Université Claude Bernard – Lyon 1
INSTITUT DE SCIENCE FINANCIERE ET D'ASSURANCES

Mémoire présenté devant l'Institut de Science Financière et d'Assurances pour l'obtention du diplôme d'Actuaire de l'Université de Lyon

le _____20/12/2006_____

Par : PHAM QUANG Hung

Titre : **PROVISIONNEMENT EN ASSURANCE NON-VIE**

Confidentialité : Annexes

Membre du jury I.A.
M. PLANCHET Frédéric
Société :Winter & Associés

Membres du jury I.S.F.A.
M. AUGROS Jean-Claude
M. BIENVENÜE Alexis
M. LAURENT Jean-Paul
M. LEBOISNE Nicolas
M. LOISEL Stéphane
M. QUITTARD-PINON François
Mme REY-FOURNIER Béatrice
M. RULLIERE Didier
M. SERANT Daniel

Invité :

Entreprise :
MAAF
1, rue de Chauray
79000 NIORT

Directeur de mémoire :
ARBOGAST Etienne
Responsable des Etudes
Actuarielles IARD

Secrétariat :
Mme GARCIA Marie-José
Mme BARTHELEMY Diane
Mme BRUNET Marie-Christine
Mme MOUCHON Marie-Claude

Bibliothèque :
Mme SONNIER Michèle

50 Avenue Tony Garnier
69366 Lyon Cedex 07

Résumé

Mots clés

Provision, Assurance Non-Vie, méthodes stochastiques, méthodes Monte-Carlo, DFA, triangles cumulés - non cumulés, Bootstrap, erreur de prédiction, l'analyse financière dynamique, Modèles Linéaires Généraux, Simulation, GAM.

Ce mémoire a pour but de présenter succinctement l'état de l'art sur les méthodes stochastiques et déterministes d'évaluation de la provision de sinistres en assurance Non-Vie. Il illustre le concept de provisionnement en assurance en utilisant une méthodologie. L'objectif principal est d'analyser différentes méthodologies en assurance Non-Vie afin d'évaluer les différences entre les modèles déterministes et stochastiques.

Une vue d'ensemble de la nature de l'assurance Non-Vie est donnée dans la Partie I pour que l'on puisse mieux comprendre le concept de provisionnement. Ce qui suit est une illustration des théories du provisionnement aussi bien qu'une discussion des types de provisions qu'une compagnie d'assurance Non-Vie est amenée à réaliser.

La partie II présente un ensemble de méthodes déterministes et stochastiques, tout particulièrement le Chain-Ladder. Ces méthodes, en général robustes, doivent impérativement être utilisées avant tout appel à des méthodes stochastiques qui ont un fort potentiel, mais demandent une totale maîtrise de leur utilisation.

Les erreurs de prédiction, les résidus, la distribution des modèles stochastiques de Chain-Ladder sont présentés dans la Partie III suivi d'exemples de la documentation de messieurs England et Verral. En conclusion, durant les vingt dernières années, il y a eu un intérêt croissant pour les modèles stochastiques, malgré cela leur utilisation est toujours limitée, due à un certain nombre de raisons dont un manque de compréhension générale des méthodes ainsi que leur manque de flexibilité.

Abstract

Key words

Reserves, general insurance, stochastic methods, Monte-Carlos Methods, cumulative-non cumulative triangles, Bootstrap, Error prediction, Dynamic Financial Analysis, Generalised Linear Model, Simulation, Generalised Additive Model.

The purpose of this memory is to briefly present the state of the art on the stochastic and deterministic methods of evaluation of the provision of disasters in general insurance. It illustrates the concept of reserving along with reserving methodologies in general insurance. The main objective is to analyze reserving methodologies in general insurance in order to assess several discrepancies between deterministic and stochastic models.

An overview of the nature of general insurance is given in the Part I for a better understanding of the concept of reserving. What follows is an illustration of the theoretical background underlying reserving as well as a discussion of the types of reserves that a general insurance company can hold.

The whole methods of stochastic claims reserving in general insurance have been introduced, especially Chain Ladder in Part II. The deterministic methods must imperatively be used before the stochastic methods asking for high skills of practitioners.

Lastly, the prediction errors, residues, distribution underlying the stochastic Chain-Ladder models are presented followed by the examples in documentation of England and Verral in Part III. In conclusion, despite the fact that there has been an increasing interest in stochastic claims reserving in the last twenty years, their use is still limited due to a number of reasons including a general lack of understanding of the methods as well as lack of flexibility in the methods.

Remerciements

Je tiens à remercier les différents collègues de l'équipe IARD du Service Actuariat, Natacha ALLUSSE, Fabien DE BAROS, Anais LE TEUFF pour leur écoute, leur soutien et les connaissances qu'ils m'ont apportées.

Je remercie particulièrement mon maître de stage, M. Etienne ARBOGAST, responsable d'étude au sein du service actuariat de la compagnie d'assurance MAAF, pour son aide et ses conseils ainsi que pour avoir accepté de m'orienter dans l'organisation de mon mémoire.

Je remercie enfin M. Sébastien COSTE, responsable de division de MAAF pour m'avoir permis de mener à bien cette étude.

Sommaire

PARTIE I : INTRODUCTION ET GENERALITES DANS ASSURANCE NON -VIE

CHAPITRE 1: OBJET DU MEMOIRE 6
§1. Cadre général du stage 6
§2. Outils actuariels utilisés 6
§3. Outils informatiques utilisés 7
CHAPITRE 2: PRESENTATION DE MAAF
§1. Organigramme du Groupe MAAF Assurances 7
§2. Présentation de MAAF Assurances 8
CHAPITRE 3: L'ASSURANCE NON VIE EN FRANCE 9
§1. Réglementation 9
§2. Méthodes d'évaluation de la provision de sinistres 11
§3. Méthodes statistiques 12

PARTIE II : RAPPEL LES METHODES DETERMINISTE ET STOCHASTICS 14

CHAPITRE 1: CHAIN LADDER-DETERMINISTES 15
§1. Principes généraux. 15
§2. Les link ratios: Chain ladder standard 20
§3. La méthode London chain ajuste par taux d'inflation 23
CHAPITRE 2: CHAIN LADDER- STOCHASTIQUES 25
§1. Induction. 25
§2. Le modèle Poisson sur dispersion 26
§3. Le Modèle binomial négatif 28
§4. Approximation Loi de Normale à Modèle binomial négatif 28
§5 .Mack' Modèle 29
§6. Log Normal 30

PARTIE III : ESTIMATION D'ERREURS PREDICTIONS ET BOOTSTRAP 31

CHAPITRE 1: CALCUL DES PROVISIONS ET ERREURS PREDICTIONS 33
§1. Introduction 33
§2.Estimation d'erreurs prédictions 34
CHAPITRE 2: BOOTSTRAP 36
§1. Introduction 36
§2. Principe de Bootstrap 37
§3.Application sur modèle de Poisson surdispersion 38

CONCLUSION 42

APPENDICE 44

ANNEXES 48

A. LES CODES DE LOGICIEL SUR R 48
B. MODELE LINEAIRE GENERALISE 52
§1. Loi de probabilité de la variable de réponse 52
§2. Fonction de lien 53
§3. Fonction de Log Vraisemblance 53
§4. Estimation des paramètres 54
§5. Qualité du modèle 54
§6. Estimation de la moyenne 54
C. TYPOLOGIE DES BRANCHES (TRAVAUX DE LA COMMISSION DOMMAGES(FFA) 54
D. BIBLIOGRAPHIE 56

Partie I : Introduction et généralités sur l'assurance Non- Vie

Chapitre I : Objet du mémoire

§1. Cadre général du stage

L'objet de ce mémoire chez MAAF Assurances France est l'analyse des risques et le calcul des provisions en assurance Non vie. La Direction de MAAF a décidé de constituer un projet pour effectuer une étude complète de ce sujet. Cette étude comprend trois grands axes : la recherche des modèles stochastiques, le calcul des provisions et l'analyse des résidus et de la distribution…. Dans la suite ces trois parties sont détaillées, tout particulièrement le calcul des provisions.

