développement sur les sinistres survenus cumulés pour toutes les années d'origine à la date j de développement.

$\mathrm{B}_i(s) = \{P_{i,1}, \mathbf{L}, P_{i,s}, I_{i,1}, \mathbf{L}, I_{i,s}\}$ est l'ensemble des sinistres payés et survenus cumulés connus.

f_j^I est le coefficient de développement des sinistres survenus de j-1 à j.

f_j^P est le coefficient de développement des sinistres payés de j-1 à j.

$OS_{i,j}$ est le montant de sinistres en suspens de l'année d'origine i au développement j.

3.2 Présentation synthétique du principe de la méthode

La méthode de provisionnement du Munich Chain Ladder (MCL) est une méthode de projection de sinistres qui utilise un ensemble d'informations B_i (l'ensemble des sinistres payés et survenus cumulés connus) dont elle dispose.

Contrairement aux méthodes « traditionnelles », le sinistre ultime n'est pas obtenu en projetant soit le triangle de sinistres payés soit celui des sinistres survenus. En effet, les résultats issus de ces deux projections ont montré qu'ils sont la plupart du temps sensiblement différents.

Le sinistre ultime du MCL est déterminé par une combinaison de projections de ces deux mesures.

$$\forall i \quad k \in \{a_{i+1}, \mathbf{K}, N+i+1\} \quad P_{i,k} = f(I_{i,j}, P_{i,j} / j \in \{1, \mathbf{K}, k-1\}) \quad et \quad I_{i,k} = g(P_{i,j}, I_{i,j} / j \in \{1, \mathbf{K}, k-1\})$$

3.3 Etude préliminaire : Evolution des sinistres survenus, payés et en suspens ainsi que leurs influences sur les ratios $R_{i,j}^{MCL}$ et $\left(R_{i,j}^{MCL}\right)^{-1}$

A. Evolution des sinistres survenus, payés et en suspens

Afin de comprendre pourquoi la projection d'un triangle de sinistres payés donne des résultats sensiblement différents d'une projection de sinistres survenus, il est nécessaire d'étudier leurs évolutions et leurs interactions.

Dans un second temps l'évolution des ratios $R_{i,j}^{MCL}$ et $\left(R_{i,j}^{MCL}\right)^{-1}$ sera présentée, pour cela posons :

$$\left(R_{i,j}^{MCL}\right)^{-1} = I_{i,j} / P_{i,j} = 1 + OS_{i,j} / P_{i,j} \quad (i)$$
$$R_{i,j}^{MCL} = P_{i,j} / I_{i,j} = 1 - OS_{i,j} / I_{i,j} \quad (ii)$$

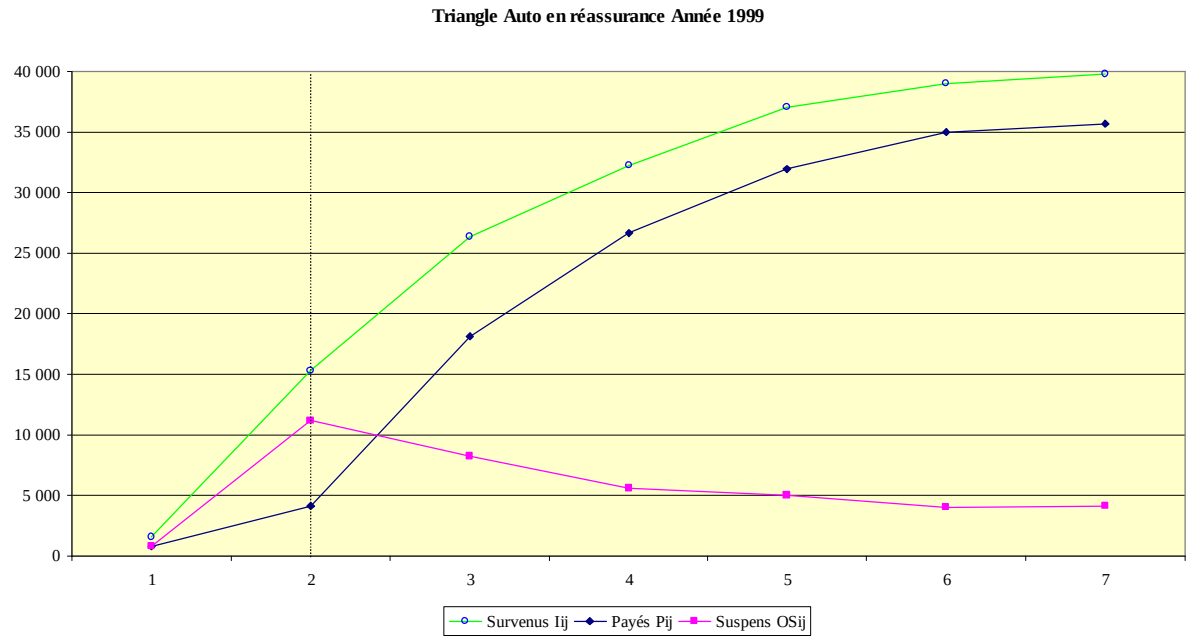
Chacune des variables présentes dans les deux équations précédentes ont leur propre évolution pour une année d'origine i donnée, tout au long de son développement j.

Lorsqu'un sinistre survient, le montant de sinistre survenu $I_{i,j}$ se développe et est rapidement stabilisé.

Les sinistres payés $P_{i,j}$ croissent généralement moins rapidement que les sinistres survenus $I_{i,j}$ sur au moins la première année de développement. Puis ils progressent « tout au long du développement du triangle » plus ou moins rapidement.

Les suspens $OS_{i,j}$ quant à eux, croissent dans un premier temps, en parallèle des sinistres survenus. Dans un univers idéal, ils décroissent une fois le sinistre survenu stabilisé dans une proportion de l'augmentation des payés $P_{i,j}$. Leur développement est généralement en forme de cloche.
Ce n'est pas le cas lorsqu'il y a des évolutions en termes de procès, par exemple un sinistre en appel dont la décision de justice change en faveur de l'assuré ou s'il y a une dégradation du sinistre par exemple en RC automobile l'évolution des coûts médicaux.

L'évolution de l'année de souscription 1999 du triangle Automobile en réassurance permet d'illustrer ces tendances :



Dans cet exemple, l'année de développement 2 est le point culminant de la « cloche » des suspens. A partir de cette date, ils décroissent et les payés croissent très fortement.

En termes de provisionnement, la projection du triangle de sinistres payés a tendance à fournir des coefficients de développement plus forts que pour une étude sur les sinistres survenus. Le sinistre ultime sera tout de même sous-évalué lorsque l'année d'origine est faiblement développée du fait de la faible assiette de sinistres payés. En effet, dans ce cas, elle prend en compte un montant d'émis faible et n'utilise pas l'information que fournit la part des sinistres en suspens.

A contrario, la projection des sinistres survenus a tendance à surestimer le sinistre ultime. En particulier lorsque le développement des sinistres est bien avancé et que les suspens décroissent ou sont faibles, le résultat de la projection surestime les provisions.

Ces deux dernières remarques expliquent l'importance des sinistres en suspens. Ils ne sont généralement pas utilisés en tant qu'information décisive dans les méthodes de provisionnement classiques. Ils restent un critère qualitatif. Parfois les IBNR sont un pourcentage des sinistres en suspens.

Aussi, les développements des variables $P_{i,j}$, $I_{i,j}$, $OS_{i,j}$ précédemment décrits montrent qu'il est judicieux de faire bénéficier aux estimateurs d'une première méthode de provisionnement de l'information contenue dans la seconde méthode. Pour cela la méthode de MCL utilise les ratios $R_{i,j}^{MCL}$ et $\left(R_{i,j}^{MCL}\right)^{-1}$.

B. Les ratios $R_{i,j}^{MCL}$ et $\left(R_{i,j}^{MCL}\right)^{-1}$

Des différentes progressions expliquées dans le paragraphe précédent, peuvent être déduits les comportements des ratios $R_{i,j}^{MCL}$ et $\left(R_{i,j}^{MCL}\right)^{-1}$.

Les seconds termes des équations (i) et (ii) permettent de situer le développement du sinistre et donc d'évaluer si le montant de provision est surestimé ou sous-estimé :

Dans l'équation (i) $\left(R_{i,j}^{MCL}\right)^{-1} = I_{i,j}/P_{i,j} = 1 + OS_{i,j}/P_{i,j}$:

Le tableau ci-dessous récapitule l'évolution du ratio $\left(R_{i,j}^{MCL}\right)^{-1}$:

Valeur des sinistres Payés	**Valeur des sinistres en Suspens**	**Valeur du ratio $\left(R_{i,j}^{MCL}\right)^{-1}$**
0	**Sinistres survenus**	**Indéfini**
1 unité	**Sinistres survenus – 1 unité**	**Suspens**
< suspens	**>payés**	**]2 ; Suspens[**
= suspens	**= payés**	**2**
> suspens	**<payés**	**[1 ; 2[**

Ainsi $\left(R_{i,j}^{MCL}\right)^{-1}$ est une valeur comprise entre 1 et les suspens si l'on prend comme hypothèse que les montants des sinistres payés ne peuvent être inférieurs à une unité.

La première partie de la courbe est une courbe croissante car le sinistre survenu augmente ainsi que les suspens. A ce niveau de développement du sinistre, $\left(R_{i,j}^{MCL}\right)^{-1}$ croît.
Lorsque $\left(R_{i,j}^{MCL}\right)^{-1}$ est compris entre 1 et 2 (dernier cas dans le tableau récapitulatif $OS_{i,j} \leq P_{i,j}$), généralement le sinistre survenu commence à se stabiliser.

Le ratio $\left(R_{i,j}^{MCL}\right)^{-1}$ donne de l'information à la projection du triangle des sinistres payés.

Dans l'équation (ii) $R_{i,j}^{MCL} = P_{i,j}/I_{i,j} = 1 - OS_{i,j}/I_{i,j}$:

Le second terme de l'équation (ii) est la part des suspens dans le sinistre survenu. Lorsque ce dernier se développe, le suspens le fait dans une même proportion. Le ratio suspens sur sinistre survenu est constant et $R_{i,j}^{MCL}$ est proche de 0.

Une fois le sinistre survenu stabilisé, les suspens vont diminuer au profit des payés et donc le ratio suspens sur sinistre survenu diminue.
La courbe des $R_{i,j}^{MCL}$ a une progression logarithmique. La valeur du $R_{i,j}^{MCL}$ est comprise entre 0 et 1.

Le ratio $R_{i,j}^{MCL}$ donne de l'information à la projection du triangle des sinistres survenus.

$R_{i,j}^{MCL}$ représente aussi le pourcentage de payés par rapport aux sinistres survenus.

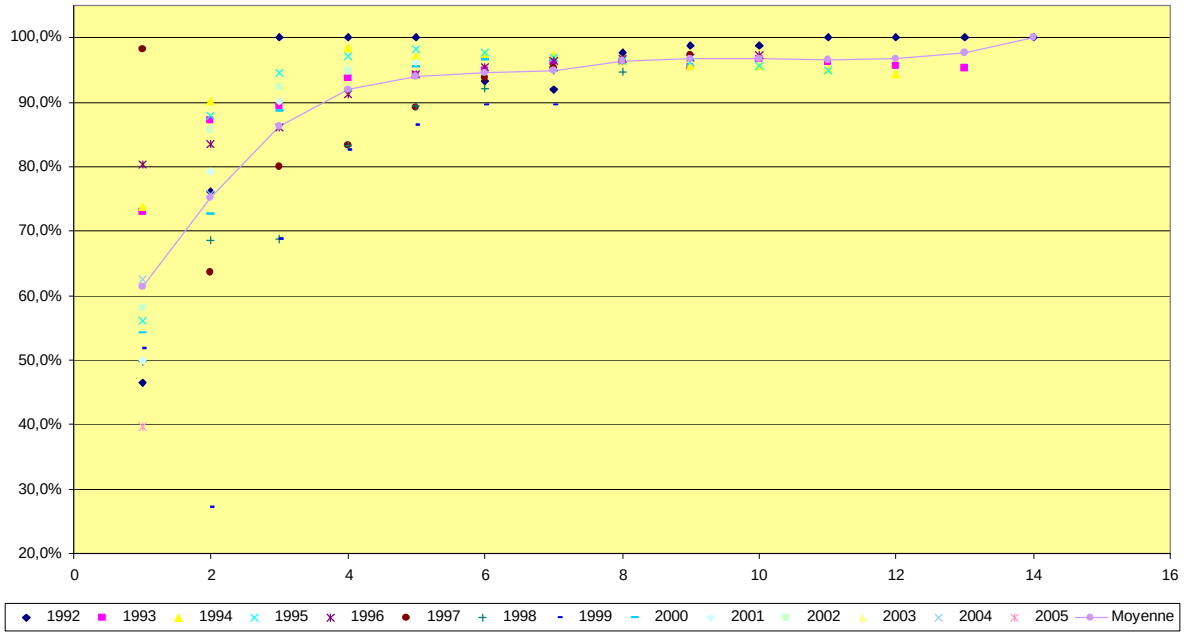
Synthèse de la signification des ratios $R_{i,j}^{MCL}$ et $\left(R_{i,j}^{MCL}\right)^{-1}$ dans le développement du sinistre :

Valeur des ratios	**Développement du sinistre**	
	$\left(R_{i,j}^{MCL}\right)^{-1}$	$R_{i,j}^{MCL}$
0	f	**Pas développé**
]0,1[	f	**En cours de développement**
1	**Terminé**	**Terminé**
]1,2]	**Développement avancé**	f
]2,+∞[	**En cours de développement**	f

3.4 La corrélation entre la diagonale et la sur diagonale d'un triangle

Les ratios $R_{i,j}^{MCL}$ et $\left(R_{i,j}^{MCL}\right)^{-1}$ permettent aussi de matérialiser la corrélation qui existe dans un triangle entre sa diagonale et sa sur diagonale.

Evolution des $R_{i,j}^{MCL}$ pour $j \leq a_{i-1}$ sur le triangle de réassurance :

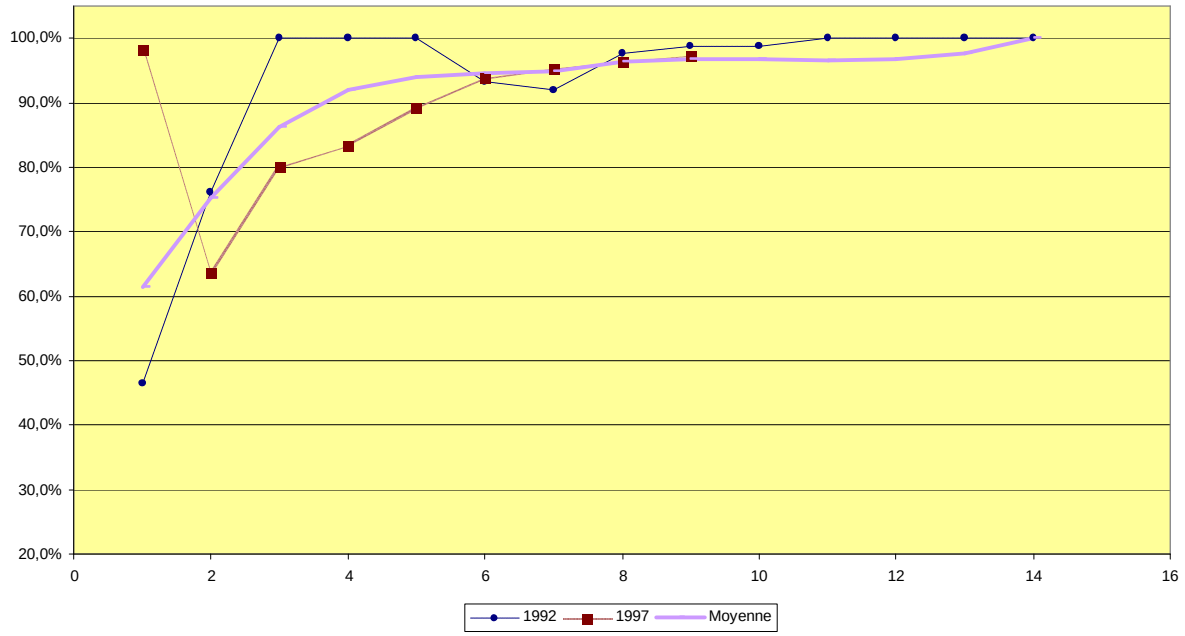


Graph 1.

Ce graphique représente les $R_{i,j}^{MCL}$ émis en fonction des années de développement, c'est-à-dire : $R_{i,j}^{MCL} = P_{i,j} / I_{i,j} \qquad \forall j \in 1,...,a_{i-1}$

Il montre que les $R_{i,j}^{MCL}$ sur des années fortement développées tendent vers 100%.
Aussi, plus une année d'origine est développée, plus les $R_{i,j}^{MCL}$ se rapprochent du $R_{i,j}^{MCL}$ moyen (moyenne des $R_{i,j}^{MCL}$ pour un j fixé).

Zoomons sur les années 1992 et 1997 afin d'illustrer les variations par rapport à la moyenne :



Graph 2.

Ce graphique permet de remarquer que si une année d'origine a un ratio $R_{i,j}^{MCL}$ relativement faible par rapport à l'historique et à la moyenne (graphiquement il sera en dessous de la moyenne), l'année suivante, son ratio sera plus fort que l'historique et que la moyenne (graphiquement il sera au dessus de la moyenne).
Ici, c'est le cas sur l'année 1997 sur les années de développement 1, 2 et 3. En première année le ratio est bien plus élevé que la moyenne, puis en seconde année il passe en dessous de la moyenne. Finalement à partir de l'année 3 il va tendre vers la moyenne. Pour l'année 1992, on constate le même phénomène dans le sens inverse (plus faible que la moyenne en 1, puis au dessus en 2 et surtout en 3).

Les ratios $R_{i,j}^{MCL}$ permettent donc d'illustrer l'existence de la corrélation dans un triangle entre la diagonale et la sur diagonale de celui-ci.

Posons

$$\tilde{R}_{i,j}^{MCL} = \frac{P_{i,j}^{CL}}{I_{i,j}^{CL}} \text{ pour } j=a_i, \ldots, N+i+1$$

et $R_{i,j}^{MCL} = \begin{cases} R_{i,j}^{MCL}, j = 1,\ldots,a_{i-1} \\ \tilde{R}_{i,j}^{MCL}, j = a_i,\ldots,N+i+1 \end{cases}$

avec $P_{i,j}^{CL}$ et $I_{i,j}^{CL}$ respectivement l'estimation des sinistres payés et survenus obtenus par la méthode Chain Ladder pour les années de développement j.

Comparons le graphique précédent [2] à un graphique représentant les $\hat{R}_{i,j}^{MCL}$ en fonction des années de développement.

Evolution des $R_{i,j}^{MCL}$ réels et projetés sur le triangle de réassurance :

Graph 3.

Les $\hat{R}_{i,j}^{MCL}$ ne tendent plus vers la moyenne sur les exercices fortement développés.

Une année d'origine ayant un ratio $R_{i,j}^{MCL}$ réel loin de la moyenne, aura des ratios $\hat{R}_{i,j}^{MCL}$ d'autant plus éloignés de la moyenne. Ceci s'illustre avec les années 2005 (année avec peu d'historique) ou 1999.

Ce constat permet d'illustrer d'une autre façon que la projection indépendante du triangle des sinistres payés et des sinistres survenus à l'aide de la méthode Chain Ladder surestime ou sous-estime le sinistre ultime. En effet, cette méthode n'utilise pas l'information de la corrélation entre la diagonale et la sur diagonale du triangle.

Les graphiques représentants les mêmes ratios en assurance sont en ANNEXES PARTIE 2 [4.1]. Ils ne sont pas très significatifs car la branche est à déroulé rapide.

3.5 La méthode du Munich Chain Ladder

Le paragraphe précédent a permis de montrer que les ratios $R_{i,j}^{MCL}$ et $\left(R_{i,j}^{MCL}\right)^{-1}$ sont des facteurs exogènes permettant d'avoir une meilleure estimation des sinistres ultimes.

Ainsi, au lieu d'appliquer la méthode de Chain Ladder pour déterminer un sinistre ultime :

$$U_{i,N}^{CL} = E\left(C_{i,N}\middle|C_{i,1},..C_{i,j}\right) = C_{i,j}\cdot E\left(\frac{C_{i,j+1}}{C_{i,j}}\right)\cdot E\left(\frac{C_{i,j+2}}{C_{i,j+1}}\right)\cdot \mathbf{L}\cdot E\left(\frac{C_{i,N}}{C_{i,N-1}}\right)$$

avec $C_{i,j}$ les sinistres payés ou les sinistres survenus.

Le MCL intègre les facteurs exogènes $R_{i,j}^{MCL}$ et $\left(R_{i,j}^{MCL}\right)^{-1}$:

$$U_{i,N} = E\left(P_{i,N}\middle|P_{i,1},..P_{i,j},\left(R_{i,j}^{MCL}\right)^{-1}\right) = P_{i,j}\cdot E\left(\frac{P_{i,j+1}}{P_{i,j}}\middle|\left(R_{i,j+1}^{MCL}\right)^{-1}\right)\cdot E\left(\frac{P_{i,j+2}}{P_{i,j+1}}\middle|\left(R_{i,j+2}^{MCL}\right)^{-1}\right)\cdot \mathbf{L}\cdot E\left(\frac{P_{i,N}}{P_{i,N-1}}\middle|\left(R_{i,N}^{MCL}\right)^{-1}\right) \quad (5)$$

ou

$$U_{i,N} = E\left(I_{i,N}\middle|I_{i,1},..I_{i,j},\left(R_{i,j}^{MCL}\right)\right) = I_{i,j}\cdot E\left(\frac{I_{i,j+1}}{I_{i,j}}\middle|\left(R_{i,j+1}^{MCL}\right)\right)\cdot E\left(\frac{I_{i,j+2}}{I_{i,j+1}}\middle|\left(R_{i,j+2}^{MCL}\right)\right)\cdot \mathbf{L}\cdot E\left(\frac{I_{i,N}}{I_{i,N-1}}\middle|\left(R_{i,N}^{MCL}\right)\right) \quad (6)$$

Les hypothèses sous-jacentes du modèle de Munich Chain Ladder sont telles que :

Les deux premières hypothèses sont communes avec celle du modèle du Chain Ladder Standard, ici il s'agit des hypothèses pour un SCL sur les payés, les hypothèses sont semblables sur les survenus :

(PE) : hypothèse d'espérance (IE sur les survenus) :

$$\forall t,s\in T,\quad t=s+1,\quad \exists f_{s->t}^{P}>0,\quad i=1,\mathbf{L},n\quad \mathrm{E}\left(\frac{P_{i,t}}{P_{i,s}}\middle|\mathrm{P}_i(s)\right) = f_{s->t}^{P}$$ (H2 dans la Partie I)

(PV) : hypothèse de variance (IV sur les survenus) :

$$\forall t,s\in T,\quad t=s+1,\quad \exists s_{s->t}^{P}>0,\quad i=1,\mathbf{L},n\quad Var\left(\frac{P_{i,t}}{P_{i,s}}\middle|\mathrm{P}_i(s)\right) = \frac{\left(s_{s->t}^{P}\right)^2}{P_{i,s}}$$

(H4 dans la Partie I)

Puis, afin de faciliter les calculs de ce modèle, il est nécessaire que (PIU) :

- $P_{i,1},\mathbf{L},P_{i,j},\left(R_{i,j}^{MCL}\right)^{-1}$ soient indépendantes et identiquement distribuées ou respectivement que
- $I_{i,1},\mathbf{L},I_{i,j},R_{i,j}^{MCL}$ soient indépendantes et identiquement distribuées

Afin de garantir le caractère indépendant et identiquement distribué des variables, la méthode du Munich Chain Ladder utilise les résidus conditionnels de ces variables. Les coefficients de développement pour les sinistres payés sont alors de la forme suivante (hypothèse PQ dans la théorie, respectivement IQ pour les sinistres survenus) :

$$\exists\, l^P \;\forall s,t \in J \quad avec\, t = s+1 \quad i = 1, \mathbf{L}, n$$

$$\mathrm{E}\left(\mathrm{Re}\, s\left(\frac{P_{i,t}}{P_{i,s}} \middle| \mathrm{P}_i(s) \right) \mathrm{B}_i(s) \right) = l^P \cdot \mathrm{Re}\, s\left(\left(R_{i,s}^{MCL}\right)^{-1} \middle| \mathrm{P}_i(s) \right)$$

ou

$$\mathrm{E}\left(\frac{P_{i,t}}{P_{i,s}} \middle| \mathrm{B}_i(s) \right) = f_{s->t}^{P} + l^P \frac{s\left(\frac{P_{i,t}}{P_{i,s}} \middle| \mathrm{P}_i(s) \right)}{s\left(\left(R_{i,s}^{MCL}\right)^{-1} \middle| \mathrm{P}_i(s) \right)} \cdot \left(\left(R_{i,s}^{MCL}\right)^{-1} - \mathrm{E}\left(\left(R_{i,s}^{MCL}\right)^{-1} \middle| \mathrm{P}_i(s) \right) \right) \qquad (7)$$

Avec l^{P} tel que :[4]

$$l^{\mathrm{P}} = corr(\mathrm{Re}\, s\left(\frac{P_{i,t}}{P_{i,s}} \middle| P_i(s) \right), \mathrm{Re}\, s\left(\left(R_{i,s}^{MCL}\right)^{-1} \middle| P_i(s) \right) \text{ soit } l^{\mathrm{P}} = \frac{\mathrm{cov}(\mathrm{Re}\, s\left(\frac{P_{i,t}}{P_{i,s}} \right), \mathrm{Re}\, s\left(\left(R_{i,s}^{MCL}\right)^{-1} \right)}{s\left(\frac{P_{i,t}}{P_{i,s}} \right) \cdot s\left(\left(R_{i,s}^{MCL}\right)^{-1} \right)} \qquad (8)$$

ou $l^{\mathrm{P}} = corr(\frac{P_{i,t}}{P_{i,s}}, \left(R_{i,s}^{MCL}\right)^{-1} \middle| P_i(s)$

[1] [4] Cf. Quarg G., Mack T. (2004). Munich Chain Ladder. A reserving method that reduces the gap between IBNR projections based on paid losses and IBNR projections based on incurred losses.