Dans un premier temps nous avons présenté les modèles stochastiques en assurance non vie. La tâche n'est pas aisée, car d'une part, les modèles présentent un niveau de technicité élevé et il est difficile d'obtenir les informations voulues. D'autre part, l'utilisation de ces modèles est toujours limitée à cause du manque de compréhension général des méthodes ainsi que du manque de flexibilité dans leur utilisation. Il est nécessaire de comprendre le mode de fonctionnement des modèles pour identifier les informations utiles. Cependant, cette phase de recherche et de validation des modèles a été très utile. Elle a permis d'avoir une vue globale sur le fonctionnement de ces modèles et nous a permis d'avoir un grand nombre d'informations sur les modèles à travers des discussions avec des collègues et tout particulièrement, Monsieur Etienne Arbogast qui a encadré ce mémoire. Toutes ces informations ont été très utiles lors du provisionnement, de la collecte des données et de l'analyse de la sinistralité. Certaines hypothèses des auteurs pour réaliser le provisionnement ont été confirmées par l'étude statistique et d'autres ont été rejetées.

Le calcul actuariel des provisions en Assurance Non Vie est très important pour la compagnie. Généralement, les résultats de ce type doivent être certifiés par des actuaires indépendants. Nous avons été présents tout au long du déroulement de cette mission. Nous avons fourni les données nécessaires au calcul des provisions ainsi que validé les méthodes actuarielles et les hypothèses qu'ils ont utilisées. Dans ce mémoire nous présentons uniquement les méthodes de provisionnement car les résultats numériques sont confidentiels pour la compagnie MAAF.

§2. Outils actuariels utilisés

Les principaux outils sont les méthodes de triangulation (Chain Ladder et Bornhuetter-Ferguson) et le Modèle Linéaire Généralisé (GLM). Le GLM est une méthode d'estimation qui est bien adaptée à la construction d'un tarif. Il est largement utilisé en Assurance Automobile où il donne généralement de bons résultats. Nous avons utilisé le GLM pour identifier les erreurs de prédictions.

§3. Outils informatiques utilisés

Les principaux outils de cette analyse ont été le logiciel Excel, SAS et plus particulièrement R. R est adapté à la gestion de grandes bases de données grâce à l'utilisation du langage SQL et permet de réaliser de très belles représentations graphiques. De plus, nous avons trouvé sur internet les lignes de code d'un certain nombre de procédures statistiques, dont la procédure Packages, nous permettant d'avancer rapidement dans sur la partie « informatique ». Cette procédure permet l'application des Modèles Linéaires Généralisés. On peut spécifier toutes les distributions de la famille exponentielle et également les fonctions de lien usuelles. Le processus d'optimisation qui en réalité est assez complexe, est fait automatiquement par R. La procédure délivre beaucoup d'informations et de tests statistiques qui permettent de juger de la qualité d'un modèle et de comparer plusieurs modèles entre eux.

A plusieurs reprises nous avons également utilisé Excel ou SAS. Nous avons développé une petite application sous Visual Basic qui est utilisée pour le Bootstrap. Le logiciel permet aussi de connaître les résultats de la distribution des provisions. Nous avons aussi utilisé Visual Basic Application pour développer des tables dans Excel qui permettent identifier les erreurs de prédiction.

Chapitre II : Présentation de MAAF

§1. Organigramme du Groupe MAAF Assurances

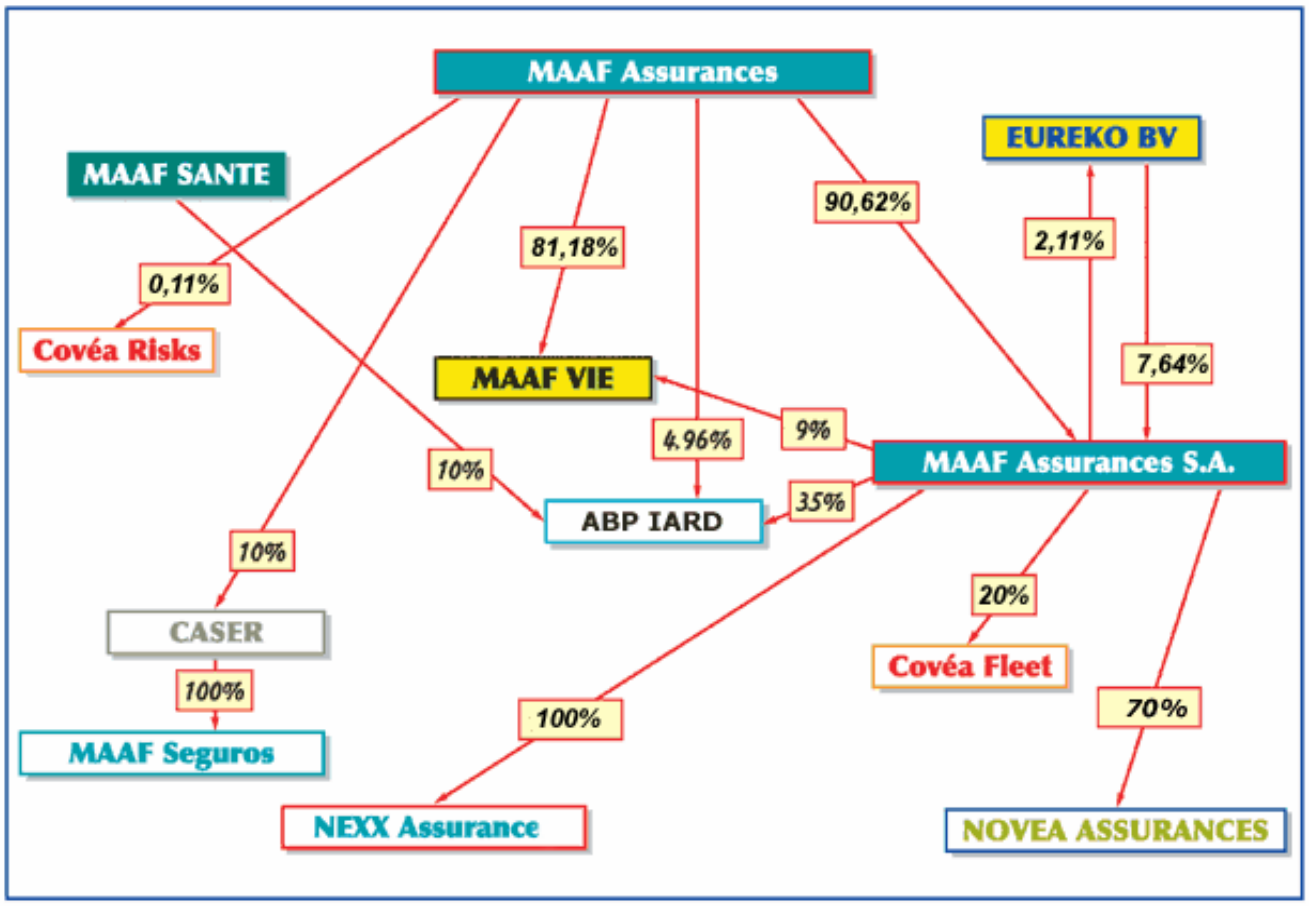
Aux premiers rangs des assureurs généralistes, le Groupe mutuel MAAF Assurances assure 3,5 millions de particuliers et de professionnels. Expert en assurances IARD (automobile, habitation, santé, risques professionnels..) comme en assistance, en crédit, en épargne et assurance vie, il est aussi devenu un assembleur de services. MAAF assurances propose des solutions globales à des problématiques de plus en plus individualisées, pour accompagner ses clients dans toutes les étapes de leur vie, au meilleur rapport qualité/prix. Son chiffre d'affaires 2005-2006 atteint près de 3.1 milliards d'euros, en hausse de plus de 6% à périmètre constant, sur un marché pourtant de plus en plus concurrentiel