3.6 La méthode du Munich Chain Ladder une crédibilisation linéaire

Le principe du MCL est d'utiliser deux indicateurs, les sinistres payés et les sinistres survenus afin d'en obtenir un troisième permettant d'obtenir une meilleure estimation de l'ultime. La méthode entre donc dans le cadre général de la crédibilité. Cela se fait via une régression linéaire.

Reprenons les équations l'équation (9) pour montrer qu'il s'agit d'une régression linéaire :

Si l'on reprend l'équation (RL) dans la Partie 2 1.3.2 dans un cas général et les formules de $\hat{a}$ et $\hat{b}$ obtenus (RLO) :

$$\hat{Y} = \hat{a} + \hat{b}Z + e \quad \text{(RL)}$$
$$\hat{a} = E(Y) - \hat{b}.E(Z)$$
$$\hat{b} = \frac{COV(Z,Y)}{V(Z)} \quad \text{(RLO)}$$

en remplaçant on a :

$$\hat{Y} = E(Y) + \hat{b}.\cdot(Z - E(Z)) + e$$
$$\hat{Y} = E(Y) + \frac{COV(Z,Y)}{V(Z)} \cdot (Z - E(Z)) + e \quad \text{(RL2)}$$

A présent, posons

$$Y = \mathrm{E}\left(\frac{P_{i,t}}{P_{i,s}} \middle| \mathrm{B}_i(s)\right) \qquad E(Y) = f_{s->t}^{P} \quad Z = \left(R_{i,s}^{MCL}\right)^{-1} \quad E(Z) = E\left(\left(R_{i,s}^{MCL}\right)^{-1} \middle| \mathrm{P}_i(s)\right)$$

En utilisant la seconde forme du l^{P} :

$$l^{\mathrm{P}} \cdot \frac{s\left(\frac{P_{i,t}}{P_{i,s}} \middle| \mathrm{P}_i(s)\right)}{s\left(\left(R_{i,s}^{MCL}\right)^{-1} \middle| \mathrm{P}_i(s)\right)} = \frac{\mathrm{cov}\left(\frac{P_{i,t}}{P_{i,s}}, \left(R_{i,s}^{MCL}\right)^{-1}\right)}{s\left(\frac{P_{i,t}}{P_{i,s}}\right) \cdot s\left(\left(R_{i,s}^{MCL}\right)^{-1}\right)} \cdot \frac{s\left(\frac{P_{i,t}}{P_{i,s}} \middle| \mathrm{P}_i(s)\right)}{s\left(\left(R_{i,s}^{MCL}\right)^{-1} \middle| \mathrm{P}_i(s)\right)} = \frac{\mathrm{cov}\left(\frac{P_{i,t}}{P_{i,s}}, \left(R_{i,s}^{MCL}\right)^{-1}\right)}{V\left(\left(R_{i,s}^{MCL}\right)^{-1} \middle| \mathrm{P}_i(s)\right)}$$

En remplaçant les variables ci-dessus dans (RL2), on retrouve bien l'équation (9) de la formule du MCL. Cette équation est donc bien une régression linéaire avec des $\hat{a}$ et $\hat{b}$ optimaux. l^{P} est la pente de la droite de régression du nuage de point des résidus. Il ne dépend pas de l'année de développement s.
En reprenant la démonstration vue en Partie 2 1.3.3, il s'agit bien d'une forme de crédibilité linéaire.

En ANNEXES PARTIE 2 [4.2], se trouve le développement de la démonstration de M Merz et M V Wüthrich dans l'article de « A credibility appraoch to the Munich Chain-Ladder Method » (Bibliographie [11]) En utilisant une s -algèbre composée des sinistres payés et survenus cumulés, M Merz et M V Wüthrich montrent que la meilleure projection linéaire affine correspond au résultat d'un MCL. Sur le plan de la crédibilité, ils prouvent que localement l'estimation du MCL est l'estimateur linéaire optimal.

4 Conclusions et postulats de cette Partie

Dans cette Partie, nous avons vu qu'une crédibilisation peut s'apparenter soit à une régression linéaire, soit à un programme d'optimisation.

Cela nous a permis de montrer qu'un certain nombre de méthodes de provisionnement utilise la crédibilité afin d'améliorer l'estimation de ses projections à l'ultime.

Par ailleurs, nous avons étudié la crédibilisation de variables assez différentes :

- Dans le premier cas (A), il s'agit de crédibiliser le résultat de différentes méthodes de provisionnement afin de combiner les points positifs de différentes méthodes. Les coefficients de crédibilité sont directement liés au pourcentage de développement des années d'origine. Ainsi, si une année d'origine est fortement ou faiblement développée, le poids d'une méthode est plus fort que la seconde, puis la tendance s'inverse au cours du développement du triangle.
- Dans le second cas (B), il ne s'agit plus de crédibiliser le résultat de différentes méthodes, mais bien le résultat d'une même méthode appliquée sur des variables différentes (ici les sinistres payés d'une part et les survenus d'autre part). Le coefficient de crédibilité ne dépend plus des pourcentages de développement des années d'origine mais d'un ratio faisant intervenir deux variables qui apportent une information supplémentaire.

Le caractère original du principe de crédibilisation du type (B) nous a amené à réfléchir à d'autres variables susceptibles d'apporter une information supplémentaire à l'estimation de la charge de sinistres ultimes. La fréquence correspond à l'une de ces variables explicatives potentielles. Nous proposons donc dans la Partie 3, une méthode de type (B) utilisant la variable fréquence. Elle utilise la même approche que celle développée par la méthode Munich Chain Ladder

PARTIE 3
ETUDE D'UNE NOUVELLE METHODE DE PROVISIONNEMENT UTILISANT LA THEORIE DE LA CREDIBILITE LINEAIRE VIA UNE INFORMATION EXOGENE

L'objectif de ce mémoire consistait à tester et éventuellement proposer une nouvelle méthode de provisionnement utilisant la théorie de la crédibilité de façon novatrice, à l'instar de ce qui a été réalisé par le développement récent de la méthode Munich Chain Ladder.

Cette troisième partie vise ainsi à présenter les résultats des travaux que nous avons conduits en suivant le cheminement des réflexions menées. Tout d'abord, les bases et les principes qui ont abouti à la formulation de la méthode sont détaillés.

L'influence des variables testées, cherchant à intégrer l'information liée aux nombres de sinistres, ainsi que le bien-fondé de celles-ci sont ensuite exposés à partir de plusieurs exemples réels.

Enfin, la qualité des estimations obtenues à partir de cette méthode est analysée et comparée avec certaines méthodes traditionnelles déjà connues. Nous verrons alors que cette dernière étape nous a amené, par itération, à adapter notre méthode initiale et à proposer finalement une approche qui contribue à enrichir, dans certains cas de figure, le modèle initial du Munich Chain Ladder.

1 Présentation des principes d'une nouvelle approche et formulations

1.1 Les bases de réflexion

Les origines du raisonnement que nous développons dans cette partie s'appuient sur l'idée que les méthodes de provisionnement dites « classiques » décrites dans la Partie I n'utilisent pas suffisamment toute l'information disponible.

En effet, certaines ne se fondent sur aucune information particulière autre qu'une connaissance a priori du risque qu'il ne convient pas ici de remettre en cause (méthode du Loss Ratio), étant encore largement répandue et bien souvent plus pertinente que d'autres comme nous l'avons déjà vu dans la Partie I.

D'autres s'appuient uniquement sur les triangles de paiements ou de sinistres survenus pouvant aboutir parfois, notamment dans le cas de risques longs et volatils, à des sous estimations ou sur estimations significatives de la charge ultime réelle du fait bien souvent de l'application de modèles multiplicatifs (méthodes fondées sur les facteurs de développement).

Une troisième catégorie de méthodes intègre davantage d'information soit par la combinaison de méthodes évoquées juste au dessus (méthodes de Bornhuetter Ferguson et par extension Cape Cod, Benktander), soit en considérant en plus des montants les nombres de sinistres survenus ou déclarés (méthode des coûts moyens).

Enfin, le dernier type de méthode est celle du Munich Chain Ladder (MCL). Celle-ci traite simultanément les triangles de paiements et de sinistres survenus en intégrant une corrélation entre le rapport de ceux-ci et les facteurs de développement observés l'année d'après.

Ainsi, la méthode du MCL nous est apparue extrêmement novatrice dans le sens où elle utilise en plus de la méthode Chain Ladder Classique (SCL), à la fois l'information sur les développements des sinistres survenus et ceux des sinistres payés. Elle intègre aussi implicitement une information supplémentaire, celle des sinistres en suspens. Généralement, l'analyse des suspens était faite de façon qualitative par l'actuaire, mais il n'y avait pas de méthode pour l'utiliser directement.

Nous avons cherché à utiliser de la même façon une information supplémentaire permettant d'améliorer la méthode de SCL. Pour cela, nous avons décidé d'utiliser les triangles de nombres de sinistre $n_{i,j}$ mais d'une manière distincte de celle employée par la méthode des Coûts Moyens.

L'idée a donc été de créer un ou plusieurs ratios et de tester d'éventuelles corrélations avec les facteurs de développement des sinistres payés ou survenus issus de la méthode de SCL. L'objectif final étant de « crédibiliser » l'estimation de ces facteurs de développement à l'instar de la méthode du MCL.

Le principe s'appuie entièrement sur la méthodologie employée dans la méthode MCL, mais en utilisant d'autres ratios que les $R_{i,j}^{MCL}$ et $\left(R_{i,j}^{MCL}\right)^{-1}$, ceux-ci intégrant l'information du nombre de sinistres.

Nous avons donc testé cette méthode en intégrant un paramètre ($n_{i,j}$) supplémentaire au SCL.

Nous avons nommé cette méthode « New Method ». Puis, au vu des résultats obtenus et des conclusions non satisfaisantes, nous avons testé une seconde méthode. Pour cela, nous avons utilisé la conclusion de la méthode de Benktander dans notre Partie 2 : Il est possible de créer un méta estimateur à partir d'un autre méta estimateur. La seconde méthode (« New Method 2 ») consiste à combiner l'information issue de la fréquence et l'estimation issu du MCL. Elle intègre ainsi deux paramètres : la fréquence et le ratio du MCL.

1.2 Formulation de l'approche proposée

1.2.1 Définition des ratios sélectionnés et justification économique

Au cours de nos réflexions, plusieurs ratios ont été abordés de manière à ne nous imposer aucune restriction. Cependant, un certain nombre d'entre eux n'apparaissaient pas pertinents ou n'avaient aucune explication concrète. Ils ont été exclus et ne sont pas présentés dans ce mémoire.

Les variables retenues notées Q_1 et Q_2 correspondent respectivement au coût moyen sur les sinistres survenus $Q_{1,ij} = \frac{I_{i,j}}{n_{i,j}}$ et au rapport paiements cumulés sur nombre de sinistres $Q_{2,ij} = \frac{P_{ij}}{n_{ij}}$ avec $I_{i,j}$, $P_{i,j}$ et $n_{i,j}$ vus dans les parties précédentes.

Ces ratios sont intuitifs et ont l'avantage d'avoir des significations claires, surtout dans le cas de Q_1. L'éventuelle corrélation entre ceux-ci et les facteurs de développement sur sinistres survenus et / ou sinistres payés a été testée sur plusieurs triangles. Ces travaux sont détaillés dans la suite de cette partie.

Dans le cas de Q_1, pour une année d'origine donnée, un coût moyen des sinistres significativement différent de celui observé en moyenne peut impacter le facteur de développement des sinistres payés. En effet, un coût moyen élevé par exemple peut traduire la survenance de davantage de sinistres de gravité supérieure et entraîner un allongement de la cadence des paiements (nécessaire délai de consolidation des conséquences d'un accident, complexité des travaux d'expertise,...). L'impact a priori que nous venons de décrire semble adapté aussi bien aux sinistres d'intensité qu'aux sinistres dits de masse.

Dans le cas de Q_2, les impacts potentiels de cette variable n'apparaissent pas évidents a priori. Nous devons cependant la conserver dans nos analyses de corrélation car celle-ci sera nécessaire lors des estimations de la charge ultime.

1.2.2 Formulation générale

Dans la Partie 2, nous avons vu que la méthode du MCL généralise le modèle de Chain Ladder introduit par Mack. Il s'agit d'un modèle joint pour les payés et les suspens qui prend en compte la corrélation des facteurs de développement sur les $R_{i,j}^{MCL}$ ou $\left(R_{i,j}^{MCL}\right)^{-1}$. Dans cette partie, nous souhaitons créer un modèle joint pour les sinistre payés ou survenus qui prend en compte la corrélation des facteurs de développement sur les ratios Q_1 et Q_2.

De manière plus formalisée, en remplaçant l'information exogène contenue dans les formules de la méthode MCL (cf. (5) et (6) Partie 2) par le ratio dépendant du nombre de sinistres, on obtient l'écriture des charges ultimes :

$$U_{1,N} = E\left(P_{i,N} \middle| P_{i,1},..P_{i,j}, Q_{1,1},..Q_{1,j}\right) = P_{i,j} \cdot E\left(\frac{P_{i,j+1}}{P_{i,j}} \middle| Q_{i_{1,j+1}}\right) \cdot E\left(\frac{P_{i,j+2}}{P_{i,j+1}} \middle| Q_{i_{1,J+2}}\right) \cdot \mathbf{L} \cdot E\left(\frac{P_{i,N}}{P_{i,N-1}} \middle| Q_{i_{1,N}}\right) \tag{9}$$

et

$$U_{2,N} = E\left(I_{i,N} \middle| I_{i,1},..I_{i,j},, Q_{2,1},..Q_{2,j}\right) = I_{i,j} \cdot E\left(\frac{I_{i,j+1}}{I_{i,j}} \middle| Q_{i_{2,j+1}}\right) \cdot E\left(\frac{I_{i,j+2}}{I_{i,j+1}} \middle| Q_{i_{2,j+2}}\right) \cdot \mathbf{L} \cdot E\left(\frac{I_{i,N}}{I_{i,N-1}} Q_{i_{2,N}}\right) \tag{10}$$

Or, afin de faciliter les calculs de ce modèle, il est nécessaire que :

- $P_{i,1}, \mathbf{L}, P_{i,j}, Q_{1,1},..Q_{1,j}$ soient indépendantes et identiquement distribuées ou respectivement que
- $I_{i,1}, \mathbf{L}, I_{i,j},, Q_{2,1},..Q_{2,j}$ soient indépendantes et identiquement distribuées

Afin de garantir le caractère indépendant et identiquement distribué des variables, nous utilisons les résidus conditionnels de ces variables.

Rappel sur la modélisation des résidus conditionnels :

Soit X une variable aléatoire et C une condition et $s(X/C) := \sqrt{Var(X/C)}$ l'écart type conditionnel de X sachant C alors :

$\mathrm{Re}\,s(X/C) := \dfrac{X - E(X/C)}{s(X/C)}$ est le résidu conditionnel de X sachant C. Il est standardisé sous contraintes de :

$$\mathrm{E}\left(\mathrm{Re}\,s(X/C)/C\right) = 0 \qquad et \qquad Var\left(\mathrm{Re}\,s(X/C)/C\right) = 1$$

En reprenant les développements du MCL (cf. (7) Partie 2) et en le transposant sur les ratios Q_1 et Q_2 on aura :

$$\exists\, l_{Q_1}^{P}\ \forall s,t \in T \quad avec\, t = s+1 \quad i = 1, \mathbf{L}, n$$
$$\mathrm{B}_i(s) = \left\{P_{i,1}, \mathbf{L}, P_{i,s}, I_{i,1}, \mathbf{L}, I_{i,s}, n_{i,1}, \mathbf{L}, n_{i,s}\right\}$$
$$\mathrm{E}\left(\mathrm{Re}\, s\left(\frac{P_{i,t}}{P_{i,s}} \middle/ \mathrm{P}_i(s) \right) \middle/ \mathrm{B}_i(s) \right) = l_{Q_1}^{P} \cdot \mathrm{Re}\, s\left(Q_{1_{i,s}} \middle/ \left\{I_i(s), n_i(s)\right\}\right)$$

ou

$$\mathrm{E}\left(\frac{P_{i,t}}{P_{i,s}} \middle/ \mathrm{B}_i(s) \right) = f_{s->t}^{P} + l_{Q_1}^{P} \frac{s\left(\frac{P_{i,t}}{P_{i,s}} \middle/ \mathrm{P}_i(s) \right)}{s\left(Q_{1_{i,s}} \middle/ \left\{I_i(s), n_i(s)\right\}\right)} \cdot \left(Q_{1_{i,s}} - \mathrm{E}\left(Q_{1_{i,s}} \middle/ \left\{I_i(s), n_i(s)\right\}\right)\right) \qquad (11)$$

Et

$$\exists\, l_{Q_2}^{I}\ \forall s,t \in T \quad avec\, t = s+1 \quad i = 1, \mathbf{L}, n$$
$$\mathrm{B}_i(s) = \left\{P_{i,1}, \mathbf{L}, P_{i,s}, I_{i,1}, \mathbf{L}, I_{i,s}, n_{i,1}, \mathbf{L}, n_{i,s}\right\}$$
$$\mathrm{E}\left(\mathrm{Re}\, s\left(\frac{I_{i,t}}{I_{i,s}} \middle/ I_i(s) \right) \middle/ \mathrm{B}_i(s) \right) = l_{Q_2}^{I} \cdot \mathrm{Re}\, s\left(Q_{2_{i,s}} \middle/ \left\{\mathrm{P}_i(s), n_i(s)\right\}\right)$$

ou

$$\mathrm{E}\left(\frac{I_{i,t}}{I_{i,s}} \middle/ \mathrm{B}_i(s) \right) = f_{s->t}^{I} + l_{Q_2}^{I} \frac{s\left(\frac{I_{i,t}}{I_{i,s}} \middle/ I_i(s) \right)}{s\left(Q_{i_{i,s}} \middle/ \left\{\mathrm{P}_i(s), n_i(s)\right\}\right)} \cdot \left(Q_{2_{i,s}} - \mathrm{E}\left(Q_{2_{i,s}} \middle/ \left\{\mathrm{P}_i(s), n_i(s)\right\}\right)\right) \qquad (12)$$

Les équations précédentes formalisent les idées suivantes :

- le coefficient de développement des sinistres survenus (payés) est une fonction monotone linéaire croissante du ratio Q_2 *(*Q_1*)*. Plus précisément, il s'agit de la somme d'un coefficient Chain Ladder Standard et d'un terme de correction linéaire à $Q_{2,s}$ *(*$Q_{1,s}$*)*.
- Le terme de correction comporte trois composantes :
  - Les facteurs $l_{Q_2}^{I}$ et $l_{Q_1}^{P}$ sont indépendants des *i* et des s et correspondent aux coefficients de corrélation entre les résidus des facteurs de développement des sinistres survenus et payés et les résidus des ratios Q_1 et Q_2 respectivement.

$$l_{Q_2}^{I} = corr\left(\mathrm{Re}\,s\left(\frac{I_{i,t}}{I_{i,s}} \middle| I_i(s) \right); \mathrm{Re}\,s\left(Q_{2_{i,s}} \middle| \{\mathrm{P}_i(s), n_i(s)\} \right) \right)$$

$$l_{Q_1}^{P} = corr\left(\mathrm{Re}\,s\left(\frac{P_{i,t}}{P_{i,s}} \middle| P_i(s) \right); \mathrm{Re}\,s\left(Q_{1_{i,s}} \middle| \{I_i(s), n_i(s)\} \right) \right)$$

Ils sont compris entre 0 et 1 et mesurent la dépendance entre ces deux valeurs. Dans le cas d'une corrélation très faible, ces coefficients sont proches de 0 et le coefficient de développement revient à celui d'un Chain Ladder Standard.

- L'écart type est le rapport des écarts type conditionnels des facteurs de développement des sinistres survenus (payés) et des ratios Q_2 (Q_1). Cela permet de réajuster l'écart de ces ratios à leur moyenne sur les facteurs de développement. Plus l'écart type des facteurs de développement est grand, plus le coefficient de correction sera grand.
- Les termes linéaires, que nous appellerons dans la suite « termes de correction », $Q_{1_{i,s}} - \mathrm{E}\left(Q_{1_{i,s}} / \{I_i(s), n_i(s)\}\right)$ et $Q_{2_{i,s}} - \mathrm{E}\left(Q_{2_{i,s}} / \{\mathrm{P}_i(s), n_i(s)\}\right)$ représentent l'écart à la moyenne des ratios Q_1 et Q_2. Ceux-ci permettent donc d'effectuer un réajustement de leurs écarts à la moyenne. Plus les ratios sont proches de la moyenne, plus les coefficients de développement de la méthode proposée se rapprochent d'un coefficient de Chain Ladder Standard.

1.2.3 Formulation des différents paramètres

Ci-dessous le tableau synthétise les formules des différents paramètres intervenant dans le calcul de la méthode proposée :

	Coefficients de développement et moyenne	Ecart type
Payés	$f_{s->t}^{P} = \frac{\sum P_{i,t}}{\sum P_{i,s}}$ (15)	$\hat{s}_s^{P} = \sqrt{\frac{1}{I-s-1} \cdot \left(\frac{\sum_i P_{i,t}}{\sum_i P_{i,s}} - f_{s->t}^{P} \right)^2 \cdot P_{i,s}}$
Survenus	$f_{s->t}^{I} = \frac{\sum I_{i,t}}{\sum I_{i,s}}$ (16)	$\hat{s}_s^{I} = \sqrt{\frac{1}{I-s-1} \cdot \left(\frac{\sum_i I_{i,t}}{\sum_i I_{i,s}} - f_{s->t}^{I} \right)^2 \cdot I_{i,s}}$
Q_1	$f_{s->t}^{Q_1} = \frac{\sum_i I_{i,t}}{\sum_i n_{i,s}}$	$\hat{r}_s^{Q_1} = \sqrt{\frac{1}{I-s-1} \cdot \sum_i \left(Q_{1_{i,t}} - f_{s->t}^{Q_A} \right)^2 \cdot I_{i,s}}$
Q_2	$f_{s->t}^{Q_1} = \frac{\sum_i P_{i,t}}{\sum_i n_{i,s}}$	$\hat{r}_s^{Q_2} = \sqrt{\frac{1}{I-s-1} \cdot \sum_i \left(Q_{2_{i,t}} - f_{s->t}^{Q_2} \right)^2 \cdot P_{i,s}}$

1.2.4 Les résidus

Ici les résidus sont sur les coefficients de développement

	Résidus sur les Coefficients de développement
$\mathrm{Re}\,s\left(\frac{P_{i,t}}{P_{i,s}} \Big/ \mathrm{P}_i(s)\right)$	$\frac{1}{\hat{s}^{P}_{s-1}}\left(\frac{P_{i,t}}{P_{i,s}} - f^{P}_{s->t}\right)\cdot\sqrt{P_{i,s}}$
$\mathrm{Re}\,s\left(\frac{I_{i,t}}{I_{i,s}} \Big/ I_i(s)\right)$	$\frac{1}{\hat{s}^{I}_{s}}\left(\frac{I_{i,t}}{I_{i,s}} - f^{I}_{s->t}\right)\cdot\sqrt{I_{i,s}}$
$\mathrm{Re}\,s\left(Q_{1_{i,s}} \Big/ I_i(s)\right)$	$\frac{1}{\hat{r}^{Q_1}_{s}}\left(Q_{1_{i,t}} - f^{Q_1}_{s->t}\right)\cdot\sqrt{I_{i,s}}$
$\mathrm{Re}\,s\left(Q_{2_{i,s}} \Big/ \mathrm{P}_i(s)\right)$	$\frac{1}{\hat{r}^{Q_2}_{s}}\left(Q_{2_{i,t}} - f^{Q_2}_{s->t}\right)\cdot\sqrt{P_{i,s}}$

1.2.5 Les Corrélations

Pour les mêmes raisons que celles énoncées dans l'article de G. Quarg, les données à disposition ont été traitées d'une manière comparable à celle utilisée dans la méthode du Munich Chain Ladder.

Pour rappel, la méthode MCL met en évidence la corrélation entre P/I (I/P) des années précédentes et les facteurs de développement SCL sur sinistres survenus (Payés) en centrant et réduisant ces variables *(selon les notations définies en (8) dans la Partie 2)*.

$$l^{P}_{MCL} = \frac{\mathrm{cov}(\mathrm{Re}\,s\left(\frac{P_{i,t}}{P_{i,s}}\right), \mathrm{Re}\,s\left(\left(R^{MCL}_{i,s}\right)^{-1}\right)}{s\left(\frac{P_{i,t}}{P_{i,s}}\right)\cdot s\left(\left(R^{MCL}_{i,s}\right)^{-1}\right)} \qquad \text{et} \qquad l^{I}_{MCL} = \frac{\mathrm{cov}(\mathrm{Re}\,s\left(\frac{I_{i,t}}{I_{i,s}}\right), \mathrm{Re}\,s\left(\left(R^{MCL}_{i,s}\right)\right)}{s\left(\frac{I_{i,t}}{I_{i,s}}\right)\cdot s\left(\left(R^{MCL}_{i,s}\right)\right)}$$

La technique est identique pour tester d'une part l'éventuelle corrélation entre les coûts moyens Q_1 et les facteurs de développement sur sinistres survenus et payés (respectivement $l^{I}_{Q_1}$ et l^{P}_{Q1}) et d'autre part celle entre le rapport paiements cumulés sur nombre de sinistres Q_2 et ces mêmes facteurs de développement (respectivement $l^{I}_{Q_2}$ et $l^{P}_{Q_2}$).

Les écritures sont les suivantes :

$$l_{Q_1}^{P} = \frac{\operatorname{cov}\left(\operatorname{Re} s\left(\frac{P_{i,t}}{P_{i,s}} \right), \operatorname{Re} s\left(Q_{1i,s}\right) \right)}{s\left(\frac{P_{i,t}}{P_{i,s}} \right) \cdot s\left(Q_{1i,s}\right)} \qquad l_{Q_2}^{P} = \frac{\operatorname{cov}\left(\operatorname{Re} s\left(\frac{P_{i,t}}{P_{i,s}} \right), \operatorname{Re} s\left(Q_{2i,s}\right) \right)}{s\left(\frac{P_{i,t}}{P_{i,s}} \right) \cdot s\left(Q_{2i,s}\right)}$$

$$l_{Q_1}^{I} = \frac{\operatorname{cov}\left(\operatorname{Re} s\left(\frac{I_{i,t}}{I_{i,s}} \right), \operatorname{Re} s\left(Q_{1i,s}\right) \right)}{s\left(\frac{I_{i,t}}{I_{i,s}} \right) \cdot s\left(Q_{1i,s}\right)} \qquad l_{Q_2}^{I} = \frac{\operatorname{cov}\left(\operatorname{Re} s\left(\frac{I_{i,t}}{I_{i,s}} \right), \operatorname{Re} s\left(Q_{2i,s}\right) \right)}{s\left(\frac{I_{i,t}}{I_{i,s}} \right) \cdot s\left(Q_{2i,s}\right)}$$

2 Analyse actuarielle de l'approche proposée

2.1 Données utilisées

Les travaux décrits dans cette Partie 3 s'appuient sur les résultats de plusieurs applications numériques. Nous avons cherché à nous assurer dans la mesure du possible que la mise en évidence d'une forte corrélation lors de l'analyse portant sur les triangles d'une branche donnée ne soit pas le simple fait du hasard et qu'elle ne constitue pas un cas isolé.

Comme nous l'avions déjà fait dans la première partie, nous avons réalisé des tests sur des triangles de natures diverses issus de deux sociétés distinctes, l'une de réassurance, l'autre d'assurances. Le but étant de vérifier si la méthode peut fonctionner sur tous les types de risque.

2.1.1 Les triangles de réassurance

Quatre triangles sont testés en réassurance. Nous détaillerons uniquement les particularités de celui dont les résultats sont utilisés : celui en Auto Non Proportionnel (dans les tableaux suivants il est indiqué comme celui en « Motor »).

N'ayant pas le triangle de nombre de sinistres sur la branche de réassurance utilisé dans les premières parties du mémoire nous avons sélectionné un triangle sur la même branche mais en non proportionnel.

Le nombre de contrats souscrits par année est à peu près homogène, avec une souscription un peu inférieure sur les années 1999 et 2000. En termes de montant de primes, la souscription est plus faible sur les années 1998 à 2000.

La répartition géographique sur cette branche peut avoir un impact dans la projection des triangles, car d'un pays à l'autre, les comportements des conducteurs, les jurisprudences ainsi que les barèmes peuvent être assez différents. Nous avons tout de même décidé de les agréger.

Dans ce portefeuille, les traités sont des excédents de sinistres indexés ou non indexés avec des tranches limitées ou illimitées. Ils peuvent concerner de la réassurance directe d'assureurs ou de réassureurs.

Les limites et priorités des traités sous-jacents sont très variables, elles peuvent être très basses ou très hautes. En moyenne par année, les traités sinistrés ont des priorités décroissantes mais assez homogènes depuis 1998 avec comme exceptions les années 2002 et 2005.

2.1.2 Les triangles d'assurance

Les données d'assurance appartiennent à une société présente sur le marché français. Celles-ci ont été modifiées par souci de confidentialité sans que cela n'ait d'influence sur les caractéristiques intrinsèques des triangles. Notons toutefois que cette société a connu un développement assez important sur la dernière décennie, ce qui explique pourquoi l'historique est assez court, les premières années étant trop peu significatives. Sur les années plus représentatives et afin de ne pas polluer l'analyse, les données ont été corrigées de l'effet taille du portefeuille.

Trois triangles ont été exploités :

- Le premier concerne les risques de responsabilité civile matériels en assurance automobile. Les données sont trimestrielles et portent sur la période septembre 2004 – décembre 2007, soit 14 trimestres d'origine.
- Le deuxième traite du risque de dommages tous accidents (DTA) toujours en assurance automobile. Les données sont annuelles et portent sur la période 2001 – 2008, soit 8 années d'origine.
- Enfin, le troisième s'intéresse à l'assurance multirisques habitations, y compris les catastrophes naturelles. Les données sont également annuelles et couvrent la même période 2001-2008, soit 8 années d'origine.

Ces trois triangles portent sur un portefeuille qui couvre de manière homogène l'ensemble du territoire national. Les données ont été déflatées selon la même méthodologie que celle exposée dans le sous paragraphe 1.2 de la partie 1 relatif à la présentation des données d'assurance.

- Pour le risque automobile les indices de prix utilisés sont ceux fournis par SRA. L'indice retenu pour « déflater » les données est une combinaison de trois indices (prix des pièces détachées, prix des ingrédients de peinture, prix de la main d'œuvre carrosserie) pondérés par leur part respective dans le coût total des réparations automobiles telle qu'indiquée dans le flash d'information SRA du mois de janvier 2010 (respectivement 48,4%, 10,3%, 41,3%)[5].
- Pour le risque multirisques habitations, il s'agit toujours de l'indice FFB.

[5] Plus d'informations à cette adresse : http://www.sra.asso.fr/cv/sraedit.nsf/HTML/ALOI-83MKB8.html/$FILE/FLASH%2074.pdf

2.2 Etude des corrélations

L'étude des corrélations ci-après s'appuie dans un premier temps sur les tableaux 1a et 1b qui reprennent de manière synthétique l'ensemble des coefficients de corrélation définis dans le paragraphe précédent et calculés sur chacun des triangles présentés. Elle est ensuite approfondie par une analyse graphique sur deux des sept exemples. Les graphiques pour l'ensemble des triangles sont disponibles en ANNEXES PARTIE 3 1.

2.2.1 Analyse statique des coefficients de corrélations

Tableau 1a : tableau des coefficients de corrélation par risque sur facteurs de développement de sinistres payés

	facteurs de développement sur paiements			
	variable I/P	variable P/I	variable I/NS (Q1)	variable P/NS (Q2)
	l^{P}_{MCL}		$l^{P}_{Q_1}$	$l^{P}_{Q_2}$
triangles de réasssurance				
motor (inclus 2001)	0,28	-0,28	0,36	0,06
motor (hors 2001)	0,32	-0,34	0,20	-0,15
marine	0,71	-0,96	0,05	-0,28
Property per Risk	0,39	-0,40	0,09	0,03
Property Cat	0,48	-1,00	0,16	0,12
triangles d'assurance (marché français)				
Automobile - RC matériels	0,51	-0,52	0,16	-0,16
Automobile dommages - annuel	0,66	-0,66	0,68	0,40
Multirisques habitations - annuel	0,32	-0,34	0,20	-0,15

Clé de lecture : dans le cas du « motor (inclus 2001) », le coefficient de corrélation entre les variables I/P et facteurs de développement des payés (centrées réduites) est de 0,28

<u>Tableau 1b</u> : tableau des coefficients de corrélation par risque sur facteurs de développement de sinistres survenus

	facteurs de développement sur Incurred			
	variable I/P	variable P/I	variable I/NS (Q1)	variable P/NS (Q2)
		l^{I}_{MCL}	$l^{I}_{Q_1}$	$l^{I}_{Q_2}$
triangles de réasssurance				
motor (inclus 2001)	-0,15	0,19	-0,15	0,08
motor (hors 2001)	-0,24	0,26	-0,15	0,14
marine	0,00	0,04	-0,07	0,05
Property per Risk	-0,38	0,37	0,00	0,01
Property Cat	-0,16	0,16	0,00	-0,02
triangles d'assurance (marché français)				
Automobile - RC matériels	0,01	0,00	0,20	0,23
Automobile dommages - annuel	-0,35	0,35	0,18	0,30
Multirisques habitations - annuel	-0,24	0,26	-0,15	0,14

Plusieurs remarques ou commentaires peuvent être faits à partir des tableaux 1a et 1b ci-dessus :

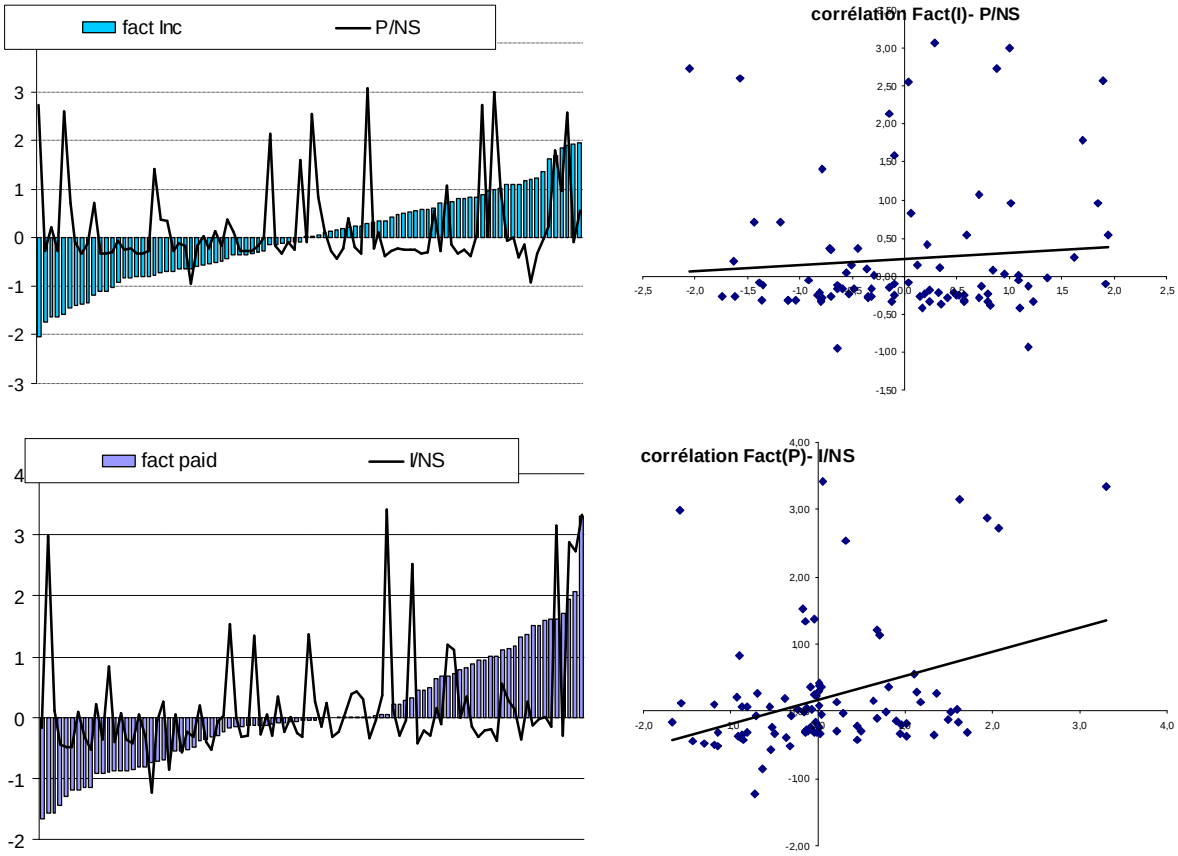
- les corrélations calculées dans le cadre de la méthode MCL apparaissent presque toujours supérieures en valeur absolue à celles testées dans notre approche sauf dans les cas de la réassurance Motor et de l'assurance automobile dommages pour les facteurs de développement sur paiements. Il convient de rappeler toutefois que la construction même des variables servant à la méthode du MCL implique mécaniquement une corrélation implicite, les montants de paiements cumulés (ou de sinistres survenus) de l'année n-1 servant à la fois à la construction de la variable à expliquer et de la variable expliquée.
- Les corrélations sur les variables testées apparaissent bien meilleures sur les triangles d'assurance que sur ceux de réassurance. Cette observation peut paraître étonnante, étant donné les variations importantes de coût moyen que nous pouvons constater d'une année sur l'autre pour l'activité de réassurance.
- Les corrélations testées sur les facteurs de développement des paiements apparaissent plus évidentes que celles réalisées sur les facteurs de développement des sinistres survenus. Ce même phénomène se retrouve sur les niveaux de corrélation constatés sur la méthode MCL. Cela tend à confirmer notre sentiment initial : le niveau de coût moyen est susceptible d'influencer la vitesse de paiement. Le signe positif de la corrélation corrobore également notre idée a priori.

- Les corrélations $l^P_{Q_1}$ et $l^I_{Q_2}$ semblent meilleures que $l^P_{Q_2}$ et $l^I_{Q_1}$ respectivement. Cela s'observe toutefois essentiellement sur les triangles d'assurance. Ce sont ces corrélations que nous retenons dans la suite notamment pour tester la qualité des estimations de cette approche.

2.2.2 Analyse graphique des corrélations

L'analyse graphique présentée ci-dessous s'attarde sur deux triangles : l'un dont la mise en évidence de corrélations ne semble pas évidente (réassurance Motor) et un autre où un lien apparaît plus présent.

<u>Tableaux 2a</u> : triangle de réassurance Motor avec 2001



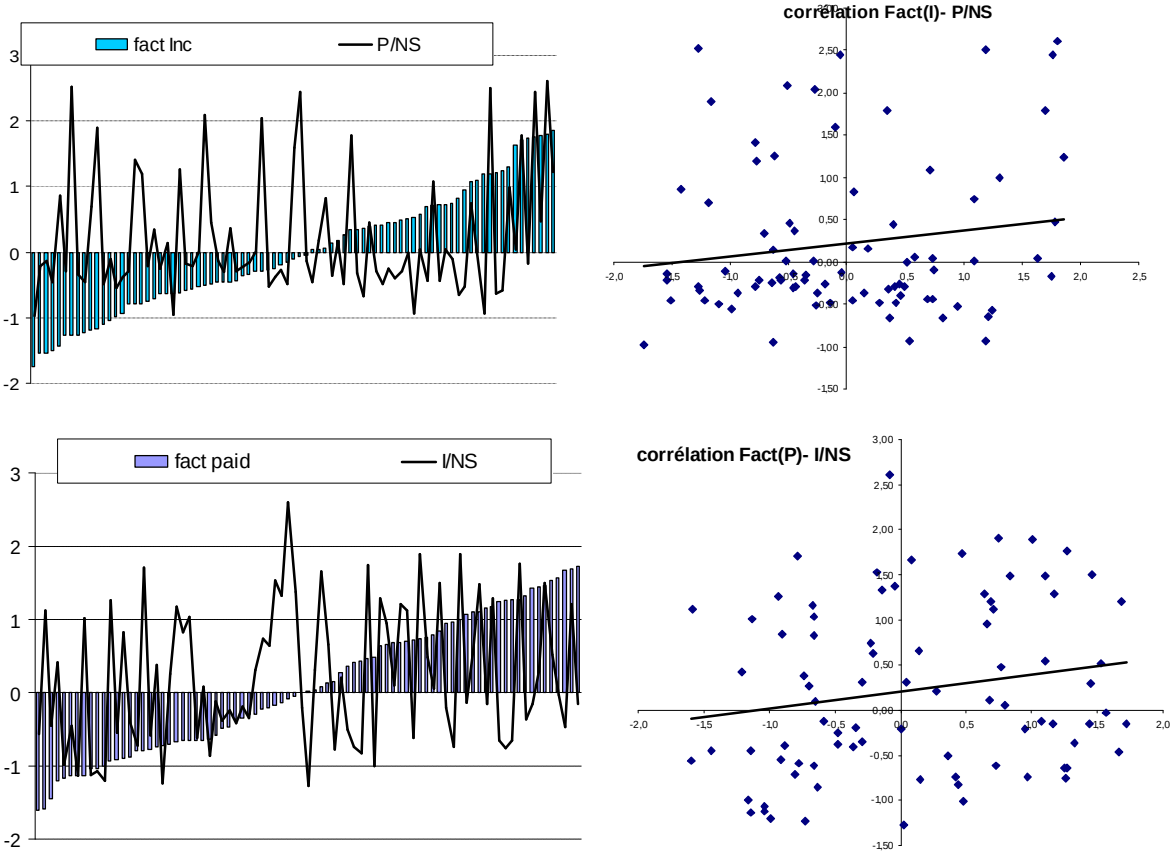
Les graphiques de gauche comparent les facteurs de développement centrés réduits triés par ordre croissant avec la valeur correspondante des variables Q_1 ou Q_2 également centrées réduites. Ce type d'illustration est un outil parmi d'autres visant à faciliter l'interprétation d'éventuels liens entre deux chroniques.

La lecture de ces graphiques illustre les niveaux de corrélation repris dans les tableaux 1a et 1b : $l^P_{Q_1}$ =0,36 et $l^I_{Q_2}$ =0,08

Les graphiques de droite représentent un nuage de points avec ces mêmes variables.

L'année 2001 semble fausser l'analyse. Nous avons donc décidé de l'exclure afin de voir si cela améliore les corrélations, ci-dessous les analyses graphiques correspondantes à cette exclusion :

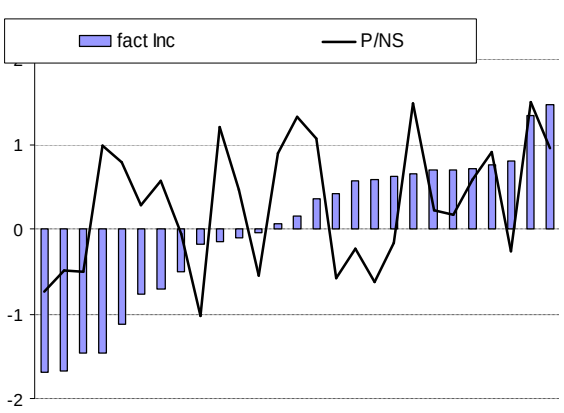
<u>Tableau 2b</u> : triangle de réassurance Motor sans 2001

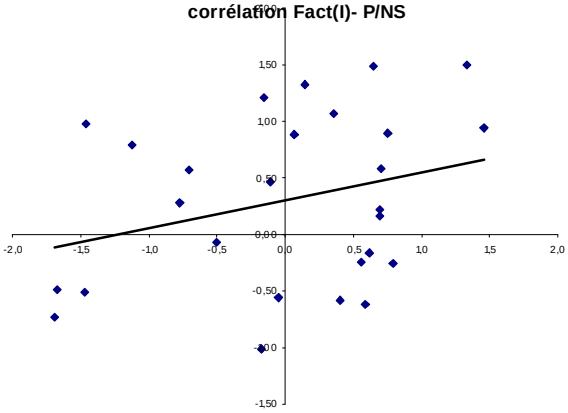


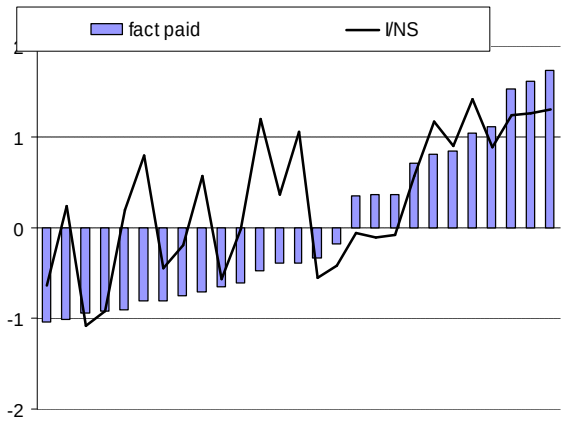
La lecture de ces graphiques illustre les niveaux de corrélation repris dans les tableaux 1a et 1b : $l_{Q_1}^{P}$ =0,20 et $l_{Q_2}^{I}$ =0,14. Nous pouvons noter toutefois une corrélation sur l'ensemble des facteurs de développement dans le cas de $l_{Q_1}^{P}$ (graphique en bas à gauche) et davantage centré sur les facteurs de développement les plus élevés dans le $l_{Q_2}^{I}$ (graphique en haut à gauche). On remarque que l'exclusion de 2001 diminue le $l_{Q_1}^{P}$ et accroît $l_{Q_2}^{I}$.

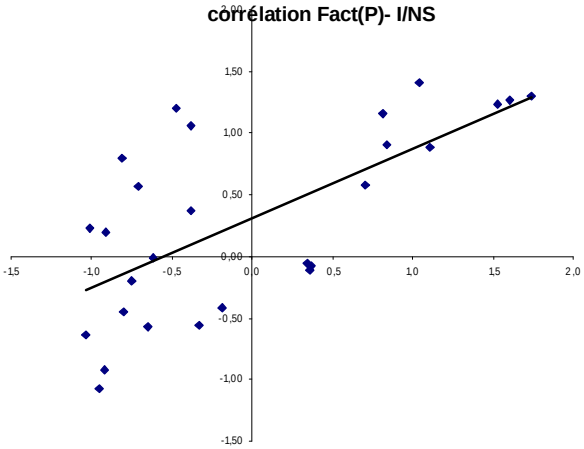
Notre approche est fondée sur l'étude des corrélations. Toutefois, une réflexion plus approfondie pourrait être menée et conduire éventuellement à identifier des types de dépendance non linaires entre ces variables. A ce titre, le graphique des tableaux 2a (en bas à droite) est un bon exemple.

<u>Tableau 3</u> : triangle d'assurance automobile dommages tous accidents









Les corrélations apparaissent plus clairement sur ces graphiques comparativement à celles du triangle de réassurance Motor et illustrent bien les coefficients $l^{P}_{Q_1}$ et $l^{I}_{Q_2}$ respectivement égaux à 0,68 et 0,30. Dans les deux cas, une corrélation semble se dessiner, de manière plus distincte dans le cas des facteurs de développement sur payés et des ratios Q_1.

Il convient toutefois de rester prudent sur l'interprétation des résultats en raison d'un nombre d'observations beaucoup plus restreint que dans le cas de la réassurance (27 contre 83 respectivement).

3 Application et qualité de l'estimation

La sous partie précédente nous a permis de mettre en évidence des liens indirects entre les niveaux de fréquence et les facteurs de développement. Cette première étape justifie pleinement de tester l'approche proposée, que nous avons détaillée dans la sous partie « formulation générale ».

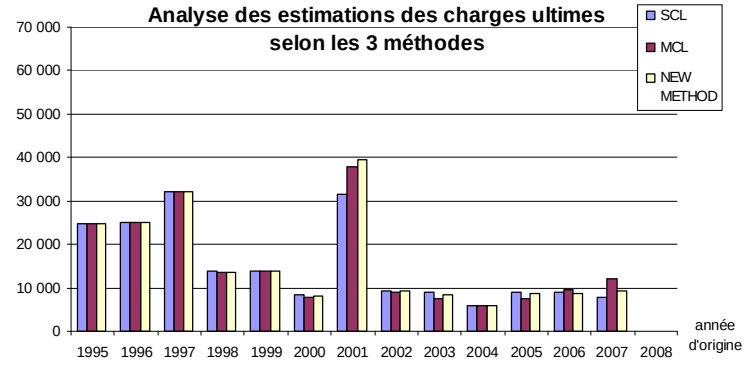
A présent, analysons les résultats obtenus par cette méthode. Puis, nous testons les qualités prédictives de la méthode à partir d'une analyse graphique des estimations à l'ultime en les comparant à d'autres méthodes. Enfin, nous réalisons une analyse en « backtesting » afin de comprendre comment se comporte la méthode sur les données passées du triangle.

3.1 Analyse graphique des projections à l'ultime

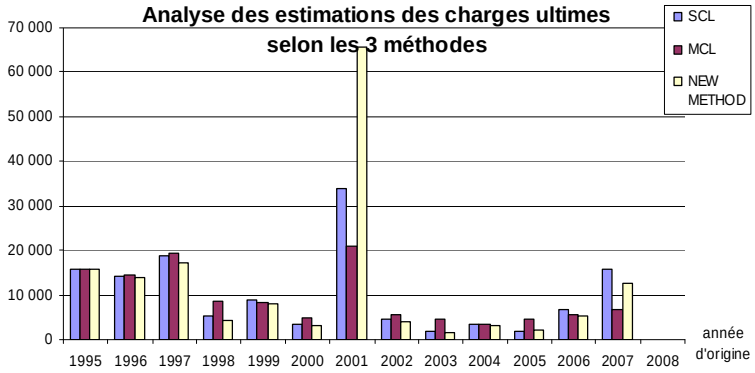
L'appréhension de la qualité des estimations fournies par la méthode proposée s'appuie sur plusieurs critères, étant entendu qu'il apparaît difficile à une date donnée d'émettre un avis sur la prévision d'une charge ultime dont nous ne connaissons pas encore le niveau définitif. Pour ce faire, nous comparons les projections obtenues avec les méthodes SCL et MCL ainsi que la convergence entre leurs estimations basées sur les triangles de sinistres survenus et payés. L'analyse graphique est présentée sur les triangles de réassurance Motor et d'assurance dommages automobiles. Les charges ultimes projetées des autres triangles sont en ANNEXES PARTIE 3 1.

Tableaux 4a : Réassurance Motor (avec 2001) : étude des projections à l'ultime

Estimations à l'ultime à partir des sinistres survenus



Estimations à l'ultime à partir des sinistres payés



Sur les années antérieures à 2001, les méthodes convergent vers un ultime identique pour l'estimation basée sur les sinistres survenus. Ceci n'est pas le cas pour l'estimation à l'ultime à partir des sinistres payés. Toutefois, pour les années 1997 à 2000, le MCL propose un ultime supérieur ou au moins égal au SCL qui se rapproche de l'estimation basée sur les sinistres survenus. Le MCL converge vers l'estimation obtenue via les sinistres survenus alors que le SCL payés les sous-estime. Ce qui n'est pas le cas de la « New Method »sur ces années.

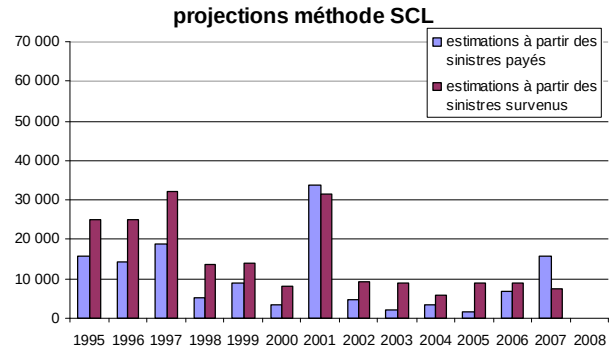
On remarque le caractère atypique de l'année 2001 au travers des résultats obtenus. Dans les deux estimations, la « New Method » est celle qui propose l'ultime le plus fort.