Le groupe MAAF Assurances est aujourd'hui, présent dans tous les domaines de la vie des particuliers, des professionnels et des entreprises. Qu'il s'agisse d'assurer une automobile ou de financer son achat, de protéger ses proches d'une disparition ou des conséquences d'un accident, d'assurer son habitation, de compléter les remboursements des ses frais de santé, de protéger son activité professionnelle ou son entreprise, d'épargner pour préparer sa retraite, MAAF Assurances se positionne aux cotés de ses sociétaires et de leur famille.

Cette organisation comprend, entre autres :

- une société d'assurance mutuelle, à la tête du groupe, qui gère les services centraux
- une société anonyme, MAAF Assurances S.A., qui conçoit, distribue et administre les produits,
- une filiale, MAAF VIE, qui fait fructifier l'épargne sous différentes formes,

- une mutuelle de prévoyance, MAAF SANTE, qui gère les activités santé et prévoyance.
- les filiales, NEXX Assurances, NOVEA Assurances, FORCE ET SANTE, Assurance Banque Populaire IARD.



§2. Présentation de MAAF Assurances

La MAAAF : Mutuelle Assurance Automobile des Artisans de France a été crée en 1951 avec l'appui de la Chambre de métiers des Deux Sèvres et devient la MAAF (Mutuelle Assurance des Artisans de France) pour des raisons commerciales. Avec la crise économique et une concurrence de plus en plus forte, 40 ans après sa naissance, la MAAF connaît de très grosses difficultés. Les années 90 sont une période de renouveau avec l'arrivée d'un nouveau président et devient MAAF Assurances. La société MAAF Assurances compte aujourd'hui 3,5 millions de sociétaires et clients, 6 635 salariés et 573 agences.

Au niveau national, la MAAF est :
- le 8ème assureur IARD.
- le 5ème assureur automobile.
- la 1ère mutuelle des professionnels.

Chiffre d'affaires 2005 par marché (en millions d'€, hors acceptations) :

Automobile	1364
Vie privée	364
Professionnels et entreprises	285
Santé – prévoyance	371.4

Epargne, assurance vie	679

Chiffre d'affaires combiné 2005 (en millions d'€, hors acceptations) :

MAAF Assurances S.A	1847
MAAF VIE	687
MAAF SANTE	303
Assurances Banque Populaire IARD	145
NEXX Assurances	48

Chapitre III : L'assurance non vie en France

§1. Réglementation :

Au bilan de l'ensemble du marché français de l'assurance non-vie, les provisions techniques représentaient en 2002 la 2/3 du passif et la provision de sinistres 50 %[1].

D'après le Code des assurances[2], la *provision pour sinistres à payer* (PSAP en abrégé) *est la valeur estimative des dépenses en principal et en frais, tant internes qu'externes*[3]*, nécessaires au règlement de tous les sinistres* ***survenus*** *(connus ou inconnus de l'assureur) et non payés*.

Comme toutes les provisions techniques, elle doit réglementairement être *suffisante*[4] *pour le règlement intégral des engagements vis-à-vis des assurés ou bénéficiaires de contrats*. Ceci a pour conséquence une approche prudentielle dans son évaluation.

Quand elle est réglementaire[5], la PSAP, calculée exercice par exercice, est déterminée en brut de recours. Les recours à recevoir sont évalués séparément sous la forme de prévision de recours. Par différence, on obtient la PSAP nette de recours figurant au passif du bilan. Cette provision est évaluée en brut de réassurance. En contrepartie la provision pour sinistres réassurés apparaît à l'actif du bilan. Ces contraintes n'ont plus lieu d'être quand la finalité du provisionnement est d'ordre tarifaire ou quand il s'agit de déterminer un capital économique pour l'entreprise.

Réglementairement la PSAP est ventilée dans l'état C10 entre les catégories ministérielles 20 à 39 de l'assurance non-vie :

20 - Dommages corporels (contrats individuels) (y compris garanties accessoires aux contrats d'assurance vie individuels) ;
21 - Dommages corporels (contrats collectifs) (y compris garanties accessoires aux contrats d'assurance vie collectifs) ;

[1] Un tel volume d'engagements vis-à-vis des assurés et bénéficiaires de contrats nécessite le contrôle de l'Etat. Il s'opère par la Commission de contrôle des assurances devenue CCAMIP et le corps des Commissaires contrôleurs, sur pièces et sur place.

[2] Article R 331-6.

[3] Frais d'expertise, frais judiciaires,…

[4] Article R 331-1.

[5] Articles R 331-15 et R 331-1.

22 - Automobile (responsabilité civile)6 ;
23 - Automobile (dommages) ;
24 - Dommages aux biens des particuliers ;
25 - Dommages aux biens professionnels ;
26 - Dommages aux biens agricoles ;
27 - Catastrophes naturelles ;
28 - Responsabilité civile générale ;
29 - protection juridique ;
30 - Assistance ;
31 - Pertes pécuniaires diverses ;
34 - Transports ;
35 - Assurance construction (dommages) ;
36 - Assurance construction (responsabilité civile) ;
37 - Crédit ;
38 - Caution ;
39 - Acceptations en réassurance (Non-vie).