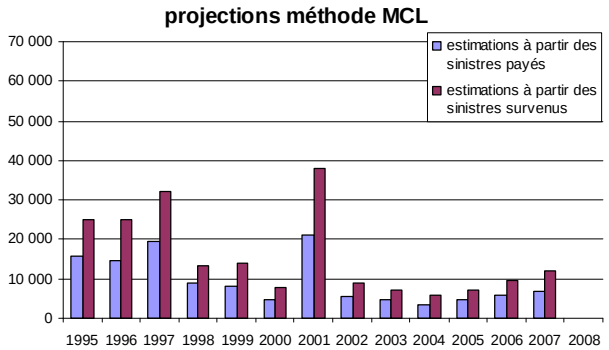
Pour l'année 2007, il est intéressant de voir que l'estimation d'ultime basée sur les sinistres survenus du MCL est celle la plus élevée alors que lorsque celle-ci est basée sur les sinistres payés elle est la plus faible. Par contre, la « New Method » propose un ultime inférieur au SCL sur Sinistres survenus et mais supérieur au MCL sur payés.

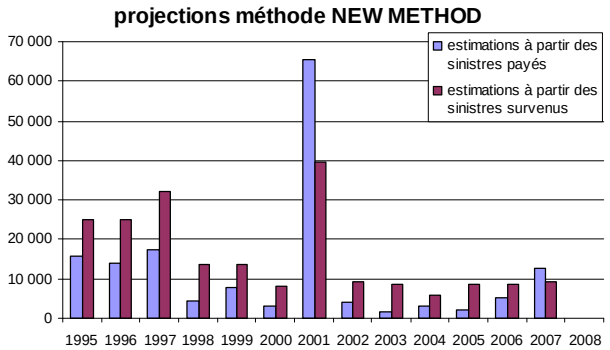
Au vu de ces observations, dans certains cas, l'estimation de la « New Method » parait plus appropriée que le SCL mais est inférieure au MCL. Dans certains cas, toutes les estimations convergent. Dans d'autres la « New Method » améliore les résultats du MCL et SCL. La première conclusion est que l'information du nombre de sinistres peut dans certains cas proposer un ultime de meilleur qualité.

Les tableaux ci-dessous mettent en regard pour chacune des méthodes, le montant d'ultime obtenu en fonction de la base du calcul (sinistres payés ou survenus)

Convergence des estimations basées sur les triangles des survenus versus payés







L'estimation des ultimes à partir des sinistres payés apparaît très inférieure au résultat obtenu quelle que soit la méthode à l'aide des sinistres survenus. On peut penser que les montants de sinistres en suspens sont particulièrement élevés en particulier sur les années les plus développées. Il y a sans doute des litiges non encore résolus.

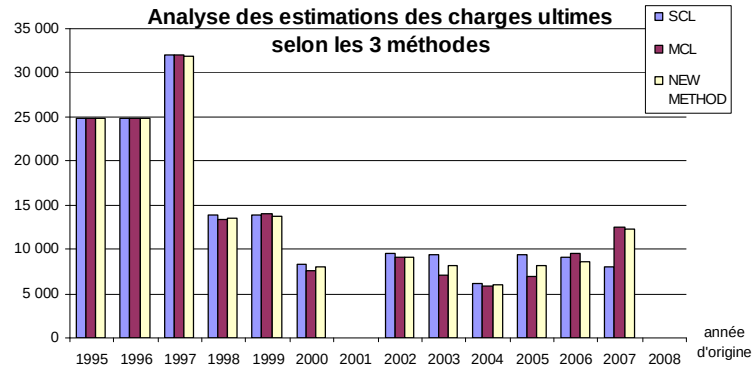
Une autre explication pourrait être que la branche est plus longue en terme de payés qu'en terme de sinistres survenus. Il faudrait donc projeter sur un plus grand nombre d'années de développement les payés. Or ceci n'est possible qu'en ajoutant des coefficients de queue. Nous ne prenons pas cette option car cela produirait un biais supplémentaire à l'estimation des ultimes.

La convergence des ultimes en fonction des sinistres payés ou survenus n'est donc pas un test pertinent sur ce triangle.

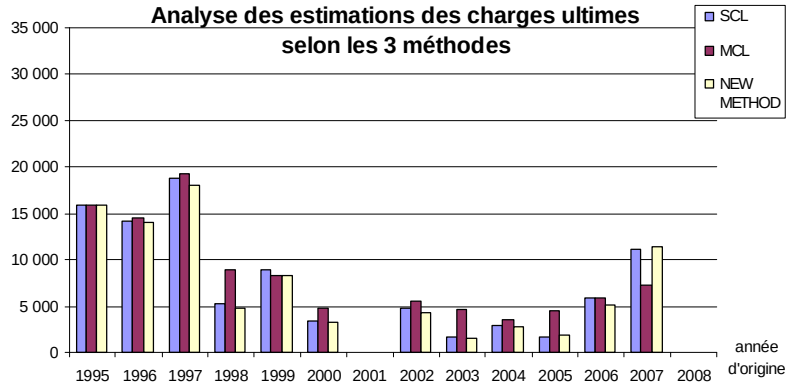
Le caractère atypique de l'année 2001 nous a amenés à tester les méthodes en excluant 2001. Voici les résultats obtenus en Tableaux 4b :

Tableaux 4b : Réassurance Motor (hors 2001) : étude des projections à l'ultime

Estimations à l'ultime à partir des sinistres survenus



Estimations à l'ultime à partir des sinistres payés

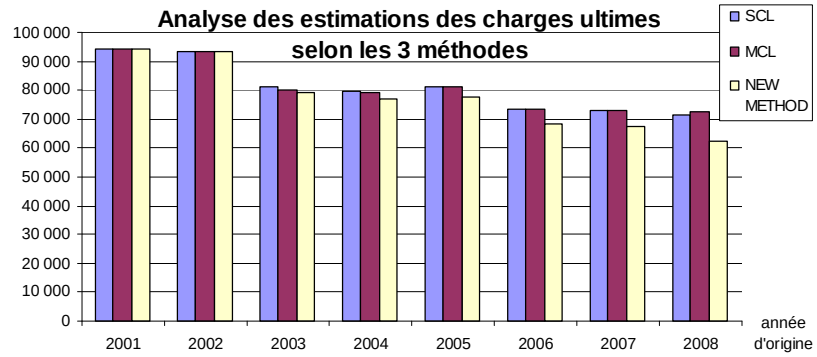


Le fait d'exclure l'année 2001 permet une estimation supérieure des ultimes des années récentes (essentiellement 2007) tant pour le MCL que pour la nouvelle méthode. Sur les autres années, cette exclusion ne semble pas avoir un gros impact sur l'estimation des ultimes.

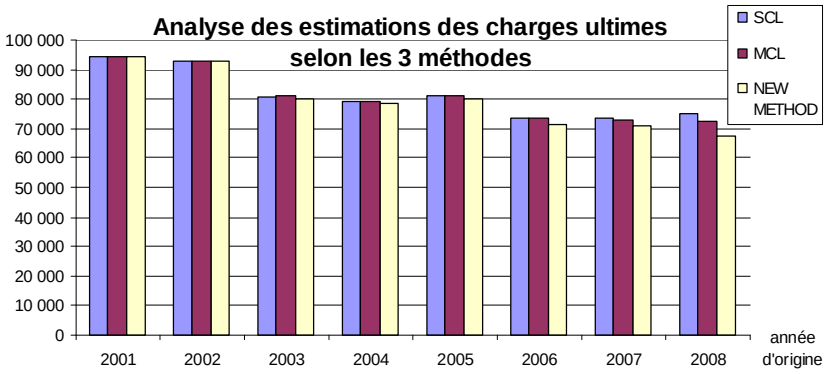
A présent, testons les résultats obtenus sur le triangle d'assurance automobile dommages afin de voir si nous obtenons des résultats semblables ou si nous constatons d'autres phénomènes.

Tableaux 5 : Assurance automobile dommages : étude des projections à l'ultime

Estimations à l'ultime à partir des sinistres survenus



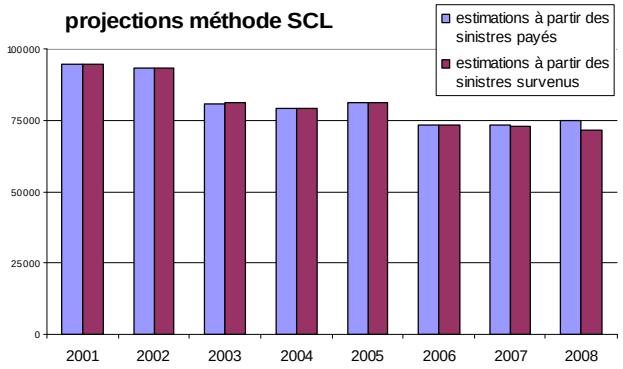
Estimations à l'ultime à partir des sinistres payés

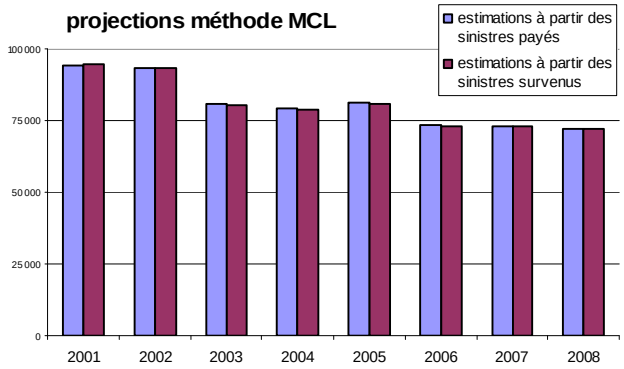


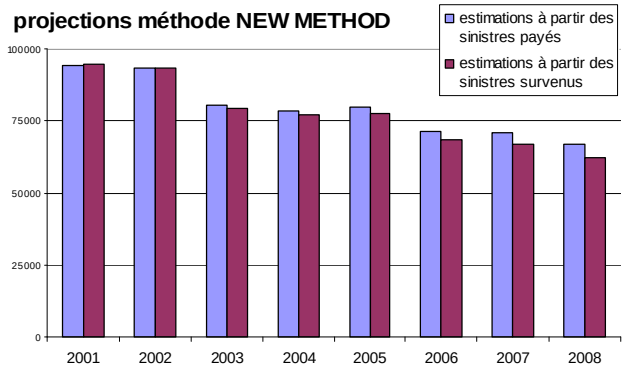
Les trois méthodes ont l'air de converger avec un décalage pour la « New Method » sur les années les plus récentes. La « New Method » parait plus optimiste que les deux autres.
L'ajout du paramètre sur le nombre de sinistres n'a pas l'air d'apporter plus d'information à la méthode.

Analysons la convergence des méthodes :

Convergence des estimations





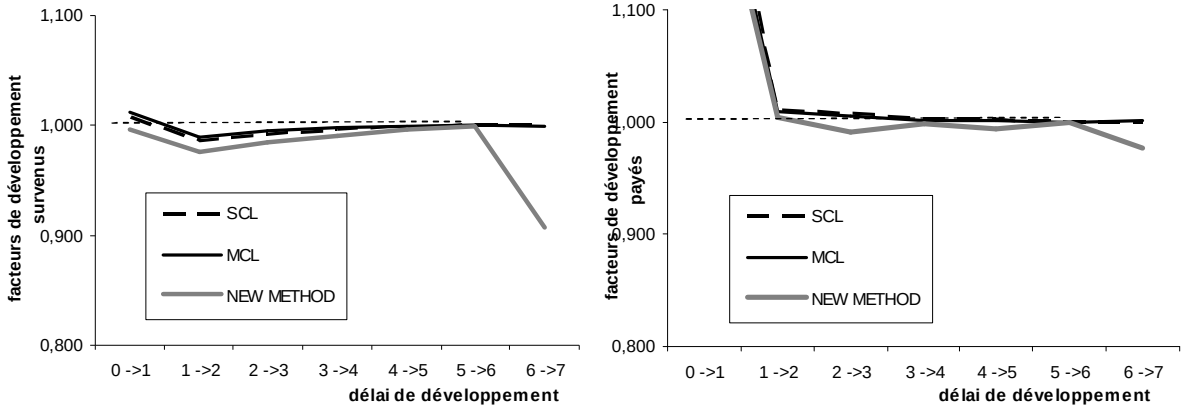


Les méthodes de SCL et MCL convergent toutes les deux.

Malgré des coefficients de corrélation très significatifs (pour rappel $l^{P}_{Q_1}$ =0,68 et $l^{I}_{Q_2}$ =0,30), les projections à l'ultime de la méthode proposée semblent traduire une sous estimation de la charge par rapport aux autres méthodes, ceci alors que les données du triangle apparaissent homogènes.

En analysant graphiquement (graphique 6 ci-dessous) la chronique des facteurs de développement des sinistres payés sur l'année la moins développée, en l'occurrence 2008, la raison de la sous estimation apparaît clairement. Cet écart est dû à l'estimation du dernier facteur, largement inférieur aux deux autres, que ce soit sur les sinistres survenus ou les sinistres payés. Après vérification, ceci se retrouve pour chacune des autres années d'origine.

<u>Graphiques 6</u> : facteurs de développement estimés pour l'année 2008



Cet écart est dû à deux points :

- premièrement, contrairement aux variables *I/P* ou *P/I* qui convergent vers 1 à mesure que l'on s'approche de la clôture de l'année d'origine (voir tableau 7 ci-dessous), les termes de correction $\left(\frac{I}{NS}\right)_{i,s} - \left(\frac{I}{NS}\right)_{s}$ (c'est-à-dire $Q_{1i,s} - Q_{1s}$), (ou $\left(\frac{P}{NS}\right)_{i,s} - \left(\frac{P}{NS}\right)_{s}$) (c'est-à-dire $Q_{1i,s} - Q_{1s}$) reportés délai après délai n'ont aucune raison de converger.

L'influence de ces variables *I/P* ou *P/I* sur la correction des facteurs de développement à estimer s'estompe avec l'avancement de la liquidation, ce qui n'est pas forcément le cas dans notre méthode.

	0	1	2	3	4	5	6
facteurs de développement des sinistres payés							
MCL							
I/P année 2008	133,6%	103,7%	101,7%	100,6%	100,3%	100,1%	100,1%
I/P moyen	139,4%	105,1%	102,6%	101,1%	100,6%	100,2%	100,0%
termes de correction	**-5,8%**	**-1,4%**	**-0,9%**	**-0,4%**	**-0,3%**	**-0,1%**	**0,0%**
NEW METHOD							
I/NS année 2008	745	707	690	679	673	670	669
I/NS moyen	864	846	845	853	865	877	892
termes de correction	**-120**	**-139**	**-155**	**-174**	**-192**	**-207**	**-223**
facteurs de développement des sinistres survenus							
MCL							
P/I année 2008	74,9%	96,4%	98,4%	99,4%	99,7%	99,9%	99,9%
P/I moyen	71,7%	95,1%	97,5%	99,0%	99,4%	99,8%	100,0%
termes de correction	**3,1%**	**1,3%**	**0,9%**	**0,4%**	**0,3%**	**0,1%**	**0,0%**
NEW METHOD							
P/NS année 2008	557	679	681	675	673	669	669
P/NS moyen	620	805	824	844	860	875	892
termes de correction	**-63**	**-126**	**-143**	**-169**	**-186**	**-206**	**-223**

- deuxièmement, la faiblesse de l'écart type de la variable *P/NS* sur le délai de développement 6 (3 370 contre 284 756 en délai 5, cf. tableau 7 ci dessus). Il s'agit là d'un problème moins structurel –c'est-à-dire lié à la formulation de la méthode proposée- et plus ponctuel qui traduit l'incertitude à laquelle nous nous exposons à mesure que les estimations s'appuient sur de moins en moins de données observées.

Dans le cas présent, une possibilité de s'affranchir de cet écueil pourrait être de définir un écart type « lissé » des variables I/NS ou P/NS à partir de l'évolution des écarts types de ces variables constatées sur les délais précédents (pour plus de détail, voir Mack (1993)*)*

3.2 Analyse en « Backtesting »

L'analyse graphique réalisée précédemment n'est pas suffisante pour affirmer que la méthode proposée apporte une amélioration certaine sur la qualité d'estimation de la charge ultime, même si nous avons constaté que dans certains cas l'estimation était améliorée. Pour tenter d'apporter d'autres éléments de réponse, nous avons cherché à tester la méthode sur les données observées des triangles, ceci permettant de comparer les valeurs estimées aux valeurs réelles.

L'idée est ici de partir de la première colonne du triangle initial et d'estimer ensuite à partir des méthodes SCL, MCL et « New Method » toutes les valeurs suivantes, dont nous connaissons à la date de l'étude le déroulement « ex post ».

Cela revient à faire les calculs suivants quelle que soit la méthode :

- sur les sinistres survenus :

$\hat{X}_{i,t} = X_{i,s} \times \hat{f}_{s->t}^{I,meth}$ lorsque *j* correspond aux données de première année

$\hat{X}_{i,t} = \hat{X}_{i,s} \times \hat{f}_{s->t}^{I,meth}$ pour les années suivantes

- sur les sinistres payés :

$\hat{P}_{i,t} = P_{i,s} \times \hat{f}_{s->t}^{P,meth}$ lorsque *j* correspond aux données de première année

$\hat{P}_{i,t} = \hat{P}_{i,s} \times \hat{f}_{s->t}^{P,meth}$ pour les années suivantes

Une fois ces calculs effectués, nous étudions la qualité d'estimation des méthodes à l'aide de la mesure de la racine carrée de l'erreur quadratique moyenne (sur les premiers délais de développement) - $RMSE\left(\hat{X}_{i,t}^{meth}\right) = \sqrt{\frac{\left(X_{i,t} - \hat{X}_{i,t}^{meth}\right)^2}{n-1}}$ -.

Les facteurs de développement selon la méthode SCL $\hat{f}_{s->t}^{P,SCL}$ et $\hat{f}_{s->t}^{I,SCL}$ sont dans le tableau de synthèse de la Partie 3 du paragraphe 1.2.3 équations (15) et (16).

Ceux du MCL $\hat{f}_{s->t}^{P,MCL}$ et $\hat{f}_{s->t}^{I,MCL}$ sont dans la Partie 2 paragraphe 3.5 équation (9).

Ceux de la nouvelle méthode proposée $\hat{f}_{s->t}^{P,NEW}$ et $\hat{f}_{s->t}^{I,NEW}$ sont dans la Partie 3 paragraphe 1.2.2 équations (11) et (12).

Les résultats sur les calculs des erreurs d'estimation des différentes méthodes au fil des délais de développement sont détaillés dans les tableaux 8 ci-dessous. La méthode SCL est en base 100. Les tableaux présentés traitent toujours des triangles Motor et assurance automobile DTA.

Tableaux 8 ; qualité des estimations des différentes méthodes sur les données passées

Ecart type des erreurs d'estimation sur triangles de réassurance Motor (y compris 2001)

Triangles des sinistres survenus

	délai 1	délai 2	délai 3	délai 4	délai 5	délai 6	délai 7	délai 8	délai 9
SCL		100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
MCL		92,7	89,1	85,7	86,7	83,1	73,8	69,8	82,1
NEW METHOD		93,3	89,6	88,7	90,9	89,6	84,4	82,6	93,8

Triangles des sinistres payés

	délai 1	délai 2	délai 3	délai 4	délai 5	délai 6	délai 7	délai 8	délai 9
SCL		100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
MCL		96,6	68,1	68,7	63,5	63,9	64,7	64,2	69,3
NEW METHOD		99,5	96,1	94,8	93,7	94,0	94,2	93,8	64,5

Ecart type des erreurs d'estimation sur triangles de réassurance Motor (hors 2001)

Triangles des sinistres survenus

	délai 1	délai 2	délai 3	délai 4	délai 5	délai 6	délai 7	délai 8	délai 9
SCL		100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
MCL		94,8	90,5	86,8	88,6	83,1	81,1	78,9	79,8
NEW METHOD		98,2	93,6	90,6	94,4	91,3	93,8	97,1	96,7

Triangles des sinistres payés

	délai 1	délai 2	délai 3	délai 4	délai 5	délai 6	délai 7	délai 8	délai 9
SCL		100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
MCL		88,8	94,7	100,7	102,6	100,7	100,1	104,9	110,7
NEW METHOD		99,4	105,1	104,4	102,5	107,3	109,4	110,1	110,9

Ecart type des erreurs d'estimation sur triangles d'assurance automobile DTA

Triangles des sinistres survenus

	délai 1	délai 2	délai 3	délai 4	délai 5	délai 6
SCL		100,0	100,0	100,0	100,0	100,0
MCL		108,6	115,8	111,9	113,2	112,9
NEW METHOD		90,9	89,6	94,1	80,1	64,1

Triangles des sinistres payés

	délai 1	délai 2	délai 3	délai 4	délai 5	délai 6
SCL		100,0	100,0	100,0	100,0	100,0
MCL		54,5	55,3	51,1	37,5	37,5
NEW METHOD		67,6	70,0	72,7	66,9	64,5

Clé de lecture : en assurance automobile, la méthode MCL voit son erreur d'estimation inférieur de 62,5% (37,5-100)/100 par rapport à la méthode SCL en délai 5.

Au vu des tableaux, la méthode MCL apparaît toujours comme celle qui s'adapte le mieux aux données de départ. Ceci est surtout vrai sur les triangles de paiements.

Il est intéressant de noter que lorsque nous travaillons sur le triangle Motor en retirant l'année 2001 de l'analyse, la méthode SCL est celle qui se comporte le mieux. Cela peut s'expliquer par le fait que plus les données tendent à s'homogénéiser (par exclusion des années aberrantes ou atypiques), moins il y a besoin d'avoir recours à des méthodes sophistiquées intégrant une information supplémentaire.

Les résultats concernant la qualité d'estimation de notre méthode n'apparaissent au final pas vraiment satisfaisants. Deux triangles sur trois permettent de mettre en évidence une amélioration des estimations mais lorsque l'on regarde les résultats sur l'ensemble des triangles étudiés (ANNEXES PARTIE 3 2), seuls 4 sur 8 mettent en évidence une amélioration à partir des triangles de paiements et 5 sur 8 à partir des triangles de sinistres survenus.

Il serait toutefois dommage de conclure à ce stade au faible intérêt représenté par l'intégration de l'information contenue dans la fréquence de sinistres. En effet, des liens ont bien été mis en évidence mais la manière de traiter ces relations n'est visiblement pas bien appréhendée. Si les variables I/P ou P/I semblent bien avoir un effet prépondérant dans l'évolution des facteurs de développement futurs, celles liées aux nombres de sinistres apportent alors peut-être une information que l'on pourrait qualifier de « second ordre ». L'idée est donc de tester ces deux variables de manière simultanée avec le recours à une approche multivariée.

3.3 Adaptation de la méthode

Comme nous l'avons dit, face aux résultats précédents décevants, la pertinence de cette méthode est discutable. Toutefois, il convient de préciser que même si cette méthode ne fonctionne pas dans toutes les circonstances, le fait qu'elle marche correctement dans certains cas et permette d'améliorer l'estimation par rapport à la méthode SCL est déjà un point positif et justifie qu'elle puisse être testée et approfondie.

Lors de l'analyse des corrélations (Partie 3, section 2.2), nous avons constaté que sur certains triangles, les variables $\frac{I_{i,s}}{P_{i,s}}$ et $\frac{I_{i,s}}{NS_{i,s}}$ (Q_1) étaient corrélées aux facteurs de développement sur payés $f_{s->t}^{P,SCL}$ mais qu'en revanche, elles ne l'étaient pas entre elles.

L'idée selon laquelle ces deux variables, $\frac{I_{i,s}}{P_{i,s}}$ et $\frac{I_{i,s}}{NS_{i,s}}$, pouvaient influencer l'évolution des facteurs de développement pour des raisons différentes est alors venue et nous avons cherché à la tester.

Les résultats sont décrits dans le tableau 9 (ci-dessous) où, à l'aide de régressions multiples, nous avons tenté d'évaluer l'apport simultané de ces deux variables dans l'explication des évolutions des facteurs de développement.

Tableau 9 : modélisation des facteurs de développement sur sinistres payés et survenus

variable expliquée	facteurs de développement sur payés			facteurs de développement sur survenus		
variables explicatives	**I/P**	**I/NS**	**R²**	**P/I**	**P/NS**	**R²**
réassurance						
Motor (inclus 2001)	0,28		0,07	0,19 (*)		0,03
		0,36	0,13		0,08 (*)	0,01
	0,27	0,35	0,19	0,25 (*)	-0,08 (*)	0,04
Motor (hors 2001)	0,31		0,09	0,26		0,06
		0,18	0,04		0,15 (*)	0,02
	0,31	0,18	0,12	0,37	-0,14 (*)	0,07
Marine	0,73		0,51	0,12 (*)		0,02
		0,03 (*)	0,00		0,03 (*)	0,00
	0,75	0,14 (*)	0,52	0,07 (*)	-0,02 (*)	0,00
Property per Risk	0,39		0,17	0,33		0,14
		0,1 (*)	0,01		0,06 (*)	0,00
	0,39	0,09	0,18	0,37	-0,01 (*)	0,15
Property Cat	0,48		0,23	0,12 (*)		0,02
		0,16 (*)	0,03		0,01 (*)	0,01
	0,53	-0,09 (*)	0,24	0,19 (*)	0,06 (*)	0,03
assurance						
automobile RC Matériels	0,51		0,29	0,01 (*)		0,00
		0,16 (*)	0,03		0,19	0,05
	0,57	-0,12 (*)	0,30	-0,02 (*)	0,23	0,06
automobile dommages	0,64		0,50	0,34		0,14
		0,75	0,42		0,33 (*)	0,08
	0,50	0,54	0,69	0,33	0,31 (*)	0,21
Multirisques habitations	0,61		0,00	0,03 (*)		0,00
		0,35	0,00		0,37	0,17
	0,67	-0,08 (*)	0,00	0,2 (*)	0,43	0,20

(*) coefficients non significatifs

Clé de lecture : dans le cas du triangle Motor (inclus 2001), le coefficient associé à la variable I/P (centrée réduite) dans le modèle de régression multiple visant à expliquer le facteur de développement des payés (centré réduit) est de 0,28 avec un R^2 de 0,07. Il est de 0,36 pour la variable I/NS, indépendamment de I/P avec un R^2 de 0,13. En revanche, ces deux coefficients sont respectivement de 0,27 et 0,35 dans un modèle multivarié avec un R^2 significativement meilleur de 0,19.

L'analyse des différentes régressions fait ressortir les points suivants :

- la mise en évidence de corrélations apparaît plus souvent dans le cas des facteurs de développement sur sinistres payés. Cela semble confirmer à la fois les commentaires faits précédemment et aussi que le principal intérêt de ces méthodes réside dans l'amélioration des facteurs de développement des sinistres payés.
- Nous constatons que le R^2 est dans de rares cas supérieur à 0,5. Cela n'empêche pas pour autant que les coefficients soient statistiquement significatifs (au seuil de 5%). Cette particularité peut également s'expliquer, comme nous l'avons déjà mentionné, par une relation non linéaire. Ce point apparaît clairement comme un axe d'étude intéressant qui pourrait être approfondi dans des travaux futurs.
- Il semble que dans 3 situations (Motor avec et sans 2001, automobile DTA), l'apport des deux variables explicatives, a priori non corrélées, améliore sensiblement l'explication des facteurs de développement sur sinistres payés. Ce dernier point est

crucial et il est absolument nécessaire de nous assurer de cette décorrélation avant l'application d'un modèle multivarié.

A la lumière de ces éléments, nous testons une nouvelle méthode (« NEW METHOD 2 » dans les tableaux et graphiques) sur les trois triangles (Motor avec et sans 2001, automobile DTA) alliant à la fois la corrélation entre les facteurs de développement des sinistres payés et les variables *I/P* et *I/NS*.

De manière plus générale, la formulation de cette seconde approche « NEW METHOD 2 » testée pour tenter d'améliorer l'estimation des facteurs de développement est la suivante :

$$\hat{f}_{s->t}^{P,NEW2} = \hat{f}_{s->t}^{P,SCL} + l_{MCL}^{P} \frac{\hat{s}\left(\frac{P_{i,t}}{P_{i,s}}\right)}{\hat{s}\left(\left(R_{i,s}^{MCL}\right)^{-1}\right)} \cdot \left(\left(R_{i,s}^{MCL}\right)^{-1} - \left(R_{s}^{MCL}\right)^{-1}\right) + l_{Q1}^{P} \frac{\hat{s}\left(\frac{P_{i,t}}{P_{i,s}}\right)}{\hat{s}\left(Q_{1i,s}\right)} \cdot \left(Q_{1i,s} - Q_{1s}\right) \quad \textbf{(éq. 1P)}$$

Et

$$\hat{f}_{s->t}^{I,NEW2} = \hat{f}_{s->t}^{I,SCL} + l_{MCL}^{I} \frac{\hat{s}\left(\frac{I_{i,t}}{I_{i,s}}\right)}{\hat{s}\left(\left(R_{i,s}^{MCL}\right)\right)} \cdot \left(\left(R_{i,s}^{MCL}\right) - \left(R_{s}^{MCL}\right)\right) + l_{Q_2}^{I} \frac{\hat{s}\left(\frac{I_{i,t}}{I_{i,s}}\right)}{\hat{s}\left(Q_{2i,s}\right)} \cdot \left(Q_{2i,s} - Q_{2s}\right) \quad \textbf{(éq. 2P)}$$

L'utilisation de ce modèle nécessite, comme nous venons de le préciser, que les variables explicatives ne soient pas corrélées entre elles (absence de multicolinéarité). L'étude de ces corrélations est illustrée au travers des tableaux 10 ci-dessous (testée sur l'ensemble des triangles) :

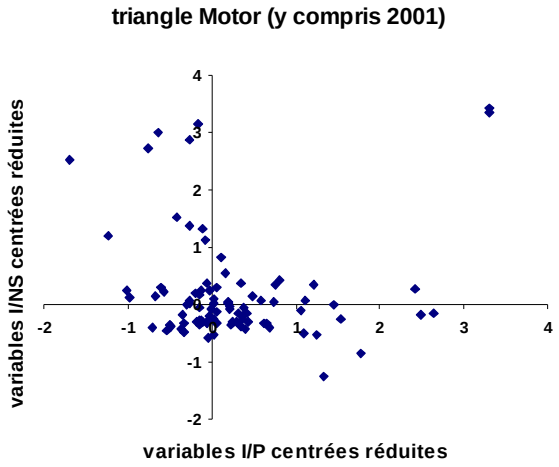
Tableaux 10 : analyse des corrélations entre variables explicatives

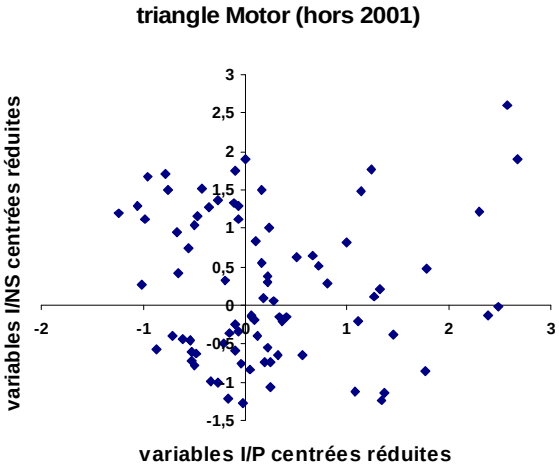
	facteurs de développement sur paiements		
	variable I/P	variable I/NS (Q1)	**corrélation entre I/P et Q1**
	l_{MCL}^{P}	$l_{Q_1}^{P}$	
triangles de réasssurance			
motor (inclus 2001)	0,28	0,36	**0,02**
motor (hors 2001)	0,32	0,20	**0,02**
marine	0,71	0,05	**-0,13**
Property per Risk	0,39	0,09	**0,01**
Property Cat	0,48	0,16	**0,49**
triangles d'assurance (marché français)			
Automobile - RC matériels	0,51	0,16	**0,49**
Automobile dommages - annuel	0,64	0,66	**0,33**
Multirisques habitations - annuel	0,32	0,20	**0,68**

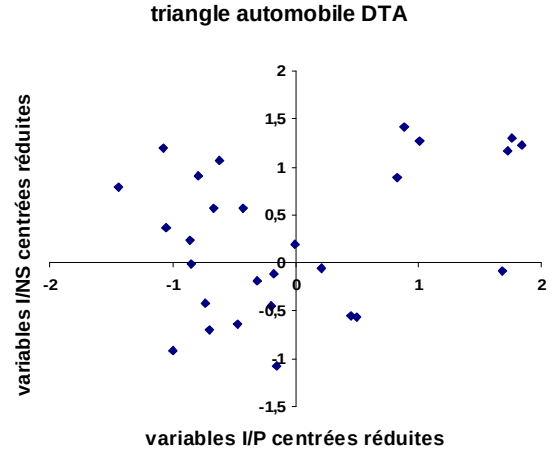
	facteurs de développement sur survenus		
	variable P/I	variable P/NS (Q2)	**corrélation entre P/I et Q2**
	l^{I}_{MCL}	$l^{I}_{Q_2}$	
triangles de réasssurance			
motor (inclus 2001)	0,19	0,08	**0,67**
motor (hors 2001)	0,26	0,14	**0,79**
marine	0,04	0,05	**0,47**
Property per Risk	0,37	0,01	**0,06**
Property Cat	0,16	-0,02	**-0,44**
triangles d'assurance (marché français)			
Automobile - RC matériels	0,00	0,23	**0,09**
Automobile dommages - annuel	0,34	0,29	**0,04**
Multirisques habitations - annuel	0,26	0,14	**-0,35**

La lecture des deux tableaux fait apparaître nos trois cas de figure intéressants où les variables $\frac{I_{i,s}}{P_{i,s}}$ et $\frac{I_{i,s}}{NS_{i,s}}$ sont à la fois corrélées aux facteurs de développement sur paiements et en même temps décorrélées l'une de l'autre.

Afin de nous assurer de manière plus certaine de l'absence de corrélation, nous confrontons les chiffres des tableaux 10 à une analyse graphique ci-dessous.







Les graphiques ci-dessus apportent bien une confirmation visuelle de l'absence de corrélation pour ces trois cas de figure.

Les résultats des régressions du tableau 9 ont montré que la variable *P/NS* n'améliore pas l'estimation des facteurs de développement des sinistres survenus dans nos trois cas. Nous décidons donc que la formule appliquée à l'estimation des facteurs de développement des sinistres survenus reste identique dans nos exemples précis à celle qu'utilise la méthode MCL.

L'ultime obtenu est toutefois différent de celui calculé par le MCL sur les sinistres survenus car il dépend de l'estimation faite des sinistres payés futurs via le ratio du MCL.

Le modèle général, décrit avec les équations 1P et 2P, se réécrit, spécifiquement à nos trois cas de figure, à l'aide des formules suivantes :

$$\hat{f}_{s->t}^{P,NEW2} = \hat{f}_{s->t}^{P,SCL} + l_{MCL}^{P} \frac{\hat{s}\left(\frac{P_{i,t}}{P_{i,s}}\right)}{\hat{s}\left(\left(R_{i,s}^{MCL}\right)^{-1}\right)} \cdot \left(\left(R_{i,s}^{MCL}\right)^{-1} - \left(R_{s}^{MCL}\right)^{-1}\right) + l_{Q1}^{P} \frac{\hat{s}\left(\frac{P_{i,t}}{P_{i,s}}\right)}{\hat{s}\left(Q_{1i,s}\right)} \cdot \left(Q_{1i,s} - Q_{1s}\right)$$

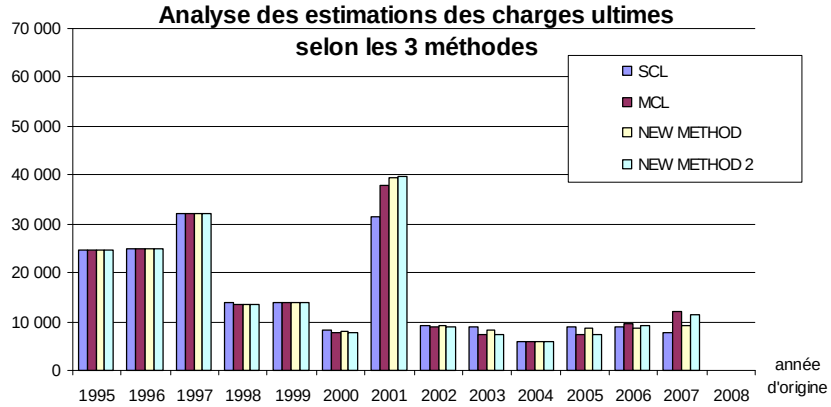
Et

$$\hat{f}_{s->t}^{I,NEW2} = \hat{f}_{s->t}^{I,SCL} + l_{MCL}^{I} \frac{\hat{s}\left(\frac{I_{i,t}}{I_{i,s}}\right)}{\hat{s}\left(\left(R_{i,s}^{MCL}\right)\right)} \cdot \left(\left(R_{i,s}^{MCL}\right) - \left(R_{s}^{MCL}\right)\right)$$

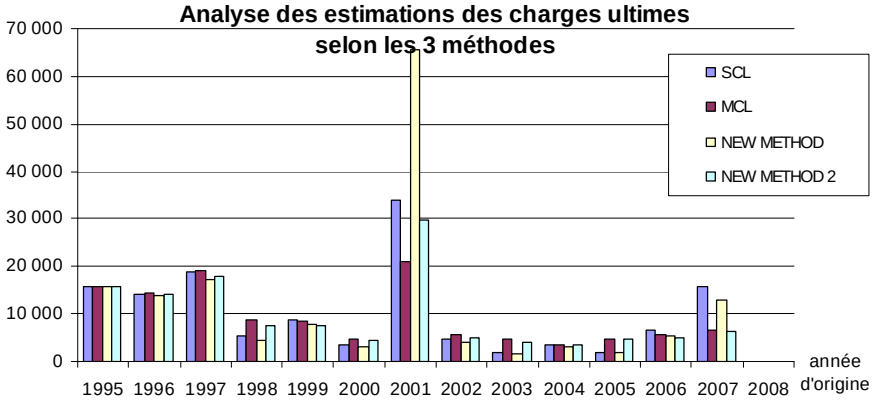
Les projections de charge ultime de sinistres effectuées à partir de cette méthode sont illustrées dans les graphiques suivants :

Tableaux 11a : Réassurance Motor (inclus 2001) : étude des projections à l'ultime

Estimations à l'ultime à partir des sinistres survenus



Estimations à l'ultime à partir des sinistres payés



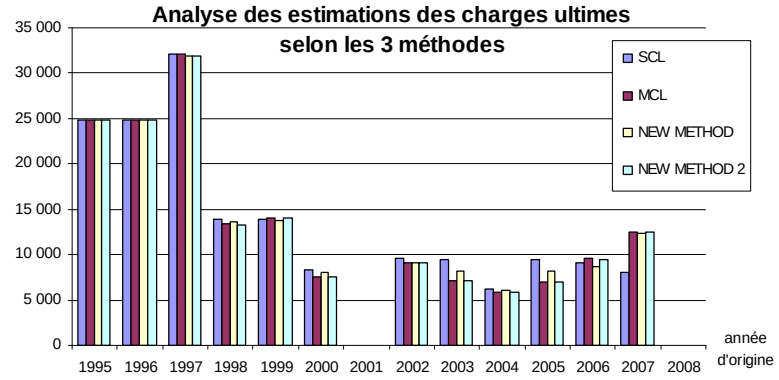
Pour les années antérieures à 2001, sur les estimations basées sur les sinistres survenus, la « New Method 2 » se comporte comme les autres. Leurs ultimes convergent. On note que sur ces années, pour les estimations basées sur les sinistres payés, la « New Method 2 » tend aussi à converger vers les méthodes SCL et MCL. Il y a donc une meilleure stabilité des méthodes.

Pour l'année 2001, on constate que l'estimation à l'ultime basée sur les payés de la « New Method 2 » parait moins aberrante que celui obtenu par la « New Method », le résultat converge vers celui du SCL et améliore le MCL. Dans ce cas, la variable de second ordre représentée par le nombre de sinistres améliore le MCL.

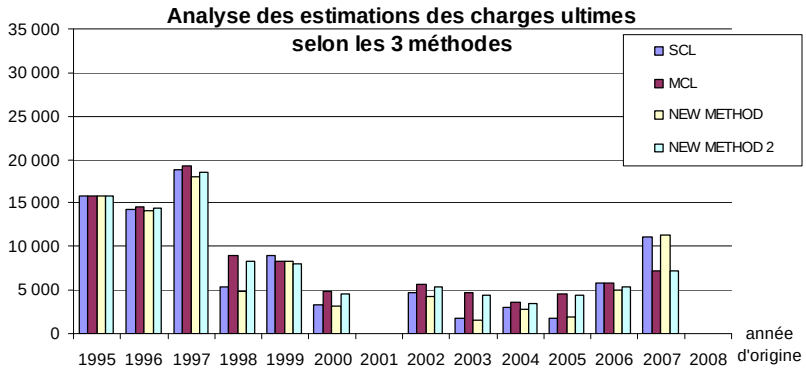
Sur les années postérieures à 2001, la variable de second ordre ne semble pas améliorer les résultats du MCL

Tableaux 11b : Réassurance Motor (hors 2001) : étude des projections à l'ultime

Estimations à l'ultime à partir des sinistres survenus



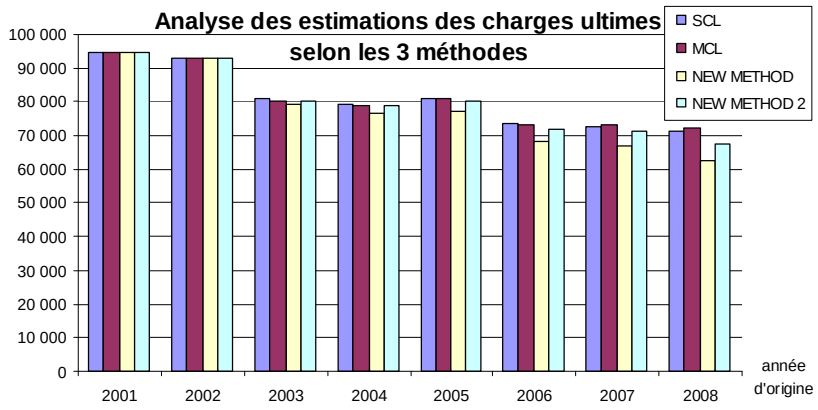
Estimations à l'ultime à partir des sinistres payés



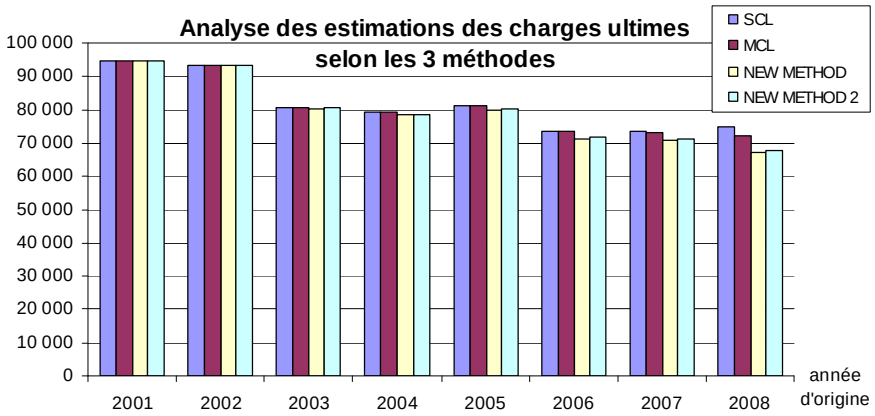
La convergence entre la méthode MCL et « NEW METHOD 2 » est encore plus flagrante lorsque l'on raisonne hors 2001

Tableaux 11c : Assurance automobile dommages tous accidents : étude des projections à l'ultime

Estimations à l'ultime à partir des sinistres survenus



Estimations à l'ultime à partir des sinistres payés



La « New Method 2 » converge plus fortement avec les autres méthodes même si elle semble toujours sous estimer la charge ultime.

De la même façon que lors de la première méthode proposée, cette méthode est soumise à un test d'adéquation des estimations sur les données connues des triangles. Les résultats sont repris dans les tableaux 12 ci-dessous :

Tableaux 12 ; qualité des estimations des différentes méthodes sur les données passées

Ecart type des erreurs d'estimation sur triangles de réassurance Motor (y compris 2001)

Triangles des sinistres survenus

	délai 1	délai 2	délai 3	délai 4	délai 5	délai 6	délai 7	délai 8	délai 9
SCL		100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
MCL		92,7	89,1	85,7	86,7	83,1	73,8	69,8	82,1
NEW METHOD		93,3	89,6	88,7	90,9	89,6	84,4	82,6	93,8
NEW METHOD 2		**92,7**	**89,1**	**86,4**	**87,5**	**84,3**	**76,7**	**73,9**	**85,2**

Triangles des sinistres payés

	délai 1	délai 2	délai 3	délai 4	délai 5	délai 6	délai 7	délai 8	délai 9
SCL		100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
MCL		96,6	68,1	68,7	63,5	63,9	64,7	64,2	69,3
NEW METHOD		99,5	96,1	94,8	93,7	94,0	94,2	93,8	64,5
NEW METHOD 2		**96,0**	**64,8**	**61,2**	**52,8**	**51,8**	**51,7**	**50,2**	**63,1**

Ecart type des erreurs d'estimation sur triangles de réassurance Motor (hors 2001)

Triangles des sinistres survenus

	délai 1	délai 2	délai 3	délai 4	délai 5	délai 6	délai 7	délai 8	délai 9
SCL		100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
MCL		94,8	90,5	86,8	88,6	83,1	81,1	78,9	79,8
NEW METHOD		98,2	93,6	90,6	94,4	91,3	93,8	97,1	96,7
NEW METHOD 2		**94,8**	**90,6**	**87,0**	**88,9**	**83,6**	**81,9**	**80,0**	**81,0**

Triangles des sinistres payés

	délai 1	délai 2	délai 3	délai 4	délai 5	délai 6	délai 7	délai 8	délai 9
SCL		100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
MCL		88,8	94,7	100,7	102,6	100,7	100,1	104,9	110,7
NEW METHOD		99,4	105,1	104,4	102,5	107,3	109,4	110,1	110,9
NEW METHOD 2		**88,0**	**100,0**	**103,9**	**103,4**	**105,7**	**107,4**	**112,7**	**116,1**

Ecart type des erreurs d'estimation sur triangles d'assurance automobile DTA

Triangles des sinistres survenus

	délai 1	délai 2	délai 3	délai 4	délai 5	délai 6	délai 7	délai 8
SCL		100,0	100,0	100,0	100,0	100,0	100,0	100,0
MCL		108,3	115,9	112,2	114,0	113,8	171,7	197,7
NEW METHOD		91,0	89,7	94,0	80,1	64,4	87,4	184,1
NEW METHOD 2		**108,3**	**103,2**	**102,0**	**99,7**	**85,4**	**112,3**	**109,6**

Triangles des sinistres payés

	délai 1	délai 2	délai 3	délai 4	délai 5	délai 6	délai 7	délai 8
SCL		100,0	100,0	100,0	100,0	100,0	100,0	100,0
MCL		54,9	55,7	51,4	38,0	37,9	37,7	36,6
NEW METHOD		68,2	70,6	73,2	67,5	65,3	62,4	74,3
NEW METHOD 2		**44,7**	**40,4**	**41,8**	**29,7**	**24,8**	**21,9**	**25,8**

Clé de lecture : en assurance automobile, NEW METHOD 2 voit son erreur d'estimation inférieur de 58,2% (41,8-100)/100 par rapport à la méthode SCL en délai 4

Au vu des tableaux, la méthode « New Method 2 » semble mieux fonctionner que la « New Method ». En effet, mis à part le triangle Motor où l'année 2001 a été exclue, les autres cas de figure mettent significativement en avant cette nouvelle méthode qui semble aussi bien s'adapter que celle de MCL. C'est vrai dans tous les cas sur les triangles de sinistres payés et survenus pour le triangle l'assurance automobile DTA. Sur le triangle de réassurance, la méthode du MCL est améliorée grâce à l'apport de l'information sur les fréquences pour l'année 2001.

Dans l'étude de nos exemples précis, aucune grosse amélioration sur le triangle des sinistres survenus n'est constatée par rapport à la méthode MCL. Ceci est par construction naturel car aucun terme de correction supplémentaire n'a été ajouté pour l'estimation des facteurs de développement des sinistres survenus. L'amélioration est due à la réestimation des sinistres payés qui intervient dans le calcul du Ratio *P/I*.

Dans l'évaluation de la qualité d'estimation de « NEW METHOD », nous avions mis en évidence un défaut que l'on pourrait qualifier de plus ou moins « constant » pour le dernier délai de développement quelle que soit l'année d'origine. Malgré ce travers a priori toujours présent dans le cas de l'assurance automobile DTA, la nouvelle méthode « NEW METHOD 2 » semble néanmoins, au regard des résultats contenus dans les tableaux 12, sensiblement améliorer la pertinence de l'estimation.

Il est intéressant de noter, dans nos cas de figure, que l'ajout simultané des deux variables permet d'intégrer des informations de nature différente, celles-ci étant nous l'avons vu faiblement voire pas du tout corrélées.

- Le terme de correction provenant de la variable *I/P* (ou *P/I*) s'apparente à une force de rappel de court terme en ce sens qu'elle corrige immédiatement l'année suivante l'écart constaté par rapport à la moyenne toutes années d'origine confondues.
- Le terme de correction provenant des variables *I/NS* (voire dans certains cas *P/NS*) tend, quant à lui, à produire des effets proches qui vont dans le même sens d'un délai de développement à l'autre pour une même année d'origine. L'ajustement engendré peut ainsi être perçu comme une force de rappel de moyen terme. A la lecture des graphiques de charge ultime projetée, cette force de rappel apparaît toutefois beaucoup plus faible que celle exercée par les variables *I/P* (ou *P/I*). Ceci montre qu'il s'agit d'une information dont l'apport est plutôt de second ordre.

Enfin, il convient de rappeler que l'application de cette méthode doit s'accompagner d'un contrôle strict des mécanismes d'ajustement conduisant à l'estimation ultime de la charge de sinistres. Dans notre cas, nous avons déjà mentionné la possibilité d'adapter la valeur de l'écart type de la variable exogène par un lissage pour les délais de développement avancés où le nombre d'observations est trop faible pour être réaliste. Une autre solution envisageable peut également être de tronquer le triangle pour l'analyse des corrélations et d'utiliser des coefficients plus standards pour l'estimation de la fin du développement de la sinistralité.

CONCLUSION

Dans le cadre du provisionnement, l'actuaire dispose d'un panel de méthodes pour effectuer ses estimations. Le premier objectif de ce mémoire était de montrer que parmi toutes celles-ci, chacune avait des qualités et des défauts. Afin de les détailler, nous les avons testées sur deux triangles assez différents. Un triangle d'assurance, simple qui se comporte assez bien, avec un développement rapide. Le second est un triangle de réassurance, moins régulier et se développant de façon beaucoup plus lente. Les caractères différents de ces deux triangles ont permis de souligner qu'une même méthode pouvait être appropriée dans un cas mais aberrante dans un autre. Ce constat, nous l'avons fait sur les résultats entre les deux triangles, mais aussi au sein d'un même triangle pour des années d'origine différentes.
Une des raisons de ces divergences repose sur l'utilisation partielle de l'information disponible. En général, les méthodes n'exploitent pas l'ensemble de celle-ci.

Nous avons utilisé des méthodes de crédibilité linéaire. Nous avons montré que le fait de combiner deux variables aléatoires via une crédibilisation permettait d'obtenir un estimateur de moindre variance. Dans la seconde Partie, nous avons mis en évidence le recours implicite aux principes de la théorie de la crédibilité, dans le cas par exemple des méthodes Bornhuetter-Ferguson, Cape Cod et Benktander, pourtant classées elles aussi comme « classiques ». Celles-ci cherchent en effet à combiner linéairement deux méthodes traditionnelles pour s'affranchir des défauts de chacune d'entre elles. Les résultats obtenus sur ces méthodes nous sont apparus plus stables et souvent de meilleure qualité.