Au sein de chacune de ces catégories, deux analyses sont possibles :

1. *la première, centrée sur le résultat technique de cette catégorie, n'utilise qu'un provisionnement global ;*
2. *la seconde, centrée sur la tarification et la rentabilité des produits, nécessite un provisionnement plus segmenté.*

Dans cette dernière analyse, la Commission Dommages de la Fédération Française des Actuaires avait émis en 1998 un ensemble de recommandations quant aux segmentations à opérer à l'intérieur de chaque catégorie (cf. annexe, chapitre 1).

Un premier critère est la longueur de déroulement des sinistres, avec segmentation déroulement court (short tail) ou long (long tail). Un second critère pourrait être l'importance des tardifs ou IBNR (Incurred But Not Reported) i.e. les sinistres survenus et non connus de l'assureur.

Le tableau 1.1 donne, pour l'exercice 1993 et les principales catégories, en première colonne le montant des provisions techniques (en MF), puis, en seconde colonne, celui-ci exprimé en % des primes.

Tableau 1.1 - *Montant absolu (col.1), montant en % des primes (col.2)*

	Col.1	Col. 2 (%)
Dommages corporels	64 671	141,3
Dommages aux biens	53 648	104,4
Automobile	113 863	146,5
Transports	13 634	147,7
R C Générale	39 693	430,0
Divers	11 657	93,3
Assurance construction (Dommages ouvrages + RC décennale)	32 961	980,7
Garantie légale des catastrophes naturelles	8 318	191,7
TOTAL AFFAIRES DIRECTES France	338 445	158,5
Acceptations, Dom Tom et étranger	63 313	170,1
TOTAL	401 758	160,2

Source : Commission de contrôle des assurances

[6] Segmentée en RC corporels et RC matériels (art. R 331-26).

Une segmentation plus fine des provisions techniques en % des primes donnerait :

Dommages aux biens :	Particuliers :	97,2 %
	Professionnels :	124,9 %
	Agricoles :	50,6 %
Automobile :	RC :	250,5 %
	Dommages :	54,0 %

Compte tenu du poids des provisions de sinistres dans les provisions techniques, celles-ci, rapportées aux primes, fournissent un premier indicateur de la longueur de liquidation dans la catégorie ou sous-catégorie concernée.

Ces données confirment la longueur de la liquidation des sinistres en RC, automobile et surtout générale. Elles montrent la place totalement à part de l'assurance construction.
La longueur de liquidation des sinistres a de plus une forte influence sur l'importance des produits financiers générés par le décalage temporel entre le paiement des primes et le règlement des sinistres.

On rappelle cependant que l'escompte des provisions n'est pas réglementairement autorisé, sauf en assurance construction (et dans des conditions définies).

§2. Méthodes d'évaluation de la provision pour sinistres:

Les provisions techniques représentent la valeur des dettes probables de l'assureur au titre des opérations d'assurance. C'est un poste du bilan. Il fait l'objet des principales écritures d'inventaire des assureurs. Les provisions sont la PSAP et la PPNA

Réglementairement la méthode de base pour l'évaluation de la PSAP est la méthode dossier-dossier[7] : le gestionnaire sinistre évalue le montant restant à payer sur chaque sinistre déclaré non clos. Une estimation de la charge des tardifs (ou IBNR[8]) doit être ajoutée à cette provision dossier-dossier.

Depuis 1991, *l'entreprise peut*[9] *utiliser des méthodes statistiques*[10] *pour l'évaluation des sinistres survenus au cours des deux derniers exercices*.
La méthode de base restant l'évaluation dossier-dossier, une autre approche est aussi largement répandue, elle consiste à mobiliser un ensemble de méthodes statistiques et à comparer les estimations de la provision qu'elles produisent.

Une convergence de ces estimations conduit à une provision qualitativement "fiable".

Une divergence de celles-ci est un avertissement sur l'instabilité des données analysées et la nécessité d'adopter une démarche "prudentielle" dans l'estimation de la provision.

On peut noter que les méthodes stochastiques, exposées dans les chapitres suivants, fournissent une mesure de la "fiabilité" de l'estimation qu'elles proposent.

---

[7] *Case estimate* en anglais.

[8] IBNR pour *Incurred But Not Reported*. Ce terme est utilisé, principalement en réassurance, pour désigner l'ensemble des sinistres survenus non encore réglés.

[9] Avec l'accord de la Commission de contrôle des assurances.

[10] En assurance automobile (art. R 331-26), il est précisé qu'il s'agit de la méthode des cadences de règlement et de celle du coût moyen.

§3. Méthodes statistiques

Contrairement à l'évaluation dossier-dossier (prospective), les méthodes statistiques reposent principalement sur les **données historiques de sinistralité**.
Elles sont d'autant plus performantes que :

1. Le passé était régulier.
2. Le présent et le futur sont structurellement peu différents du passé.
3. La branche considérée est peu volatile.
4. Les données sont nombreuses.
5. Les données sont fiables.

Les risques atomique et spatial, à données rares, ne se prêteront donc pas à l'utilisation de ces méthodes.

Dans un souci de stabilité, on séparera les sinistres récurrents (ou attritionnels) des sinistres catastrophiques qui bénéficieront d'un traitement spécifique.

Comme pour la tarification des contrats, la survenance, rare, de sinistres catastrophiques perturbent gravement le processus de provisionnement. La pratique actuelle consiste à les écrêter au-delà d'un seuil donné, celui-ci étant fonction de la (sous) branche étudiée, et d'appliquer un traitement spécifique aux montants sur crête. La difficulté majeure reste le choix du seuil d'écrêtement, la sensibilité des résultats à ce choix devant être analysée.

On ne saurait suffisamment insister sur l'évidence suivante. La qualité de l'estimation de la provision repose d'abord sur la qualité et la stabilité des données. Il est donc essentiel de savoir comment ont été créées les données et de comprendre les facteurs internes et externes susceptibles de les influencer :

- **Facteurs internes**
  - évolution du portefeuille (variation de la croissance),
  - évolution des politiques de souscription, tarification et réassurance,
  - évolution de la politique de gestion des sinistres, en particulier de la cadence de règlement des sinistres,
  - modifications comptables.

- **Facteurs externes**
  - évolution des pratiques du marché
  - cycles économiques ou d'assurance
  - évolution de la sinistralité (nombre, montant)
  - évolutions réglementaires, comptables, sociales, jurisprudentielles,…
  - changements dans l'inflation des montants de sinistre etc.

Outre la sinistralité historique, l'estimation de la provision peut intégrer des informations exogènes à celle-ci, telles que :

- indicateurs d'exposition au risque (primes, nombre de contrats, …) ou de sinistralité,
- benchmarks de marché,
- plan de réassurance, pour la provision des sinistres réassurés etc.

Partie II : Rappel sur les méthodes déterministes et stochastiques pour le calcul des provisions mathématiques

En général, le processus de provisionnement des sinistres se décompose en de deux parties. La première partie est l'analyse des données disponibles, et le calage des modèles appropriés. La deuxième partie consiste à utiliser les résultats du processus de modélisation pour se projeter dans le futur et estimer les sinistres futurs.