En faisant le lien entre cette théorie et le modèle de régression linéaire, nous avons également montré que la méthode Munich Chain Ladder s'appuie sur ce même principe. Dans la forme, elle se distingue des méthodes précédentes en faisant appel à une information exogène connue, pour « crédibiliser » l'estimation des facteurs de développement. En effet, cette dernière crédibilise deux estimations basées sur la méthode SCL, l'une appliquée sur les triangles de sinistres payés et l'autre sur les survenus. Cette méthode est très novatrice, dans le sens où elle se base uniquement sur les informations existantes connues, sans avoir à utiliser des informations exogènes (de type Loss Ratio).

Le dernier objectif de ce mémoire consistait à tester et éventuellement proposer une nouvelle méthode de provisionnement. L'approche novatrice du MCL a été la source des développements que nous avons menés et présentés dans la troisième partie.

En effet, l'ensemble de nos travaux exploratoires a consisté selon la même approche à utiliser une information supplémentaire. Nous avons cherché dans un premier temps quelles pouvaient être les autres informations à la disposition d'un actuaire permettant d'améliorer son estimation de provision. Nous nous sommes concentrés sur les nombres de sinistres.

Ces réflexions ont abouti finalement à proposer un prolongement de la méthode Munich Chain Ladder dont nous avons vu qu'il pouvait significativement améliorer la qualité d'estimation, notamment sur les projections de flux futurs de paiements.

Des faiblesses de la méthode ont été identifiées en particulier lorsque l'on travaille sur les délais de développement les plus lointains, où par définition les historiques de données sont aussi les plus réduits. Ce type de défaut est évidemment présent dans bien d'autres méthodes. Il requiert de mener les analyses nécessaires en amont, lors des étapes intermédiaires de calcul, de manière à s'assurer au mieux de la robustesse des corrélations mises en évidence.

Enfin, en guise de conclusion, rappelons que nous nous sommes contraints dans le cadre de ce mémoire à rester sur une approche fondée sur la crédibilité linéaire. Toutefois, au cours des travaux, des suspicions de dépendances autres que linéaires sont parfois apparues. Celles-ci pourraient faire l'objet de développements futurs intéressants.

ANNEXES PARTIE 1

1 Présentation des données

1.1 Illustration des triangles

a. Triangle de développement décumulé

		Délai de développement *j*						
		0	…	j	j+1	…	…	N
Année d'origine i	1	$X_{1,0}$	…	$X_{1,j}$	$X_{1,j+1}$	…	…	$X_{1,N}$
	⋮	…	…	⋮	⋮	…	…	
	i	…	…	$X_{i,j}$	$X_{i,j+1}$	…		
	⋮	…	…	⋮	⋮			
	N-j	…	…	$X_{N-j,j}$				
	⋮	…						
	I	$X_{I,0}$						

Annexe 1 : Triangle de développement décumulé

b. Triangle de développement cumulé

		Délai de développement *j*						
		0	…	j	j+1	…	…	N
Année d'origine *i*	1	$C_{1,0}$	…	$C_{1,j}$	$C_{1,j+1}$	…	…	$C_{1,N}$
	⋮	…	…	⋮	⋮	…	…	
	i	…	…	$C_{i,j}$	$C_{i,j+1}$	…		
	⋮	…	…	⋮	⋮			
	I	$C_{I,0}$						

Annexe 2 : Triangle de développement cumulé

2 Présentation des méthodes

2.1 Les différents estimateurs de Chain Ladder

Pour $a = 0$, l'estimateur Chain Ladder correspond à la moyenne arithmétique des facteurs de développement individuel $f_{i,j}$.

$$\hat{f}_j = \frac{\sum_{i=1}^{I-j-1} w_{i,j}.C_{i,j}^0.f_{i,j}}{\sum_{i=1}^{I-j-1} w_{i,j}.C_{i,j}^0} = \frac{1}{(I-j-1)} \sum_{i=1}^{I-j-1} \frac{C_{i,j+1}}{C_{i,j}}$$

Pour $a = 1$, l'estimateur Chain Ladder correspond à la moyenne des facteurs de développement $f_{i,j}$ pondérée par les montants cumulés $C_{i,j}$. C'est-à-dire que le poids du développement des années ayant de gros montants de charge influe sur les facteurs de développement. On retrouve l'estimateur Chain Ladder standard.

$$\hat{f}_j = \frac{\sum_{i=1}^{I-j-1} w_{i,j}.C_{i,j}^1.f_{i,j}}{\sum_{i=1}^{I-j-1} w_{i,j}.C_{i,j}^1} = \sum_{i=1}^{I-j-1} \left(\frac{C_{i,j}}{\sum_{i=1}^{I-j-1} C_{i,j}} \right) \frac{C_{i,j+1}}{C_{i,j}}$$

Pour $a = 2$, l'estimateur Chain Ladder correspond au coefficient de la régression des $C_{i,j+1}$ en fonction des $C_{i,j}$ sans constante[6].

$$\hat{f}_j = \frac{\sum_{i=1}^{I-j-1} w_{i,j}.C_{i,j}^2.f_{i,j}}{\sum_{i=1}^{I-j-1} w_{i,j}.C_{i,j}^2} = \frac{\sum_{i=1}^{I-j-1} C_{i,j}.C_{i,j+1}}{\sum_{i=1}^{I-j-1} C_{i,j}^2}$$

[4] Parmi les extensions possibles, il est possible d'ajouter une constante dans la régression. On parle alors du modèle London Chain.

2.2 Les loss ratio ultimes

2.2.1 Les loss ratios ultimes du triangle d'assurance estimés en N

	1995	1996	1997	1998	2000	2001	2002	2003	2004
Chain Ladder									
$\alpha = 0$	81,5%	75,2%	80,7%	74,5%	69,2%	92,6%	67,7%	75,5%	61,4%
$\alpha = 1$	81,5%	75,2%	80,7%	74,5%	69,2%	92,6%	67,8%	75,5%	61,2%
$\alpha = 2$	81,5%	75,2%	80,7%	74,5%	69,2%	92,6%	67,8%	75,5%	61,0%
Loss Ratio	81,5%	75,2%	80,7%	74,5%	69,2%	95,0%	75,0%	75,0%	70,0%
coût moyen	81,5%	75,2%	80,7%	74,5%	69,2%	90,4%	69,2%	75,9%	62,4%
Craighead	81,5%	75,2%	80,7%	74,5%	69,2%	92,5%	68,1%	77,7%	48,3%
Sherman	81,5%	75,2%	80,7%	74,5%	69,2%	92,7%	67,9%	75,0%	62,2%
Hertig	81,5%	75,2%	80,7%	74,5%	69,2%	92,6%	67,7%	75,5%	61,5%
Bornhuetter Ferguson	81,5%	75,2%	80,7%	74,5%	69,2%	92,6%	67,9%	75,4%	65,1%
Cape Cod	81,5%	75,2%	80,7%	74,5%	69,2%	92,6%	67,9%	75,3%	66,6%
Benktander	81,5%	75,2%	80,7%	74,5%	69,2%	92,6%	67,8%	75,5%	62,9%
Generalized Linear Model	81,5%	75,2%	80,7%	74,5%	69,2%	92,6%	67,8%	75,5%	61,2%

Annexe 3 : loss ratios ultimes estimés en N selon les méthodes en assurance

2.2.2 Les loss ratios ultimes du triangle d'assurance estimés en N-1

	1995	1996	1997	1998	2000	2001	2002	2003
Chain Ladder								
$\alpha = 0$	81,5%	75,2%	80,7%	74,5%	69,5%	92,0%	69,1%	78,9%
$\alpha = 1$	81,5%	75,2%	80,7%	74,5%	69,5%	92,0%	69,1%	78,6%
$\alpha = 2$	81,5%	75,2%	80,7%	74,5%	69,4%	92,0%	69,0%	78,4%
Loss Ratio	81,5%	75,2%	80,7%	74,5%	70,0%	95,0%	75,0%	75,0%
coût moyen	81,5%	75,2%	80,7%	74,5%	68,9%	95,3%	74,3%	84,2%
Craighead	81,5%	75,2%	80,7%	74,5%	69,4%	92,6%	68,1%	77,8%
Sherman	81,5%	75,2%	80,7%	74,5%	69,4%	92,2%	69,0%	78,6%
Hertig	81,5%	75,2%	80,7%	74,5%	69,5%	92,0%	69,1%	79,0%
Bornhuetter Ferguson	81,5%	75,2%	80,7%	74,5%	69,5%	92,1%	69,5%	77,0%
Cape Cod	81,5%	75,2%	80,7%	74,5%	69,5%	92,0%	69,5%	77,1%
Benktander	81,5%	75,2%	80,7%	74,5%	69,5%	92,0%	69,1%	77,9%
Generalized Linear Model	81,5%	75,2%	80,7%	74,5%	69,5%	92,1%	69,1%	78,6%

Annexe 4 : loss ratios ultimes estimés en N-1 selon les méthodes en assurance

2.2.3 Les loss ratios ultimes du triangle de réassurance estimés en N

	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Chain Ladder														
$\alpha = 0$	111,5%	82,0%	80,4%	94,0%	77,9%	89,0%	91,0%	100,2%	84,8%	75,9%	61,9%	55,7%	73,5%	5,6%
$\alpha = 1$	111,5%	82,0%	80,6%	94,4%	78,6%	88,9%	90,6%	100,3%	84,9%	75,3%	61,0%	55,3%	72,2%	4,8%
$\alpha = 2$	111,5%	82,0%	80,6%	94,5%	78,6%	88,8%	90,3%	99,8%	84,4%	74,8%	60,2%	54,9%	73,0%	4,7%
Loss Ratio coût moyen	111,5%	82,0%	80,0%	95,0%	80,0%	90,0%	91,0%	99,0%	90,0%	90,0%	80,0%	80,0%	80,0%	80,0%
Craighead	111,5%	82,0%	80,2%	93,3%	77,0%	87,1%	88,7%	98,0%	82,8%	73,3%	60,2%	55,0%	71,7%	2,3%
Sherman	111,5%	82,2%	80,6%	94,0%	77,8%	88,4%	90,4%	100,1%	84,8%	74,8%	60,8%	54,2%	69,0%	4,7%
Hertig	111,5%	81,0%	79,4%	93,2%	77,2%	88,3%	90,2%	99,4%	84,1%	75,3%	61,4%	55,3%	73,1%	5,6%
Bornhuetter Ferguson	111,5%	82,0%	80,6%	94,5%	78,6%	88,9%	90,7%	100,2%	85,1%	76,2%	62,7%	58,9%	74,4%	62,5%
Cape Cod	111,5%	82,0%	80,6%	94,4%	78,6%	89,0%	90,7%	100,0%	84,3%	75,1%	62,0%	57,8%	72,2%	56,6%
Benktander	111,5%	82,0%	80,6%	94,4%	78,6%	88,9%	90,6%	100,3%	84,9%	75,3%	61,2%	55,8%	72,8%	49,0%
Generalized Linear Model	111,6%	81,0%	81,2%	96,1%	82,0%	87,1%	74,9%	78,0%	76,4%	68,9%	65,2%	47,2%	72,8%	4,5%

Annexe 5 : loss ratios ultimes estimés en N selon les méthodes en réassurance estimés en N

2.2.4 Les loss ratios ultimes du triangle de réassurance estimés en N-1

	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Chain Ladder													
$\alpha = 0$	111,5%	81,5%	79,6%	93,0%	78,8%	89,9%	89,9%	99,0%	85,5%	77,2%	65,3%	43,6%	108,9%
$\alpha = 1$	111,5%	81,5%	79,9%	93,9%	78,7%	89,7%	90,1%	99,1%	84,8%	76,5%	65,3%	44,1%	95,7%
$\alpha = 2$	111,5%	81,5%	79,9%	94,0%	78,7%	89,5%	89,8%	98,7%	84,2%	75,5%	64,8%	44,6%	92,0%
Loss Ratio coût moyen	111,5%	82,0%	80,0%	95,0%	80,0%	90,0%	91,0%	99,0%	90,0%	90,0%	80,0%	80,0%	80,0%
Craighead	111,5%	82,0%	80,2%	93,3%	77,1%	87,3%	89,0%	98,5%	83,5%	74,3%	61,4%	56,7%	75,1%
Sherman	111,5%	81,5%	79,6%	93,1%	78,1%	88,9%	89,5%	98,7%	84,0%	75,5%	62,9%	41,5%	91,0%
Hertig	111,5%	81,0%	79,4%	92,7%	78,6%	89,6%	89,6%	98,6%	85,2%	77,0%	65,2%	43,6%	110,1%
Bornhuetter Ferguson	111,5%	81,5%	79,9%	93,9%	78,7%	89,7%	90,1%	99,1%	85,1%	77,7%	67,6%	54,4%	83,5%
Cape Cod	111,5%	81,5%	79,9%	93,9%	78,7%	89,8%	90,2%	98,9%	84,3%	76,5%	67,1%	53,3%	80,8%
Benktander	111,5%	81,5%	79,9%	93,9%	78,7%	89,7%	90,1%	99,1%	84,8%	76,6%	65,7%	47,0%	86,3%
Generalized Linear Model	111,6%	80,9%	81,3%	96,2%	82,4%	86,8%	72,3%	73,7%	74,6%	67,6%	68,7%	42,6%	88,1%

Annexe 6 : loss ratios ultimes estimés en N-1 selon les méthodes en réassurance

2.3 Test de stabilité

2.3.1 Test de stabilité des méthodes sur le triangle d'assurance

	1995	1996	1997	1998	2000	2001	2002	2003
Chain Ladder								
$\alpha = 0$	0,0%	0,0%	0,0%	0,0%	-0,3%	0,5%	-1,3%	-3,4%
$\alpha = 1$	0,0%	0,0%	0,0%	0,0%	-0,3%	0,5%	-1,3%	-3,2%
$\alpha = 2$	0,0%	0,0%	0,0%	0,0%	-0,3%	0,6%	-1,3%	-2,9%
Loss Ratio	0,0%	0,0%	0,0%	0,0%	-0,8%	0,0%	0,0%	0,0%
coût moyen	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Craighead	0,0%	0,0%	0,0%	0,0%	-0,2%	0,0%	0,0%	0,0%
Sherman	0,0%	0,0%	0,0%	0,0%	-0,2%	0,5%	-1,1%	-3,6%
Hertig	0,0%	0,0%	0,0%	0,0%	-0,3%	0,5%	-1,3%	-3,5%
Bornhuetter Ferguson	0,0%	0,0%	0,0%	0,0%	-0,3%	0,5%	-1,5%	-1,5%
Cape Cod	0,0%	0,0%	0,0%	0,0%	-0,3%	0,5%	-1,6%	-1,7%
Benktander	0,0%	0,0%	0,0%	0,0%	-0,3%	0,5%	-1,3%	-2,4%
Generalized Linear Model	0,0%	0,0%	0,0%	0,0%	0,0%	0,5%	-1,3%	-3,2%

Annexe 7 : Test de stabilité des méthodes sur le triangle d'assurance

2.3.2 Test de stabilité des méthodes sur le triangle de réassurance

	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Chain Ladder													
$\alpha = 0$	0,0%	0,5%	0,7%	0,9%	-1,0%	-0,9%	1,1%	1,2%	-0,7%	-1,3%	-3,5%	12,1%	-35,4%
$\alpha = 1$	0,0%	0,5%	0,7%	0,6%	-0,1%	-0,8%	0,6%	1,2%	0,1%	-1,2%	-4,3%	11,1%	-23,5%
$\alpha = 2$	0,0%	0,5%	0,7%	0,5%	-0,1%	-0,8%	0,5%	1,2%	0,2%	-0,7%	-4,6%	10,3%	-19,0%
Loss Ratio	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
coût moyen	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Craighead	0,0%	0,0%	0,0%	-0,1%	-0,1%	-0,2%	-0,3%	-0,5%	-0,7%	-1,0%	-1,3%	-1,7%	-3,4%
Sherman	0,0%	0,7%	0,9%	0,9%	-0,2%	-0,5%	0,9%	1,4%	0,7%	-0,6%	-2,1%	12,7%	-21,9%
Hertig	0,0%	0,0%	0,0%	0,4%	-1,4%	-1,3%	0,6%	0,7%	-1,1%	-1,7%	-3,8%	11,7%	-37,0%
Bornhuetter Ferguson	0,0%	0,5%	0,7%	0,6%	-0,1%	-0,8%	0,6%	1,1%	0,0%	-1,5%	-5,0%	4,5%	-9,2%
Cape Cod	0,0%	0,5%	0,7%	0,6%	-0,1%	-0,8%	0,6%	1,2%	0,1%	-1,4%	-5,1%	4,4%	-8,5%
Benktander	0,0%	0,5%	0,7%	0,6%	-0,1%	-0,8%	0,6%	1,2%	0,1%	-1,2%	-4,5%	8,8%	-13,5%
Generalized Linear Model	0,0%	0,1%	-0,1%	-0,2%	-0,4%	0,3%	2,6%	4,3%	1,8%	1,4%	-3,5%	4,6%	-15,3%

Annexe 8 : Test de stabilité des méthodes sur le triangle de réassurance

2.4 La méthode de SHERMAN : les paramètres de lissage

Les paramètres sur le triangle d'assurance sont :

N	N-1
a = 0,04	a = 0,06
b=-1,26	b=-1,68
c=-1,9	c=-1,77
R^2=0,995	R^2=0,996

Les paramètres sur le triangle de réassurance sont :

N	N-1
a=0,25	a=0,25
b=1,85	b=1,74
c=-1,69	c=-1,71
R^2=0,85	R^2=0,96

N correspond aux paramètres obtenus avec la totalité du triangle.
N-1 correspond aux paramètres obtenus avec le triangle dont on a retiré la dernière diagonale.

2.5 La méthode de CRAIGHEAD : les paramètres de lissage

Les paramètres sur le triangle d'assurance sont :

N	N-1
b=1,08	b=1,04
c=1,29	c=1,24
R^2=0,942	R^2=0,943

Les paramètres sur le triangle de réassurance sont :

N	N-1
b=1,13	b=1,2
c=0,74	c=0,75
R^2=0,77	R^2=0,77

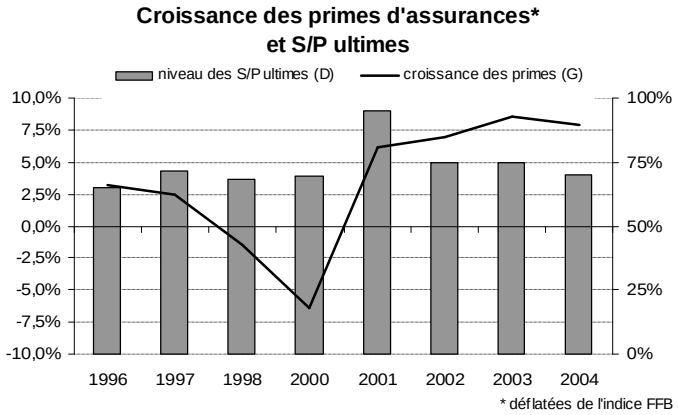
N correspond aux paramètres obtenus avec la totalité du triangle.
N-1 correspond aux paramètres obtenus avec le triangle dont on a retiré la dernière diagonale.

2.6 La méthode du LOSS RATIO : les S/P retenus

2.6.1 S/P retenus sur le triangle d'assurance

Quelques précisions sur les S/P retenus :

Le travail fondamental dans cette méthode réside dans le choix des S/P ultimes : dans notre cas, la société a fortement augmenté ses primes ces dernières années (graphique ci-contre), dû à la nécessité de rééquilibrer les déficits techniques passés (notamment 1999).
Au vu de cette évolution, nous retenons une diminution des S/P ultimes sur les années récentes, conséquence de ces réévaluations tarifaires (tableau ci-dessous).



UWY	**1995**	**1996**	**1997**	**1998**	**2000**	**2001**	**2002**	**2003**	**2004**
Loss Ratio	81%	75%	82%	75%	70%	95%	75%	75%	70%

Annexe 9 : Hypothèses des S/P Ultimes en assurance

2.6.2 S/P retenus sur le triangle de réassurance

UWY	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**
Loss Ratio	112%	75%	114%	75%	75%	75%	79%	75%	82%	75%	75%	75%	75%	75%

Annexe 10 : Hypothèses des S/P Ultimes en réassurance

2.7 La méthode GLM : les coefficients estimés

2.7.1 Coefficients sur le triangle d'assurance

estimation des coefficients - modèle GLM - triangle en N

UWY/ délai	0	1	2	3	4	5	6	7	8
c_0	6,749	//	//	//	//	//	//	//	//
α_i	//	-0,012	0,111	0,047	-0,003	0,348	0,104	0,293	0,159
β_j	//	-0,361	-2,601	-3,634	-4,154	//	//	//	//

estimation des coefficients - modèle GLM - triangle en N-1

UWY/ délai	0	1	2	3	4	5	6	7	8
c_0	6,741	//	//	//	//	//	//	//	//
α_i	//	-0,012	0,111	0,047	0,001	0,342	0,123	0,334	//
β_j	//	-0,361	-2,601	-3,634	-4,154	//	//	//	//

2.7.2 Coefficients sur le triangle de réassurance

estimation des coefficients - modèle GLM - triangle en N

UWY/ délai	0	1	2	3	4	5	6	7	8	9	10	11	12	13
c_0	5,038	//	//	//	//	//	//	//	//	//	//	//	//	//
α_i	//	3,058	3,055	3,394	4,343	4,008	3,563	3,937	5,185	5,6048	4,8475	2,2201	4,229	1,5056
β_j	//	1,131	1,303	1,370	1,415	1,434	1,429	1,391	1,363	1,351	1,379	1,395	1,410	1,398

estimation des coefficients - modèle GLM - triangle en N-1

UWY/ délai	0	1	2	3	4	5	6	7	8	9	10	11	12	13
c_0	5,031	//	//	//	//	//	//	//	//	//	//	//	//	//
α_i	//	3,057	3,056	3,396	4,348	4,005	3,528	3,881	5,161	5,585	4,8995	2,1186	4,4195	//
β_j	//	1,152	1,317	1,416	1,437	1,438	1,396	1,365	1,355	1,377	1,394	1,412	1,405	//

2.8 Comparaison des loss ratios ultimes en fonction de la méthode de cadencement utilisée

- Appliquée au triangle d'assurance :

	Chain Ladder					Bornhuetter Ferguson					Cape Cod				
	$\alpha=0$	$\alpha=1$	$\alpha=2$	*M*	*E*	$\alpha=0$	$\alpha=1$	$\alpha=2$	*M*	*E*	$\alpha=0$	$\alpha=1$	$\alpha=2$	*M*	*E*
1995	81,5%	81,5%	81,5%	*81,5%*	*0,0%*	81,5%	81,5%	81,5%	*81,5%*	*0,0%*	81,5%	81,5%	81,5%	*81,5%*	*0,0%*
1996	75,2%	75,2%	75,2%	*75,2%*	*0,0%*	75,2%	75,2%	75,2%	*75,2%*	*0,0%*	75,2%	75,2%	75,2%	*75,2%*	*0,0%*
1997	80,7%	80,7%	80,7%	*80,7%*	*0,0%*	80,7%	80,7%	80,7%	*80,7%*	*0,0%*	80,7%	80,7%	80,7%	*80,7%*	*0,0%*
1998	74,5%	74,5%	74,5%	*74,5%*	*0,0%*	74,5%	74,5%	74,5%	*74,5%*	*0,0%*	74,5%	74,5%	74,5%	*74,5%*	*0,0%*
1999	69,2%	69,2%	69,2%	*69,2%*	*0,0%*	69,2%	69,2%	69,2%	*69,2%*	*0,0%*	69,2%	69,2%	69,2%	*69,2%*	*0,0%*
2000	92,6%	92,6%	92,6%	*92,6%*	*0,0%*	92,6%	92,6%	92,6%	*92,6%*	*0,0%*	92,6%	92,6%	92,6%	*92,6%*	*0,0%*
2001	67,7%	67,8%	67,8%	*67,8%*	*0,0%*	67,9%	67,9%	68,0%	*67,9%*	*0,0%*	67,9%	67,9%	67,9%	*67,9%*	*0,0%*
2002	75,5%	75,5%	75,5%	*75,5%*	*0,0%*	75,4%	75,4%	75,4%	*75,4%*	*0,0%*	75,3%	75,3%	75,3%	*75,3%*	*0,0%*
2003	61,4%	61,2%	61,0%	*61,2%*	*0,2%*	65,3%	65,1%	65,0%	*65,1%*	*0,1%*	66,8%	66,6%	66,5%	*66,6%*	*0,1%*

Annexe 11 : Loss Ratio ultimes estimés en fonction de la méthode de cadencement en assurance

- Appliquée au triangle de réassurance :

	Chain Ladder					Bornhuetter Ferguson					Cape Cod				
	$\alpha=0$	$\alpha=1$	$\alpha=2$	*M*	*E*	$\alpha=0$	$\alpha=1$	$\alpha=2$	*M*	*E*	$\alpha=0$	$\alpha=1$	$\alpha=2$	*M*	*E*
1992	111,5%	111,5%	111,5%	*111,5%*	*0,0%*	111,5%	111,5%	111,5%	*111,5%*	*0,0%*	111,5%	111,5%	111,5%	*111,5%*	*0,0%*
1993	82,0%	82,0%	82,0%	*82,0%*	*0,0%*	82,0%	82,0%	82,0%	*82,0%*	*0,0%*	82,0%	82,0%	82,0%	*82,0%*	*0,0%*
1994	80,4%	80,6%	80,6%	*80,5%*	*0,1%*	80,4%	80,6%	80,6%	*80,5%*	*0,1%*	80,4%	80,7%	80,7%	*80,6%*	*0,2%*
1995	94,0%	94,4%	94,5%	*94,3%*	*0,3%*	94,0%	94,5%	94,5%	*94,3%*	*0,3%*	93,9%	94,3%	94,3%	*94,2%*	*0,3%*
1996	77,9%	78,6%	78,6%	*78,4%*	*0,4%*	77,9%	78,6%	78,7%	*78,4%*	*0,4%*	78,1%	79,0%	79,0%	*78,7%*	*0,5%*
1997	89,0%	88,9%	88,8%	*88,9%*	*0,1%*	89,1%	88,9%	88,8%	*88,9%*	*0,1%*	88,9%	88,8%	88,6%	*88,8%*	*0,1%*
1998	91,0%	90,6%	90,3%	*90,6%*	*0,3%*	91,0%	90,7%	90,3%	*90,7%*	*0,3%*	90,7%	90,4%	90,2%	*90,4%*	*0,3%*
1999	100,2%	100,3%	99,8%	*100,1%*	*0,2%*	100,1%	100,2%	99,8%	*100,1%*	*0,2%*	99,6%	99,7%	99,3%	*99,5%*	*0,2%*
2000	84,8%	84,9%	84,4%	*84,7%*	*0,2%*	85,0%	85,1%	84,7%	*84,9%*	*0,2%*	84,3%	84,4%	84,0%	*84,2%*	*0,2%*
2001	75,9%	75,3%	74,8%	*75,3%*	*0,5%*	76,9%	76,2%	75,7%	*76,3%*	*0,6%*	75,7%	75,2%	74,7%	*75,2%*	*0,5%*
2002	61,9%	61,0%	60,2%	*61,0%*	*0,8%*	63,7%	62,7%	61,7%	*62,7%*	*1,0%*	63,0%	62,1%	61,2%	*62,1%*	*0,9%*
2003	55,7%	55,3%	54,9%	*55,3%*	*0,4%*	59,5%	58,9%	58,5%	*58,9%*	*0,5%*	58,5%	57,9%	57,5%	*58,0%*	*0,5%*
2004	73,5%	72,2%	73,0%	*72,9%*	*0,6%*	75,4%	74,4%	75,0%	*74,9%*	*0,5%*	73,5%	72,5%	73,1%	*73,0%*	*0,5%*
2005	5,6%	4,8%	4,7%	*5,0%*	*0,5%*	65,1%	62,5%	61,9%	*63,2%*	*1,7%*	59,9%	57,4%	56,9%	*58,1%*	*1,6%*