Les principales étapes du processus de provisionnement sont les suivantes :

1. Établir le but de l'exercice, le mandat donné, le client et les autres destinataires probables des résultats.

2. Obtenir les informations de fond et les bases de données pour effectuer des projections. Ceci inclura des données quantitatives (données numériques) et qualitatives (informations descriptives concernant le développement et les objectifs des différentes branches).

3. Analyser et vérifier les données pour identifier tous les points anormaux et réconcilier les données à d'autres points de référence.

4. Obtenir des données plus détaillées pour expliquer les points anormaux sur les données. Une discussion sera probablement nécessaire avec les responsables des provisions pour sinistres, souscripteurs, régulateurs de perte, IT…

5. Effectuer des projections en utilisant plusieurs techniques de provisionnement. Ceci implique de construire un modèle et de faire des hypothèses. En outre, l'utilisation des observations passées permet d'adapter le modèle.

6. Analyser et interpréter les résultats des projections et obtenir des commentaires du destinataire des calculs.

7. Terminer les projections et documenter les calculs et les éléments importants qui ont conduit à faire certaines hypothèses.

Il y a un certain nombre de méthodes déterministe et statistiques disponibles pour estimer les provisions. Chaque méthode a besoin d'hypothèses différentes et sera réalisée avec un degré différent de prudence. La méthode de provisionnement la plus largement utilisée est le Chain-Ladder et elle est étudiée en détail dans le chapitre suivant.

Chapitre 1: Chain-Ladder déterministes

§1. Principes généraux

Ces méthodes sont parmi les plus populaires car elles sont faciles à mettre en œuvre, et facile à comprendre. L'idée est que le déroulement des paiements sont fonctions de facteurs de déroulement λ_j qui ne dépendant que de l'année de déroulement : le modèle sous-jacent est alors de la forme $E(C_{ij}) = \lambda_j C_{ij-1}$ Les paramètres intervenant présentent l'avantage d'avoir une interprétation, et son facile à estimer. L'inconvénient étant que cette estimation est relativement peu robuste.

Les données analysées sont de nature très diverses :

- montants : paiements de sinistres, charges des sinistres ou sinistres survenus, recours, sinistres réassurés,…
- primes : émises ou acquises
- nombre de sinistre : déclarés, réglés, tardifs,…
- loss ratios : montants / primes acquises, …
- montants moyens,…

Les sinistres (ou les primes) sont rapportés à des périodes : année, semestre, trimestre, …Nous utiliserons par la suite l'année.

L'année récurrente se déroule du 1/1/n au 31/12/n. Le 31/12/n est la date d'inventaire (ou de fin d'exercice).

Les sinistres sont rattachés à l'année d'origine (*origin year*) qui peut être :

- l'année de survenance (*accident year*)
- l'année de souscription (*underwriting year*)
- l'année de déclaration (*report year*)

Le rattachement le plus fréquent est celui de l'année de survenance.

Remarque : en assurance construction, les sinistres sont ventilés en croisant l'année de DROC (Date Réglementaire d'Ouverture du Chantier) et l'année de survenance (éventuellement de déclaration)

On considère une (sous) branche dont les sinistres se déroulent sur (n+1) années. On note :

i l'année d'origine (origin year), $i = 0,...,n$

j le délai de règlement (development year), $j = 0,...,n$

x_{ij} la mesure de sinistralité[11] correspondant à l'année d'origine *i* et au délai *j*.

Dans la suite on supposera, que nous traitons de paiements de sinistres. Se plaçant au 31/12/*n*, les paiements de sinistres antérieurs à cette date sont classiquement mis sous la forme d'un triangle complet de liquidation[12] de montants **non cumulés** (ou incréments, en utilisant le terme dérivé de l'anglais)[13], tableau 1.2:

Les diagonales du triangle correspondent à des règlements calendaires. L'année calendaire *n* a donné lieu au paiement total $\sum_{i=0}^{n} x_{i,n-i}$, toutes années d'origine confondues.

On peut noter que ce triangle contient $1+2+...+(n+1)=\frac{(n+1)(n+2)}{2}$ données soit 21 données pour $n=5$ (short tail) et 136 pour $n=15$ (long tail).

[11] Ou d'exposition au risque.

[12] Run-off triangle en anglais

[13] Nous utilisons les notations internationales : année d'origine en colonne, délai en ligne.

On y associe, dans le tableau 1.3, le triangle des montants cumulés (*cumulative*) au 31/12/*n*, où $C_{ij} = \sum_{h=0}^{j} x_{ih}$ est le montant **cumulé** des règlements pour l'année d'origine *i* **jusqu'au délai** de règlement *j*.

On note que $x_{io} = C_{io}$, $x_{ij} = C_{ij} - C_{i,j-1}$ $(j \geq 1)$.

Tableau 1.1 : *Triangle des montants non cumulés*

d'origine \ Délai de règlement	0	1	.....	*j*	.....	*n-i*	.....	*n*-1	*n*
0	x_{00}	x_{01}	.....	x_{0j}	.....	.....	.....	$x_{0,n-1}$	x_{0n}
1	x_{10}	x_{11}	.....	x_{1j}	.....	.....	.....	$x_{1,n-1}$	
⋮	⋮	⋮	.....	⋮		⋮			
i	.....	.....	.....	x_{ij}	.....	$x_{i,n-i}$			
⋮	⋮	⋮		⋮					
⋮	.....	.....	.....	$x_{n-j,j}$					
⋮									
n-1	$x_{n-1,0}$	$x_{n-1,1}$							
n	x_{n0}								

paiements de l'année calendaire $(i+j)$

paiements de l'année calendaire n

De manière analogue, on traitera le triangle des charges des sinistres où S_{ij} $(0 \le i, j \le n)$ désignera la charge de la $i^{\text{ème}}$ année d'origine vue après un délai j [soit à la $(i+j)^{\text{ème}}$ année].

Tableau 1.2 – *Triangle des montants cumulés*

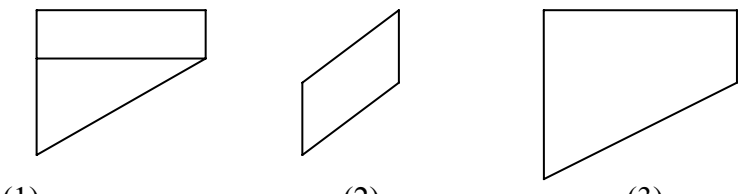
d'origine	0	……	j	j+1	……	n-i	……	n
0	C_{00}	……	C_{0j}	$C_{0,j+1}$	……	$C_{0,n-i}$	……	C_{0n}
1	C_{10}	……	C_{1j}	$C_{1,j+1}$	……	⋮	……	
⋮	⋮		⋮	⋮		⋮		
i	C_{i0}		C_{ij}	$C_{i,j+1}$	……	$C_{i,n-i}$		
⋮	⋮		⋮					
n-j	⋮	……	$C_{n-j,j}$					
n-1	$C_{n-1,0}$							
N	C_{n0}							

A partir des données de paiements du triangle supérieur (et éventuellement d'informations exogènes), on cherche à obtenir une estimation au 31/12/n de :

- la charge sinistres $S_i = C_{in}$ de chaque année d'origine i $(i = 0,...,n)$. L'année 0 étant complètement déroulé, $S_0 = C_{0n}$ est connue,
- la provision $(R_i = C_{in} - C_{i,n-i})$ à constituer pour chaque année d'origine i $(i = 0,...,n)$ avec $R_0 = 0$,
- la provision globale $R = \sum_{i=1}^{n} R_i$ (toutes années d'origine confondues).

Ces estimations sont primordiales dans la détermination du résultat technique de la (sous) branche considérée. Accessoirement, pour des finalités comptables, on peut estimer les paiements futurs au-delà du 31/12/n en fonction de l'année d'origine i et du délai j sous la forme cumulée C_{ij} ou non cumulée x_{ij} $(i + j > n)$.

Remarque : La majeure partie des méthodes statistiques de provisionnement s'applique à d'autres figures de liquidation (données) que des triangles complets :



(1) (2) (3)

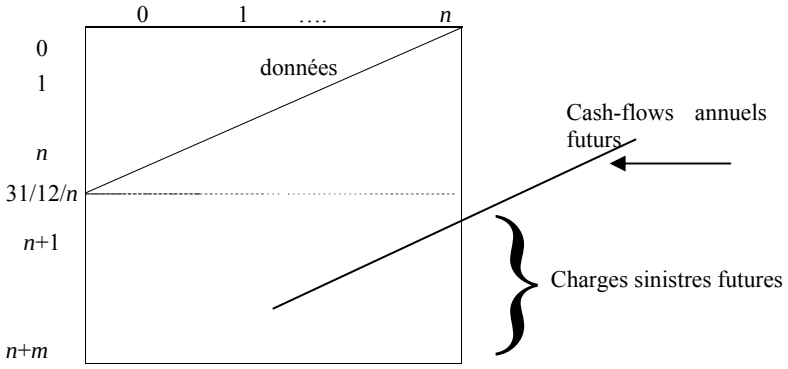
Le trapèze (3) est la figure la plus fréquente, étant celle disponible en externe par les états C.

On peut noter que l'évaluation selon les normes IFRS d'un groupe de contrats, supposant le portefeuille fermé (pas de renouvellements ni d'affaires nouvelles), fait intervenir la somme des cash-flows futurs (actualisés) de paiements de sinistres.

En négligeant l'actualisation, les cash-flows futurs au cours de l'année $(n+k)$, au titre des années d'origine 0 à *n*, donnent :

$$CF_{n+k} = \sum_{i+j=n+k} x_{ij}$$

Dans le même esprit, les méthodes récentes d'évaluation des stratégies et de gestion des risques (modèles internes - DFA, Gestion Actif Passif,…) privilégient l'approche prospective. Sont analysées les projections, au delà du 31/12/*n*, de charges sinistres, de cadences de règlement et de cash-flows futurs. Cette approche peut-être représentée par la figure suivante :



Facteurs de développement, cadences de règlement

Dans la plupart des méthodes de cadences, les calculs sont conduits à l'aide des facteurs de développement mais les résultats sont interprétés en cadences cumulées. La correspondance facteurs cadences est explicitée ci-dessous.

On considère l'année d'origine *i* dont on décrit le développement complet en paiements cumulés au 31/12/*n* :

Tableau 1.3 : Le développement complet en paiements

	0	……	*n-i*	……	*j*	*j* +1	……	*n*
I	C_{i0}	……	$C_{i,n-i}$	……	C_{ij}	$C_{i,j+1}$	……	C_{in}

Facteur f_j

Dernier montant connu

Cadence pc_j

$j^{\text{ième}}$ facteur de développement[14] :

$$f_j = \frac{C_{i,j+1}}{C_{ij}} \quad (j = 0,...,n-1)$$

$j^{\text{ième}}$ cadence cumulée de règlement[15] :

$$pc_j = \frac{C_{ij}}{C_{in}} \quad (j = 0,...,n-1)$$

Il existe une correspondance biunivoque entre les cadences $(pc_0, pc_1,..., pc_{n-1})$ et les facteurs de développement $(f_0, f_1,..., f_{n-1})$ associés à une année d'origine par

$$f_j = \frac{pc_{j+1}}{pc_j} \text{ (Avec } pc_n = 1)$$

$$pc_j = \frac{1}{f_j \times f_{j+1} \times ... \times f_{n-1}}$$

Dans l'hypothèse où l'on a choisi un vecteur de cadences $(pc_0,..., pc_{n-1})$ ou un vecteur de facteurs de développement$(f_0,..., f_{n-1})$, ces deux vecteurs étant en correspondance, le calcul de la $i^{\text{ème}}$ provision $R_i = C_{in} - C_{i,n-i}$ s'écrit

$$R_i = \frac{1 - pc_{n-i}}{pc_{n-i}} C_{i,n-i} = (f_{n-i} \times ... \times f_{n-1} - 1) C_{i,n-i}$$

Remarque : Il est important d'analyser l'impact d'une modification d'un facteur de développement, par exemple le premier. Le passage de f_0 à $f_0' = \alpha f_0$, $\alpha \succ 0$ conduit aux cadences cumulées

$$pc_0' = \frac{1}{f_0' \times f_1 \times ... \times f_{n-1}} = \frac{pc_0}{\alpha}$$

Dans l'hypothèse où les facteurs f_j , $j \geq 1$ restent inchangés, on a :

$$pc_j' = pc_j$$

Il en résulte de plus une charge sinistres et une provision pour l'année *n*

$$S_n' = C_{n0} f_0' f_1 ... f_{n-1} = \alpha S_n$$

[14] Link ratio ou development factor ou age to age factor en anglais

[15] Lag factor en anglais

§2. Chain-Ladder standard :

Considérons les délais génériques j et j+1 d'un triangle de paiements cumulés (ou de charges sinistres).

Tableau 2.1 : Un triangle de paiements cumulés :

Année D'origine	0	……	J	j+1	……	……	n
0	……	……	$C_{0\,j}$	$C_{0,\,j+1}$	……	……	
⋮			⋮	⋮			
I	……	……	C_{ij}	$C_{i,\,j+1}$			
⋮			⋮	⋮			
n-j-1	……	……	$C_{n-j-1,\,j}$	$C_{n-j-1,\,j+1}$			
n-j	……	……	$C_{n-j,\,j}$				
⋮							
N							

La méthode Chain-Ladder s'appuie sur l'hypothèse (forte) suivante :

Hypothèse : *Pour* $j = 0,..,n-1$, *les ratios des facteurs adjacents* $\dfrac{C_{i,\,j+1}}{C_{ij}}$ *sont indépendants de l'année d'origine i.*

D'où, pour $j = 0,...,n-1$:

$$\frac{C_{0,\,j+1}}{C_{0\,j}} = \frac{C_{1,\,j+1}}{C_{1\,j}} = ... = \frac{C_{i,\,j+1}}{C_{ij}} = ... = \frac{C_{n-j-1,\,j+1}}{C_{n-j-1,\,j}} \quad (1)$$

La valeur commune de ces rapports est aussi

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

Dans la pratique, les égalités (1) n'étant, au mieux, qu'approximativement vérifiées, il est naturel de choisir comme facteur commun :

$$f_j = \frac{\sum_{i=0}^{n-j-1} C_{i,\,j+1}}{\sum_{i=0}^{n-j-1} C_{ij}} \qquad (0 \le j \le n-1) \ .$$

On en déduit les évaluations des charges ultimes

$$S_i = C_{in} = C_{i,n-i} f_{n-i} \cdots f_{n-1} = C_{i,n-i} \prod_{h=n-i}^{n-1} f_h$$

conduisant aux provisions $R_i = S_i - C_{i,n-i}$ $(i = 1,...,n)$ par exercice puis à la provision globale $R = \sum_{i=1}^{n} R_i$.