Annexe 12 : Loss Ratio ultimes estimés en fonction de la méthode de cadencement en réassurance

ANNEXES PARTIE 2

3 La crédibilité dans les méthodes dites classiques

3.1 Démonstration de résultats intermédiaires dans le Bornhuetter-Ferguson

$$
\begin{aligned}
&COV\left(C_{i,N}, C_{i,j}.F_{i,j}\right) = F_{i,j} \cdot COV\left(C_{i,N}, C_{i,j}\right) \\
&= F_{i,j} \cdot COV\left(\left(C_{i,N} - C_{i,j}\right) + C_{i,j}, C_{i,j}\right) \\
&= F_{i,j} \cdot COV\left(\left(C_{i,N} - C_{i,j}\right), C_{i,j}\right) + F_{i,j} \cdot COV\left(C_{i,j}, C_{i,j}\right) \\
&= F_{i,j} \cdot 0 + F_{i,j} \cdot COV\left(C_{i,j}, C_{i,j}\right) = F_{i,j} \cdot Var\left(C_{i,j}\right) \\
&= F_{i,j} \cdot \frac{Var\left(C_{i,N}\right)}{F_{i,j}} = n^2
\end{aligned}
$$

3.2 Démonstration de résultats intermédiaires dans le Benktander

Démonstration sur la Covariance :

$$
\begin{aligned}
COV\left(\left(C_{i,N} - C_{i,j}.F_{i,j}\right), \left(C_{i,N} - \hat{D}_{i,j}\right)\right) &= COV\left(C_{i,N}, C_{i,N}\right) - COV\left(C_{i,j}.F_{i,j}, C_{i,N}\right) \\
&\quad - COV\left(C_{i,N}, \hat{D}_{i,j}\right) + COV\left(C_{i,j}.F_{i,j}, \hat{D}_{i,j}\right) \\
&= n^2 - n^2 - COV\left(C_{i,N}, \frac{1}{F_{i,j}} \cdot C_{i,j}.F_{i,j} + \left(1 - \frac{1}{F_{i,j}}\right) \cdot C_{i,N}\right) \\
&\quad + COV\left(C_{i,j}.F_{i,j}, \frac{1}{F_{i,j}} \cdot C_{i,j}.F_{i,j} + \left(1 - \frac{1}{F_{i,j}}\right) \cdot C_{i,N}\right) \\
&= \frac{1}{F_{i,j}} \cdot COV\left(C_{i,j}.F_{i,j}, C_{i,j}.F_{i,j}\right) + \left(1 - \frac{1}{F_{i,j}}\right) \cdot COV\left(C_{i,j}.F_{i,j}, C_{i,N}\right) \\
&\quad - \frac{1}{F_{i,j}} \cdot COV\left(C_{i,N}, C_{i,j}.F_{i,j}\right) - \left(1 - \frac{1}{F_{i,j}}\right) \cdot COV\left(C_{i,N}, C_{i,N}\right) \\
&= 0
\end{aligned}
$$

Démonstration sur la variance :

$$V\left(C_{i,N}-\hat{D}_{i,j}\right)=V\left(\frac{1}{F_{i,j}}\cdot\left(C_{i,N}-C_{i,j}.F_{i,j}\right)\right)+V\left(\left(1-\frac{1}{F_{i,j}}\right)\left(C_{i,N}-\hat{C}_{i,N}\right)\right)$$
$$+2\cdot\frac{1}{F_{i,j}}\cdot\left(1-\frac{1}{F_{i,j}}\right)\cdot COV\left(\left(C_{i,N}-C_{i,j}.F_{i,j}\right),\left(C_{i,N}-\hat{C}_{i,N}\right)\right)$$

On a vu dans la partie précédente sur la méthode du Bornhuetter-Ferguson que :

$$V\left(\frac{1}{F_{i,j}}\cdot\left(C_{i,N}-C_{i,j}.F_{i,j}\right)\right)=\left(\frac{1}{F_{i,j}}\right)^2\cdot n^2\cdot\left(F_{i,j}-1\right)$$
$$V\left(\left(1-\frac{1}{F_{i,j}}\right)\left(C_{i,N}-\hat{C}_{i,N}\right)\right)=\left(1-\frac{1}{F_{i,j}}\right)^2 n^2$$
$$COV\left(\left(C_{i,N}-C_{i,j}.F_{i,j}\right),\left(C_{i,N}-\hat{C}_{i,N}\right)\right)=0$$

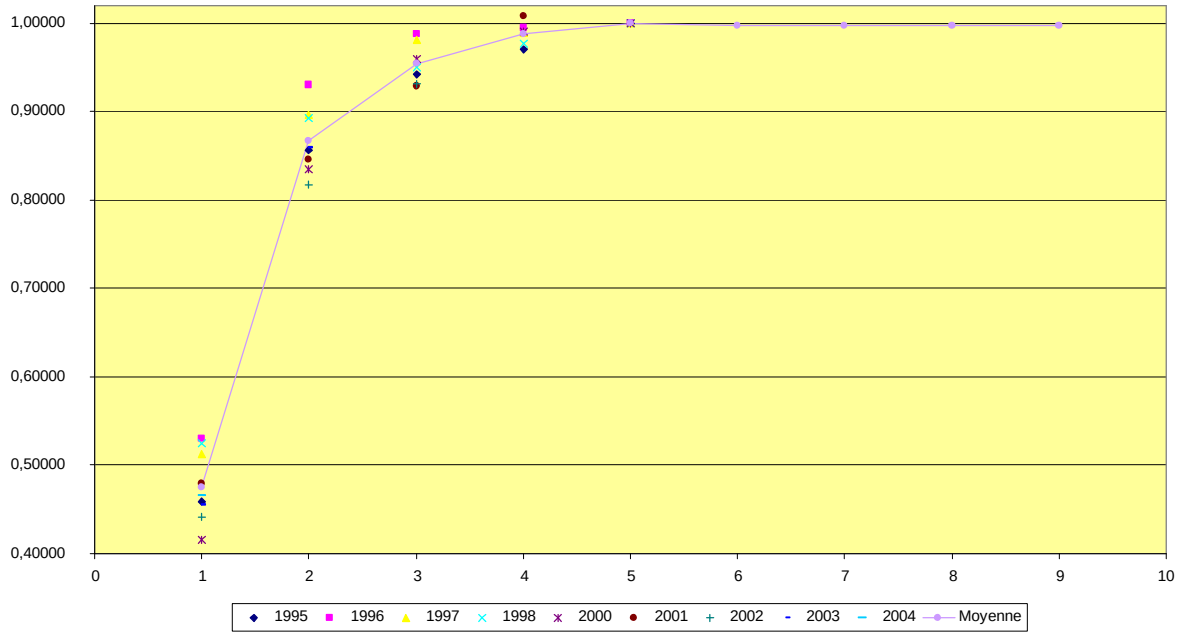
En remplaçant :

$$V\left(C_{i,N}-\hat{D}_{i,j}\right)=\left(\frac{1}{F_{i,j}}\right)^2\cdot n^2\cdot\left(F_{i,j}-1\right)+\left(1-\frac{1}{F_{i,j}}\right)^2 n^2$$
$$=n^2\cdot\left(\frac{1}{{F_{i,j}}^2}\cdot\left(\left(F_{i,j}-1\right)+\left(F_{i,j}-1\right)^2\right)\right)$$
$$=n^2\cdot\frac{1}{{F_{i,j}}^2}\cdot\left(F_{i,j}-1\right)\cdot\left(1+\left(F_{i,j}-1\right)\right)$$
$$=\frac{n^2}{F_{i,j}}\cdot\left(F_{i,j}-1\right)$$

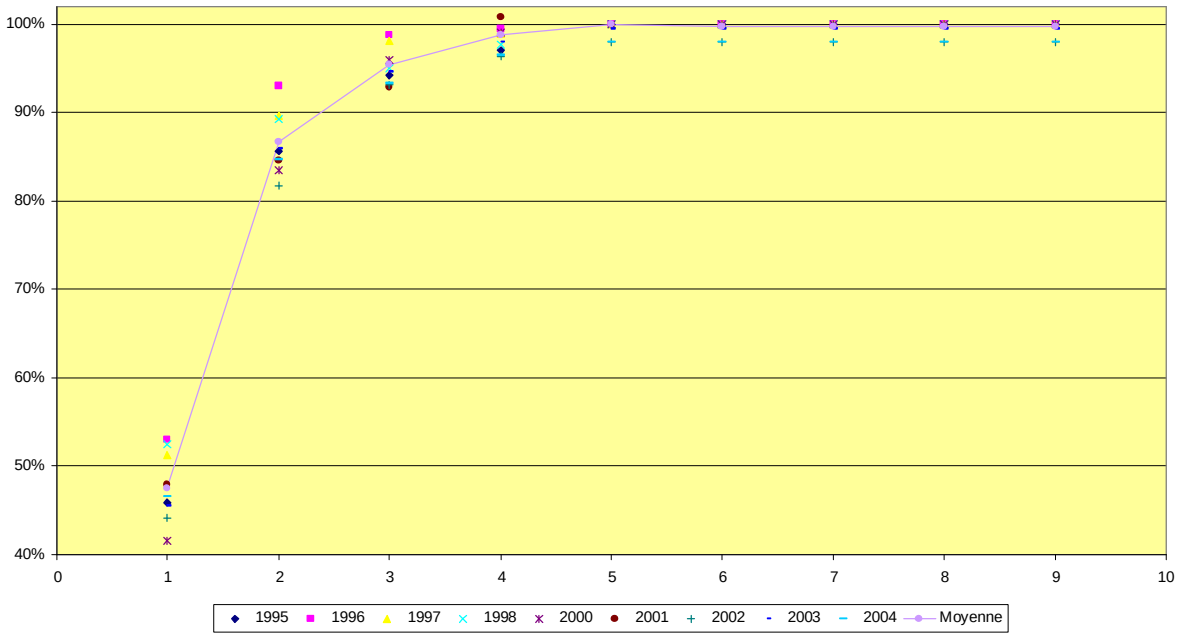
4 Le Munich Chain Ladder

4.1 Représentation graphique de l'évolution des $R_{i,j}^{MCL}$ sur le triangle d'assurance

Evolution des $R_{i,j}^{MCL}$ pour $j \leq a_{i-1}$ sur le triangle d'assurance :



Evolution des $R_{i,j}^{MCL}$ réels et projetés sur le triangle d'assurance :



4.2 Démonstration de M Merz et M V Wüthrich sur la crédibilité du MCL

Les variables aléatoires suivantes sont supposées de carré intégrable, définies dans l'espace probabilisé sous-jacent (Ω, F, P). Elles font donc partie de l'espace d'Hilbert $L^2(\Omega, F, \mathrm{P})$, comprenant le produit scalaire $\langle u, v \rangle = E[uv]$ et la norme $\|u\| = \sqrt{\langle u, u \rangle}$.

En reprenant les notations de payés et de survenus précédentes,

$$P_{i,j} \in L^2(\Omega, F, \mathrm{P}) \text{ et } I_{i,j} \in L^2(\Omega, F, \mathrm{P}).$$

Posons les deux s -algèbres :

$$D^P = s\{P_{i,j}; i + j \le I\} \subseteq F \text{ et } D^I = s\{I_{i,j}; i + j \le I\} \subseteq F$$

En reprenant les équations de projection des sinistres futures du Chain Ladder classique, puis les notations du MCL et en y intégrant les s -algèbres ci-dessus; elles peuvent être réécrites comme suit :

Le Chain Ladder :

$$U_{i,N}^{P\ CL} = E\left(P_{i,N} \middle| P_{i,1}, .. P_{i,j}\right) = E\left(P_{i,N} \middle| D^P\right) \text{ et } U_{i,N}^{I\ CL} = E\left(I_{i,N} \middle| I_{i,1}, .. I_{i,j}\right) = E\left(I_{i,N} \middle| D^I\right)$$

Les informations disponibles utilisées dans le MCL :

$\mathrm{P}_i(s) = \{P_{i,1}, \mathbf{L}, P_{i,s}\}$ peut être représenté par : $\mathrm{B}_s^P = s\{P_{i,j}; j < s\} \subseteq F$

$\mathrm{I}_i(s) = \{I_{i,1}, \mathbf{L}, I_{i,s}\}$ peut être représenté par : $\mathrm{B}_s^I = s\{I_{i,j}; j < s\} \subseteq F$

$\mathrm{B}_i(s) = \{P_{i,1}, \mathbf{L}, P_{i,s}, I_{i,1}, \mathbf{L}, I_{i,s}\}$ peut être représenté par : $\mathrm{B}_s = s\{P_{i,j}, I_{i,j}; j < s\} \subseteq F$

Soient les deux sous-espaces :

$$L_{i,j}^P = \left\{Z \in L^2(\Omega, F, P) \middle| Z = A_0 + A_1 \cdot I_{i,j-1} \quad avec \quad A_0, A_1 \quad B_{j-1}^P _ mesurable\right\}$$
$$L_{i,j}^I = \left\{Z \in L^2(\Omega, F, P) \middle| Z = A_0 + A_1 \cdot P_{i,j-1} \quad avec \quad A_0, A_1 \quad B_{j-1}^I _ mesurable\right\}$$

Afin de s'affranchir de l'hypothèse [PQ] du MCL[7], M Merz et M V Wüthrich montrent que cette hypothèse est équivalente à une crédibilisation, où les paiements cumulés et les sinistres survenus cumulés sont approximés par le meilleur estimateur à l'intérieur des sous-ensembles suivants :

$$\overline{L}_{i,j}^{P} = \left\{ Z \in L^2(\Omega, F, P) \middle| \exists (Z_n)_{n\in\mathrm{N}} \subseteq L_{i,j}^{P} \quad avec \quad \lim_{n\to\infty} \|Z - Z_n\| = 0 \right\}$$

$$\overline{L}_{i,j}^{I} = \left\{ Z \in L^2(\Omega, F, P) \middle| \exists (Z_n)_{n\in\mathrm{N}} \subseteq L_{i,j}^{I} \quad avec \quad \lim_{n\to\infty} \|Z - Z_n\| = 0 \right\} \qquad [**]$$

$\overline{L}_{i,j}^{P}$ et $\overline{L}_{i,j}^{I}$ sont des sous-ensembles clos.

Ils estiment $P_{i,j}$ à l'aide du meilleur estimateur de carré intégrable qui est la limite d'une suite $(Z_n)_{n\in\mathrm{N}} \subseteq L_{i,j}^{I}$ des estimations linéaires affines dans $I_{i,j-1}$, sachant que $L_{i,j}^{P}$ ne contient pas $I_{i,j-1}$ pour $l \neq j$. De plus, A_0 et A_1 étant des variables aléatoires $B_{j-1}^{P}_mesurable$, $P_{i,j}$ est estimé étant donné l'information B_{j-1}^{P}. Le même cheminement est fait pour $I_{i,j}$ et $L_{i,j}^{I}$.

Observons la moyenne du carré des écarts (msep) et minimisons la afin de trouver nos estimateurs optimaux ($Z_{i,j}^{P} \in \overline{L}_{i,j}^{P}$ et $Z_{i,j}^{I} \in \overline{L}_{i,j}^{P}$):

$$msep_{P_{i,j}}\left(Z_{i,j}^{P}\right) = \inf_{Z\in\overline{L}_{i,j}^{P}} msep_{P_{i,j}}(Z) = \inf_{Z\in\overline{L}_{i,j}^{P}} E\left(Z - P_{i,j}\right)$$

et

$$msep_{I_{i,j}}\left(Z_{i,j}^{I}\right) = \inf_{Z\in\overline{L}_{i,j}^{I}} msep_{P_{i,j}}(Z) = \inf_{Z\in\overline{L}_{i,j}^{I}} E\left(Z - I_{i,j}\right)$$

$Z_{i,j}^{P}$ et $Z_{i,j}^{I}$ sont les estimateurs linéaires affines optimaux de $P_{i,j}$ et $I_{i,j}$, sachant respectivement B_{j-1}^{P} et B_{j-1}^{I}. Ils sont uniques car $\overline{L}_{i,j}^{P}$ et $\overline{L}_{i,j}^{I}$ sont des sous-ensembles clos de l'espace d'Hilbert $L^2(\Omega, F, \mathrm{P})$ et $P_{i,j}, I_{i,j} \in L^2(\Omega, F, \mathrm{P})$ (Théorème 2.3.1 de Brockwell & Davis).

De plus d'après le théorème de Bühlmann & Gisler :

$$E\left[\left(P_{i,j} - Z_{i,j}^{P}\right)\cdot Z\right] = 0 \qquad \forall Z \in \overline{L}_{i,j}^{P} \qquad [1]$$

Une équation équivalente est obtenue pour $Z_{i,j}^{I}$.

[7] Dans leur article il s'agit de l'hypothèse [A3]

A partir de cette équation, M Merz et M V Wüthrich donnent une représentation des $Z_{i,j}^{P}$ $Z_{i,j}^{I}$. Sous les hypothèses PU, IU, PE, PV, IE et IV[8] du MCL les estimateurs linéaires affines optimaux sont tels que :

$$Z_{i,j}^{P} = f_{j-1}^{P} \cdot P_{i,j-1} + Corr\left[P_{i,j}, I_{i,j-1} \middle| B_{j-1}^{P}\right] \cdot Var\left[P_{i,j} \middle| B_{j-1}^{P}\right]^{1/2} \cdot \frac{I_{i,j-1} - E\left[I_{i,j-1} \middle| B_{j-1}^{P}\right]}{Var\left[I_{i,j-1} \middle| B_{j-1}^{P}\right]^{1/2}}$$

et

$$Z_{i,j}^{I} = f_{j-1}^{I} \cdot I_{i,j-1} + Corr\left[I_{i,j}, P_{i,j-1} \middle| B_{j-1}^{I}\right] \cdot Var\left[I_{i,j} \middle| B_{j-1}^{I}\right]^{1/2} \cdot \frac{P_{i,j-1} - E\left[P_{i,j-1} \middle| B_{j-1}^{I}\right]}{Var\left[P_{i,j-1} \middle| B_{j-1}^{I}\right]^{1/2}}$$

pour $O \le i \le I \quad et \quad O \le j \le N$

Démonstration pour les sinistres payés :

Posons

$$Z_{i,j}^{P} = f_{j-1}^{P} \cdot P_{i,j-1} + Corr\left[P_{i,j}, I_{i,j-1} \middle| B_{j-1}^{P}\right] \cdot Var\left[P_{i,j} \middle| B_{j-1}^{P}\right]^{1/2} \cdot \frac{I_{i,j-1} - E\left[I_{i,j-1} \middle| B_{j-1}^{P}\right]}{Var\left[I_{i,j-1} \middle| B_{j-1}^{P}\right]^{1/2}} \qquad [2]$$

alors

$Z_{i,j}^{P} \in \overline{L}_{i,j}^{P}$ ce qui implique l'équation [1].

Fixons arbitrairement un Z donné tel que $Z \in \overline{L}_{i,j}^{P}$.
Grâce aux propriétés du produit scalaire de l'espace d'Hilbert $L^2(\Omega, F, \mathrm{P})$, nous avons :

$$E\left[\left(P_{i,j} - Z_{i,j}^{P}\right) \cdot Z\right] = \lim_{n \to \infty} E\left[\left(P_{i,j} - Z_{i,j}^{P}\right) \cdot Z_n\right]$$

avec $(Z_n)_{n \in \mathrm{N}} \subseteq L_{i,j}^{P} \quad avec \quad \lim_{n \to \infty} \|Z - Z_n\| = 0$.

Comme $Z_n = A_{n,0} + A_{n,1} \cdot I_{i,j-1} \quad avec \quad A_{n,0}, A_{n,1} \quad B_{j-1}^{P} _ mesurable \quad \forall n \in N$

Alors :

$$\begin{aligned} E\left[\left(P_{i,j} - Z_{i,j}^{P}\right) \cdot Z_n\right] &= E\left[\left(P_{i,j} - Z_{i,j}^{P}\right) \cdot A_{n,0}\right] + E\left[\left(P_{i,j} - Z_{i,j}^{P}\right) \cdot A_{n,1} \cdot I_{i,j-1}\right] \\ &= E\left[A_{n,0} \cdot E\left[\left(P_{i,j} - Z_{i,j}^{P}\right) \middle| B_{j-1}^{P}\right]\right] + E\left[A_{n,1} \cdot E\left[\left(P_{i,j} - Z_{i,j}^{P}\right) \cdot I_{i,j-1} \middle| B_{j-1}^{P}\right]\right] \end{aligned}$$

En remplaçant le premier terme de l'équation par le résultat de [2], ce terme devient nul.

[8] Dans leur article il s'agit des hypothèses [A1] et [A2].

Pour le second terme :

$$
\begin{aligned}
&E\left[\left(P_{i,j} - Z_{i,j}^{P}\right)\cdot I_{i,j-1}\middle|B_{j-1}^{P}\right] \\
&= E\left[\left(P_{i,j}\cdot I_{i,j-1}\right)\middle|B_{j-1}^{P}\right] - E\left[\left(P_{i,j}\right)\middle|B_{j-1}^{P}\right]\cdot E\left[\left(I_{i,j-1}\right)\middle|B_{j-1}^{P}\right] \\
&- \frac{Cov\left[\left(P_{i,j}\cdot I_{i,j-1}\right)\middle|B_{j-1}^{P}\right]}{Var\left[\left(I_{i,j-1}\right)\middle|B_{j-1}^{P}\right]}\cdot E\left[\left(I_{i,j-1} - E\left[\left(I_{i,j-1}\right)\middle|B_{j-1}^{P}\right]\right)\cdot\left(I_{i,j-1}\right)\middle|B_{j-1}^{P}\right] \\
&= Cov\left[\left(P_{i,j}\cdot I_{i,j-1}\right)\middle|B_{j-1}^{P}\right] - \frac{Cov\left[\left(P_{i,j}\cdot I_{i,j-1}\right)\middle|B_{j-1}^{P}\right]}{Var\left[\left(I_{i,j-1}\right)\middle|B_{j-1}^{P}\right]}\cdot Var\left[\left(I_{i,j-1}\right)\middle|B_{j-1}^{P}\right] \\
&= 0
\end{aligned}
$$

Ainsi : $\forall n \in N, \qquad E\left[\left(P_{i,j} - Z_{i,j}^{P}\right)\cdot Z_n\right]$ et donc $E\left[\left(P_{i,j} - Z_{i,j}^{P}\right)\cdot Z\right]$

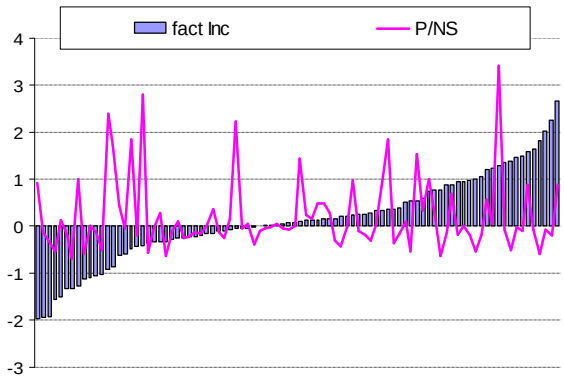
[**] a donc bien été prouvé.

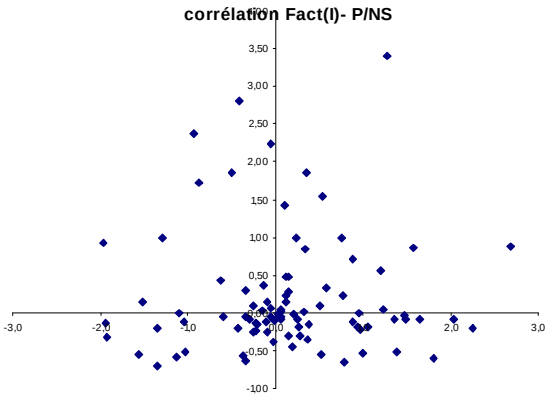
On reconnaît dans ces sous-ensembles une forme de crédibilité. Le MCL est donc bien une crédibilisation linéaire.