Si nécessaire, les facteurs de développement fournissent un rectangle complet de liquidation par

$$C_{ij} = C_{i,n-i} f_{n-i} \cdots f_{j-1} = C_{i,n-i} \prod_{h=n-i}^{j-1} f_h \ .$$

Exemple 1 : Du triangle des paiements cumulés bruts de recours donné, on déduit, sous l'hypothèse ci-dessus, les facteurs et cadences du tableau 1.2 et 1.3.

Tableau 2.2 : Les données d'England et Verral

	Dvlpt									
Surv	J=1	2	3	4	5	6	7	8	9	10
i=1	5012	3257	2638	898	1734	2642	1828	599	54	172
2	106	4179	1111	5270	3116	1817	-103	673	535	
3	3410	5582	4881	2268	2594	3479	649	603		
4	5655	5900	4211	5500	2159	2658	984			
5	1092	8473	6271	6333	3786	225				
6	1513	4932	5257	1233	2917					
7	557	3463	6926	1368						
8	1351	5596	6165							
9	3133	2262								
10	2063									

Tableau 2.3-*Facteurs de développement et cadences, exemple 1*

Facteurs de développements Chain-Ladder								
2.999	1.624	1.271	1.172	1.113	1.042	1.033	1.017	1.009

Tableau 2.4-*Les valeurs prévus par la méthode Chain Ladder standard*

	Dvlpt									
Surv	j=1	2	3	4	5	6	7	8	9	10
i=1	5012	8269	10907	11805	13539	16181	18009	18608	18662	18834
2	106	4285	5396	10666	13782	15599	15496	16169	16704	16858
3	3410	8992	13873	16141	18735	22214	22863	23466	23863	24083
4	5655	11555	15766	21266	23425	26083	27067	27967	28441	28703
5	1092	9565	15836	22169	25955	26180	27278	28185	28663	28927
6	1513	6445	11702	12935	15852	17649	18389	19001	19323	19501
7	557	4020	10946	12314	14428	16064	16738	17294	17587	17749
8	1351	6947	13112	16664	19525	21738	22650	23403	23800	24019
9	3133	5395	8759	11132	13043	14521	15130	15634	15898	16045
10	2063	6188	10046	12767	14959	16655	17353	17931	18234	18402

Tableau 2.4-*Rectangle de liquidation des paiements cumulés bruts de recours et réserves, exemple 1*

Dvlpt										Ultimes	Réserves
J=1	2	3	4	5	6	7	8	9	10		
5012	8269	10907	11805	13539	16181	18009	18608	18662	18834	18834	
106	4285	5396	10666	13782	15599	15496	16169	16704		16858	154
3410	8992	13873	16141	18735	22214	22863	23466			24083	617
5655	11555	15766	21266	23425	26083	27067				28703	1636
1092	9565	15836	22169	25955	26180					28927	2747
1513	6445	11702	12935	15852						19501	3649
557	4020	10946	12314							17749	5435
1351	6947	13112								24019	10907
3133	5395									16045	10650
2063										18402	16339
										TOTAL :	**52135**

La provision globale estimée est de 52135. On note que la provision afférente aux trois dernières années de survenance représente 73% de la provision globale donc cette branche est « long tail ». Au contraire, si la provision de la dernière année uniquement représentait la majorité de la provision (par exemple 85%), ce serait caractéristique d'une branche « short tail ».

Il y a beaucoup de variations sur la méthode « Chain-Ladder standard », par exemple : la méthode « Chain-Ladder pondérée », la méthode « London-Chain », la méthode « London-Pivot », les méthodes basées sur des ratios, la méthode « Cape-Cod », la méthode de « Bornhuetter-Fergusonmais »…mais toutes ont le même objectif : extraire du triangle de développement de les années suivantes

§3. Chain-Ladder standard- traitement de l'inflation :

La méthode de « Chain-Ladder standard » ne tient compte explicitement d'aucun effet d'année calendaire tel que l'inflation. Les facteurs de développement calculés sont basés sur des paiements. L'inflation passée est implicite dans les facteurs dérivés, mais les hypothèses sur l'inflation future ne sont pas clairement explicites. Si les anticipations sur l'inflation future sont très différentes des tendances passées, le taux implicite d'inflation pourra être inadéquate pour des projections de règlements futurs.

Il est simple de faire des ajustements pour l'inflation si les taux d'inflation des règlements connus sont disponibles. Cependant, pour analyser le modèle, il est essentiel de considérer des données non cumulées, plutôt que cumulées, pour que l'influence de chaque année calendaire soit isolée. Le modèle « Chain-Ladder standard » peut être modifié en utilisant la formule suivante :

$$X_{ij} = S_i r_j \lambda_{i+j-1} + E_{ij}$$

λ_{i+j-1} : Taux d'inflation pour l'année calendaire i+j-1

E_{ij} : Le terme d'erreur

r_j : La proportion de S_i observé durant la période de développement j

S_i : Charges de sinistres

Afin de compléter la projection, il y a deux étapes supplémentaires. Premièrement, les taux d'inflation passés sont demandés pour ajuster les règlements sur des valeurs monétaires courantes.

Puis, des hypothèses sont exigées pour les futurs taux d'inflation, qui sont appliqués aux futures valeurs non cumulés projetées.

Les paiements non cumulés ajustés de l'inflation sont présentés dans l'exemple 1 (basé sur les données non cumulés du tableau 2.2), supposant des taux d'inflation passés donnés par le tableau 3.1. Par exemple:

$$X_{2,9} = 535*1.11=541$$

On note que les valeurs cumulées des ratios développements sont également montrées dans le tableau 1.2. Les Ratios développements sont calculés comme pour la méthode « Chain-Ladder standard ». En plus, ces facteurs sont inférieurs à ceux calculés en utilisant la méthode « Chain-Ladder standard ». C'est parce que les facteurs de la méthode « Chain-Ladder standard » font une hypothèse implicite sur la valeur de l'inflation future, basée sur les tendances passées.