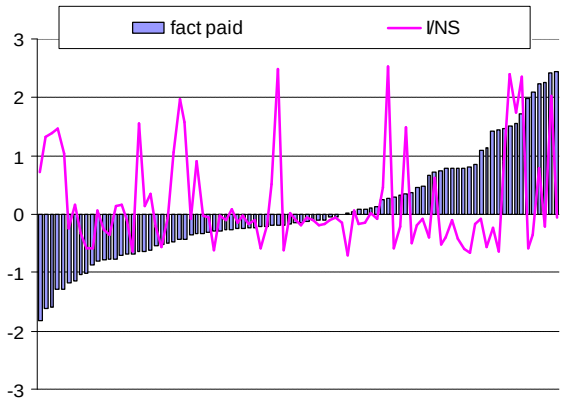
ANNEXES PARTIE 3

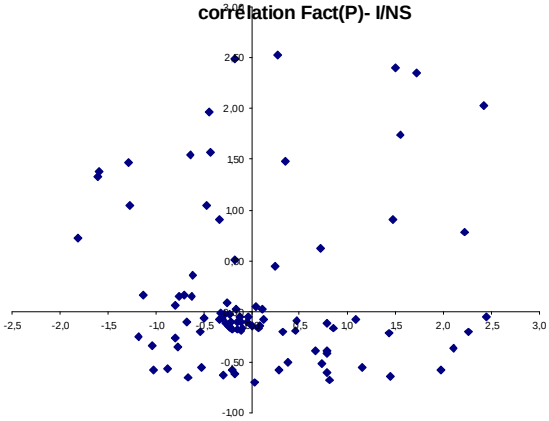
1 Graphiques de corrélation des autres triangles et projections ultimes

Triangle de réassurance : Marine

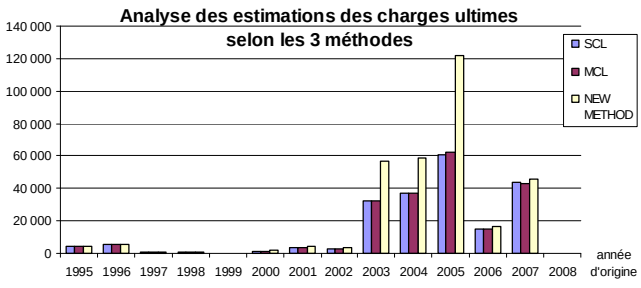




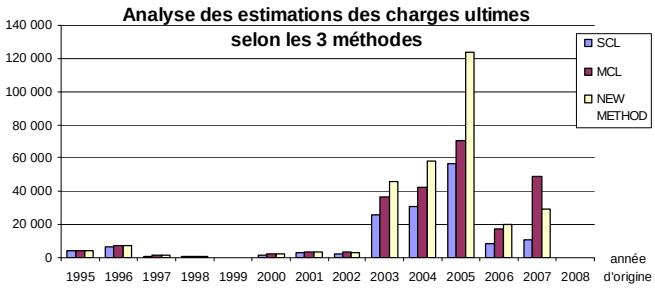




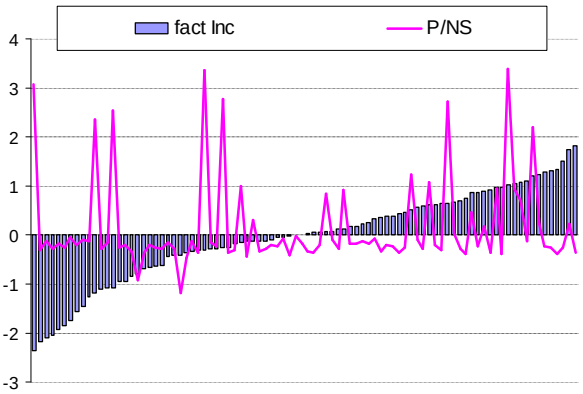
Estimation sur triangles des sinistres survenus

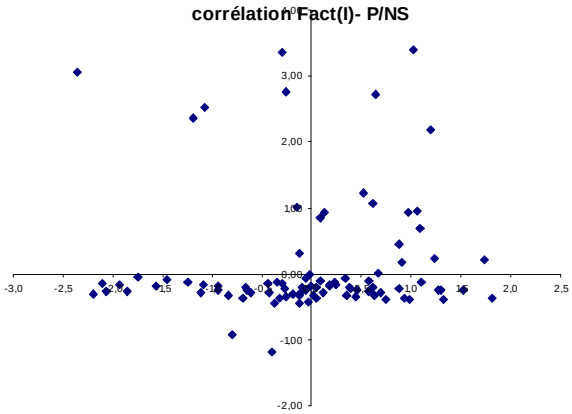


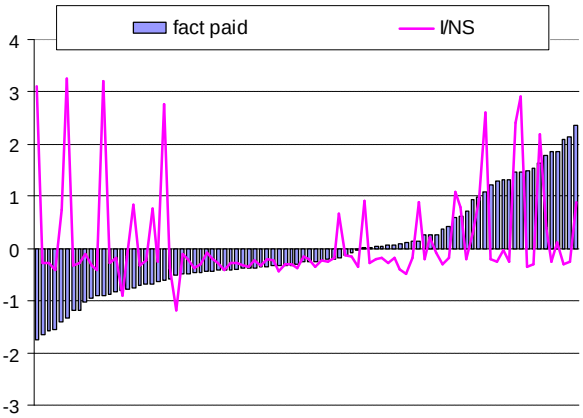
Estimation sur triangles des sinistres payés

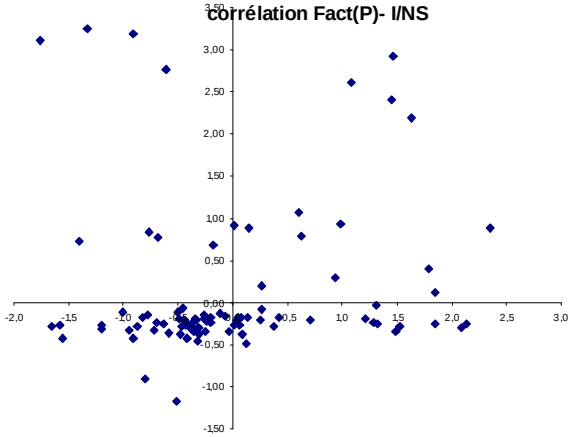


Triangle de réassurance : Property per Risk

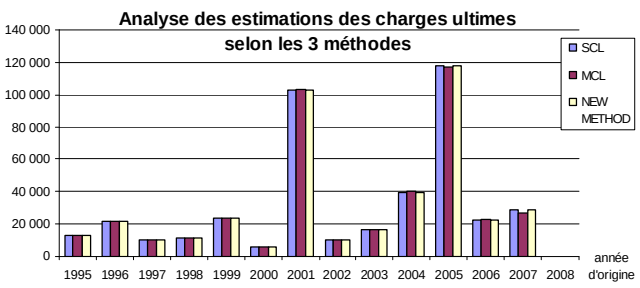




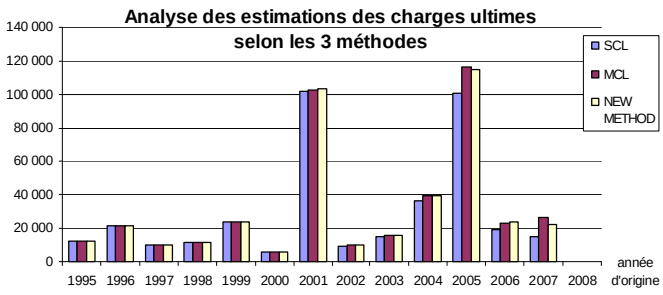




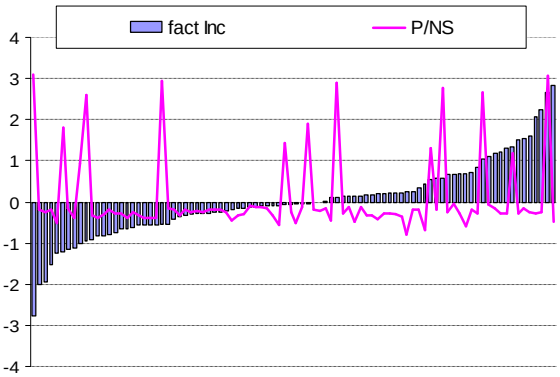
Estimation sur triangles des sinistres survenus

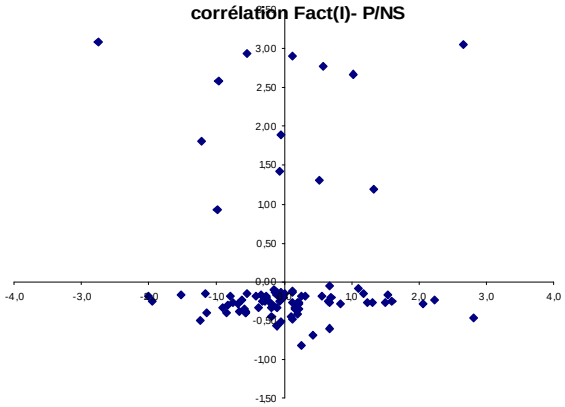


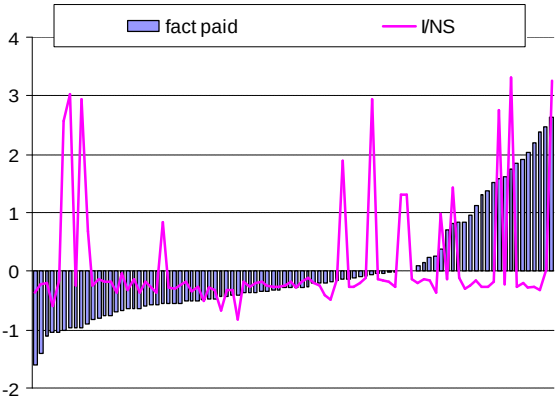
Estimation sur triangles des sinistres payés

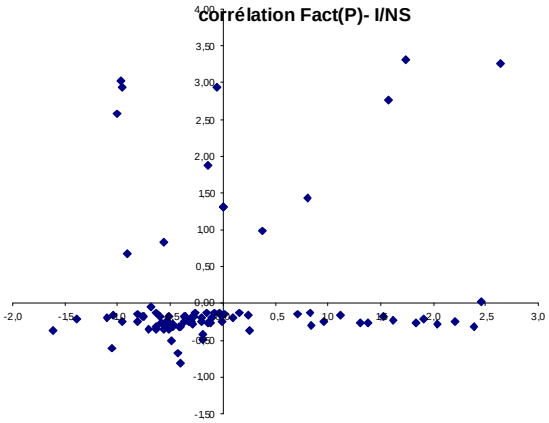


Triangle de réassurance : Property Cat

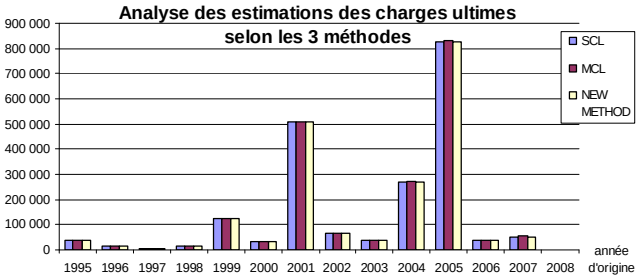




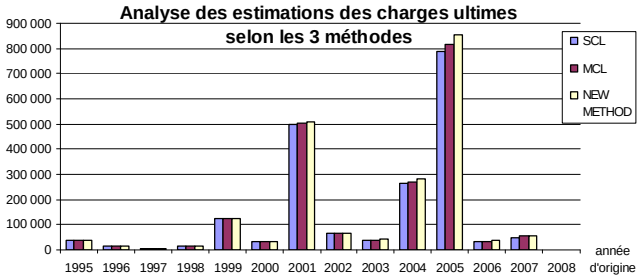




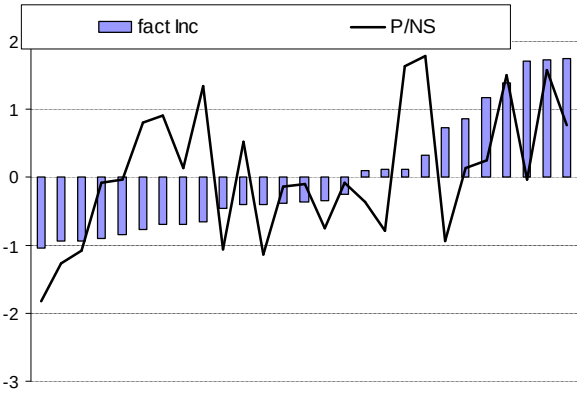
Estimation sur triangles des sinistres survenus

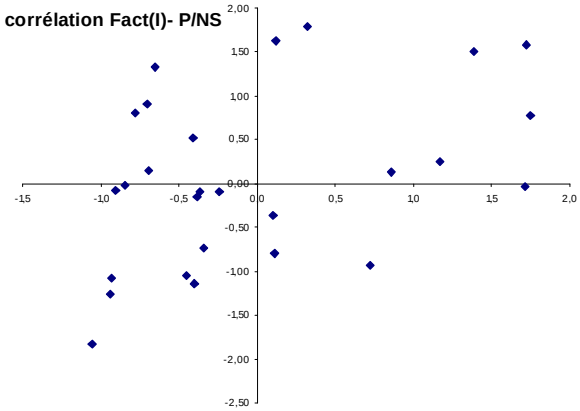


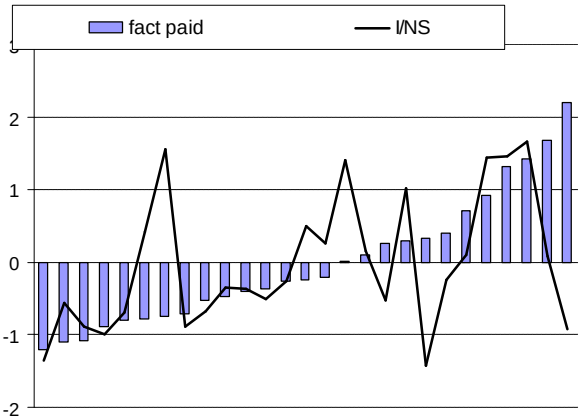
Estimation sur triangles des sinistres payés

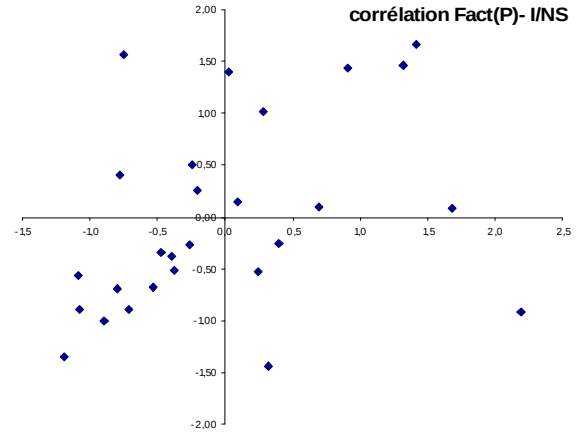


Triangle d'assurance : Multirisques habitations

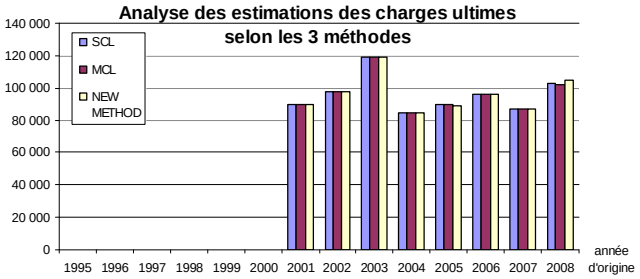




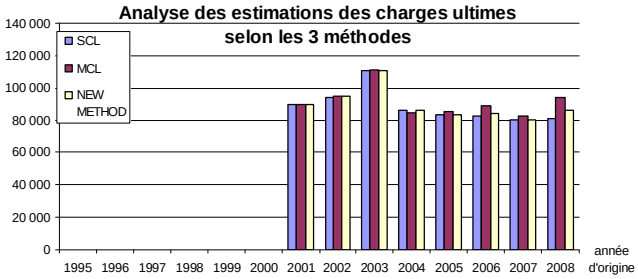




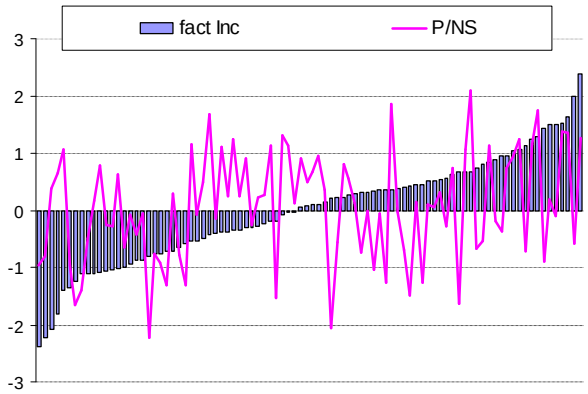
Estimation sur triangles des sinistres survenus

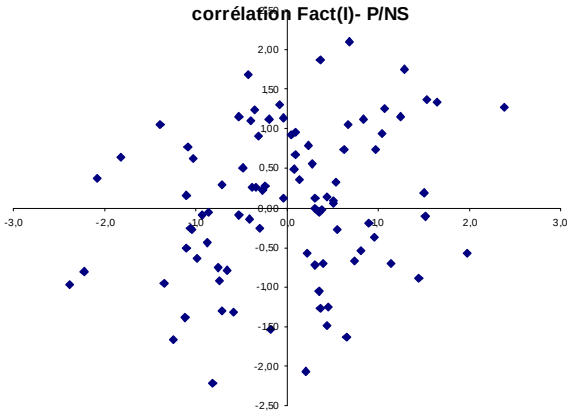


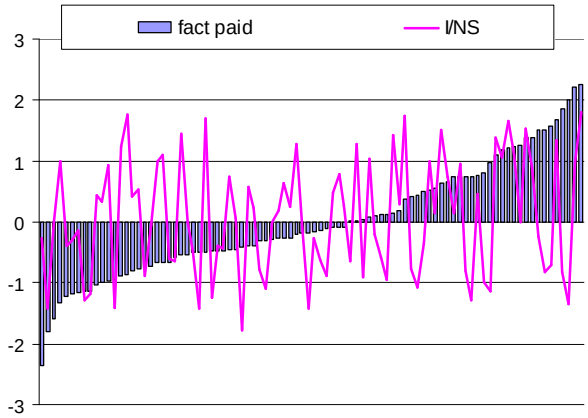
Estimation sur triangles des sinistres payés

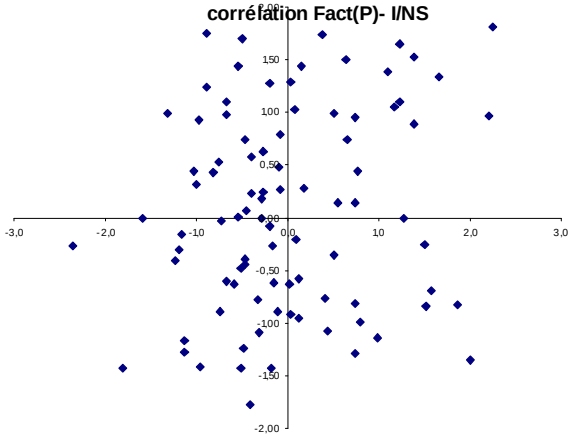


Triangle d'assurance : Responsabilité civile Matériels

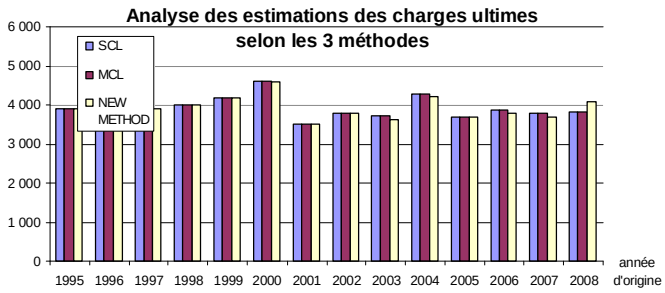




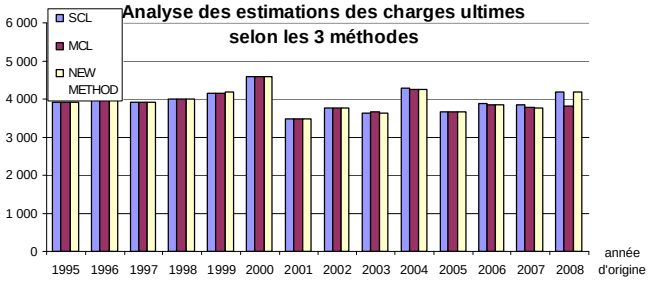




Estimation sur triangles des sinistres survenus



Estimation sur triangles des sinistres payés



2 Qualité des estimations des différentes méthodes sur les autres triangles

Ecart type des erreurs d'estimation sur triangles de réassurance Marine

Triangles des sinistres survenus

	délai 1	délai 2	délai 3	délai 4	délai 5	délai 6	délai 7	délai 8	délai 9
SCL		100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
MCL		99,1	96,5	94,7	95,4	93,4	91,8	92,1	89,0
NEW METHOD		99,9	96,0	92,4	88,6	69,1	98,7	98,5	96,6

Triangles des sinistres payés

	délai 1	délai 2	délai 3	délai 4	délai 5	délai 6	délai 7	délai 8	délai 9
SCL		100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
MCL		46,4	20,5	16,0	30,9	15,6	12,7	11,5	10,9
NEW METHOD		100,1	99,9	99,7	101,0	101,2	94,6	94,4	94,3

Ecart type des erreurs d'estimation sur triangles de réassurance : Property per Risk

Triangles des sinistres survenus

	délai 1	délai 2	délai 3	délai 4	délai 5	délai 6	délai 7	délai 8	délai 9
SCL		100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
MCL		92,3	106,3	131,5	137,0	118,9	149,1	186,0	77,4
NEW METHOD		98,7	98,9	100,0	100,3	99,0	99,9	101,8	99,3

Triangles des sinistres payés

	délai 1	délai 2	délai 3	délai 4	délai 5	délai 6	délai 7	délai 8	délai 9
SCL		100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
MCL		77,2	64,3	58,7	46,0	43,9	43,4	37,9	42,7
NEW METHOD		104,1	105,5	107,8	114,9	116,1	116,8	118,7	95,9

Ecart type des erreurs d'estimation sur triangles de réassurance : Property Cat

Triangles des sinistres survenus

	délai 1	délai 2	délai 3	délai 4	délai 5	délai 6	délai 7	délai 8	délai 9
SCL		100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
MCL		115,8	80,3	187,9	173,8	177,8	177,8	194,5	111,8
NEW METHOD		99,9	95,7	98,0	101,3	101,2	102,5	102,8	98,8

Triangles des sinistres payés

	délai 1	délai 2	délai 3	délai 4	délai 5	délai 6	délai 7	délai 8	délai 9
SCL		100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
MCL		49,2	40,5	26,2	23,2	23,4	24,8	23,9	55,7
NEW METHOD		94,1	92,9	92,5	82,1	81,3	80,9	80,7	101,9

Ecart type des erreurs d'estimation sur triangles d'assurance : multirisques habitations

Triangles des sinistres survenus

	délai 1	délai 2	délai 3	délai 4	délai 5	délai 6
SCL		100,0	100,0	100,0	100,0	100,0
MCL		98,5	98,2	98,4	98,4	97,4
NEW METHOD		103,7	100,2	87,0	79,7	74,9

Triangles des sinistres payés

	délai 1	délai 2	délai 3	délai 4	délai 5	délai 6
SCL		100,0	100,0	100,0	100,0	100,0
MCL		61,9	61,5	46,8	51,2	77,2
NEW METHOD		105,5	108,9	115,6	117,1	107,3

Ecart type des erreurs d'estimation sur triangles d'assurance : RC Matériels

Triangles des sinistres survenus

	délai 1	délai 2	délai 3	délai 4	délai 5	délai 6	délai 7	délai 8	délai 9
SCL		100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
MCL		100,0	100,0	100,1	100,1	100,1	100,1	100,2	100,1
NEW METHOD		111,0	122,1	123,8	159,9	166,1	152,8	154,5	194,8

Triangles des sinistres payés

	délai 1	délai 2	délai 3	délai 4	délai 5	délai 6	délai 7	délai 8	délai 9
SCL		100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
MCL		52,5	36,6	25,9	22,5	20,4	22,2	24,6	31,1
NEW METHOD		103,9	100,5	95,7	98,2	97,9	88,9	92,5	104,5

Bibliographie

[1] Beneteau G. (2004) Modèle de provisionnement sur données détaillées en assurance non-vie, – Mémoire ENSAE promotion 2004

[2] Bowmaker R et Drevillon Y.(1998) Méthode Stochastiques d'estimation des provisions IBNR en Réassurance, Mémoire du CEA promotion 1998.

[3] Clarke. H. E.(1988) Recent Developments in reserving for Losses in the London Reinsurance Market. PCAS LXXV, p. 1.

[4] DAHL P (2003) Introduction to reserving, corrected edition

[5] Hertig J. (1985) A statistical approach to IBNR-Reserves in marine reinsurance. ASTIN Bulletin 15, 2, 171-184.

[6] Hülimann W Credible Loss Ratio claims reserves : The Benktander, Neuhaus and Mack Methods revisited.

[7] Goovaerts M.J. et Hoogstad W.J. Credibility theory. Surveys of Actuarial Studies No.4.

[8] Mack,T. (1993) Distribution-free Calculation of the Standard Error of Chain Ladder Reserve Estimates ASTIN Bulletin 23, 2,213-225

[9] Mack T. (1999).The standard error of Chain Ladder reserves estimates: recursive calculation and inclusion of a tail factor. ASTIN Bulletin 29, 2, 361-366.

[10] Mack T. (2000). Credible claims reserves: The Benktander method. ASTIN Bulletin 30, 2, 333-347.

[11] Merz M. , Wüthrich M. V. (2006) A credibility approach to the Munich Chain-Ladder Method. German Actuarial Bulletin, n°2.

[12] Nelder J. A. & Wedderburn R. W. M. (1972) Generalized Linear Models. Journal of the Royal Statistical Society, Series A, 135, 370-384

[13] Partrat C. (2005) Evaluation de la provision pour sinistre à payer, cour proposé par CARITAT.

[14] Petauton P. Théorie de l'assurance dommages - manuel et exercices corrigés, Dunod

[15] Quarg G., Mack T. (2004). Munich Chain Ladder. A reserving method that reduces the gap between IBNR projections based on paid losses and IBNR projections based on incurred losses.

[16] Sherman R. E., Extrapolating, Smoothing and Interpolating Development Factors. PCAS LXXI, 1984, p. 122.

[17] Therond P.E. (2004-2005) Théorie de la crédibilité– cours de l'ISFA 2004-2005.