Table 3.1 : Taux d'inflation chaque année

Année	Taux d'inflation
1	1.025
2	1.015
3	1.034
4	1.047
5	1.024
6	1.067
7	1.072
8	1.021
9	1.011

Table 3.2 : Les donnes non cumulatives ajustées pour inflation

6825	4327	3453	1137	2096	3119	2023	618	55	172
144	5552	1454	6671	3767	2145	-114	695	541	
4644	7416	6389	2871	3136	4108	718	622		
7701	7838	5512	6962	2610	3138	1089			
1487	11257	8208	8017	4577	266				
2060	6552	6881	1561	3527					
758	4601	9065	1732						
1840	7434	8069							
4266	3005								
2809									

Table 3.3 : Les provisions avec les taux d'inflation :

Dvlpt J=1	2	3	4	5	6	7	8	9	10	Ultimes	Réserves
6825	11152	14605	15742	17838	20957	22980	23599	23653	23825	23825	
144.3	5696	7150	13821	17589	19734	19620	20315	20856		21007	152
4644	12059	18448	21319	24455	28563	29281	29903			30529	626
7701	15539	21051	28013	30623	33761	34850				36538	1687
1487	12744	20952	28968	33546	33811					36727	2916
2060	8613	15493	17054	20581						24658	4077

758.5	5359	14425	16156
1840	9274	17343	
4266	7271		
2809			

22412	6256
30271	12927
20428	13156
23286	20477
TOTAL :	**62274**

Chapitre II : Les méthodes stochastiques

§1. Introduction

La technique de « Chain-Ladder » a été conçue comme une méthode déterministe pour calculer les paiements futurs. Un modèle déterministe de provisionnement est principalement considéré un modèle qui ne fait que des hypothèses sur les valeurs prédites des paiements futurs (Claims Reserving Manual, Vol.2, 1997).

Comme déjà énoncé, le processus de provision permettant d'arriver à une évaluation des paiements futurs est décrit comme un élément à part entière de la construction du modèle, en l'adaptant à quelque ensemble d'observations passées, et l'utilisant pour inférer des résultats sur les paiements futurs.

Les modèles stochastiques font des hypothèses au sujet des valeurs prévues des paiements futurs et modélisent aussi la variation de ceux-ci. Les modèles stochastiques permettent non seulement de tester statistiquement les hypothèses mais aussi d'obtenir une estimation de la valeur de la provision des paiements futurs avec un intervalle de confidence spécifique.

Les modèles stochastiques ont plusieurs avantages et des inconvénients par rapport aux modèles déterministes, comme illustrés ci-dessous.

Les modèles déterministes peuvent souvent être appliqués sans une connaissance claire des hypothèses. L'avantage fondamental d'un modèle stochastique est que les hypothèses sont totalement explicites. De plus, il permet de réaliser des tests sur ces hypothèses par une variété de techniques pour s'assurer que ces hypothèses sont pertinentes.

Un modèle déterministe fait simplement une estimation des paiements futurs espérés pour une période donnée. Les paiements effectifs seront différents de ceux estimés, et les modèles déterministes ne donnent pas d'informations sur la significativité de cette différence. Par ailleurs, les modèles stochastiques permettent de construire un intervalle de confiance pour les paiements espérés et il peut être utilisé comme indicateur pour confirmer si les hypothèses peuvent être considérées comme bonnes.

Un inconvénient d'une méthode provision stochastique est qu'il modélise une série très complexe d'événements avec seulement quelques paramètres, et par conséquent cette méthode peut être critiquées du fait que ses hypothèses sont beaucoup trop simples et donc, irréalistes. Parce que les modèles stochastiques sont assez clairs et rigides, il n'y a que peu de place pour incorporer le jugement, ou d'autres facteurs étrangers dans le modèle.

De plus, les modèles stochastiques peuvent être tout à fait complexes à exécuter, et peuvent demander une capacité statistique et de calcul plus grand qu'un modèle déterministe simple. Cela peut vouloir dire que les résultats sont plus difficiles à communiquer quand on les compare à quelques modèles simples déterministes.

En général, l'avantage fondamental des modèles stochastiques de provisionnement est la disponibilité de mesure de précision des évaluations de réserve. Dans la Partie III de cette étude, nous porterons notre attention sur l'étude de la moyenne des carrés des écarts à la prédiction (MCEP).

Un bon modèle stochastique est celui qui a assez de paramètres pour décrire les caractéristiques des données, mais pas trop ce qui limitera son pouvoir descriptif. En outre, plus le nombre de paramètres du modèle est augmenté, plus son pouvoir de prédiction est diminué. Le modèle commence à adhérer de plus en plus aux données initiales. Tout changement dans les données peut engendrer de grands changements dans les paramètres du modèle et rendre ainsi toute production du modèle instable. De plus, le modèle devrait nous permettre de prendre conscience de l'influence des données, ainsi que de l'influence de l'aléatoire. Un bon modèle stochastique devrait être capable de tester les hypothèses sous-jacentes. Cela permettra au chercheur de gagner une plus grande compréhension des caractéristiques des données et d'avoir le meilleur contrôle sur les valeurs projetées.

Les modèles stochastiques sont considérés comme pouvant fournir des solutions aux problèmes quand les modèles déterministes échouent. Cependant, ce n'est pas le cas. L'utilité des modèles stochastiques sont qu'ils peuvent, dans beaucoup de circonstances, fournir plus d'informations qui peuvent être utile dans le processus de réserve et dans la gestion totale de la compagnie que les modèles déterministes.

Intuitivement, la technique de Chain-Ladder paraît naturelle et a toujours été considérée comme un modèle non-stochastique. Cependant, il a été prouvé que cette technique peut être basée sur un modèle stochastique. Tout d'abord, nous calculons les résultats des modèles stochastiques puis nous les comparons à la technique de Chain-Ladder, ce qui permet à celle-ci d'être regardé dans un plus large cadre de modèle. En outre, la technique de Chain-Ladder ne peut pas être regardée séparément du principe de la modélisation statistique utilisée pour les provisions pour sinistres (Renshaw et Verrall, 1994).

La section suivante illustre les modèles stochastiques qui reproduisent les estimations de réserves obtenues par la technique de Chain-Ladder.

§2. Modèle Poisson sur-dispersion :

Sous la distribution de Poisson sur-dispersion, la variance est proportionnelle à la moyenne. Dans la provision pour sinistre, le modèle de Poisson sur-dispersion suppose que les sinistres non cumulés, X_{ij}, sont distribués suivant une loi de Poisson sur-dispersion avec les variables aléatoires indépendantes, avec la moyenne et la variance suivante :

$$E(X_{ij}) = m_{ij} = x_i y_j \text{ et } Var(X_{ij}) = \phi x_i y_j$$

Où,

$$\sum_{k=1}^{n} y_k = 1$$

x_i : les sinistres ultimes espérés et y_j : la proportion de sinistres ultimes de chaque année du développement j; ϕ un paramètre inconnu et estimé à partir des données.

y_j apparaît dans la variance, et est imposé automatiquement comme positif. Donc, la somme des sinistres non cumulés dans la colonne j doit aussi être positive ce qui est une limite du modèle. On notera que quelques sinistres non cumulés négatifs sont permis tant que la somme de la colonne n'est pas négative.

Dans cette formulation, la moyenne a une structure multiplicative, en étant le produit d'un effet de la ligne i et d'un effet de la colonne j. Une fonction de lien logarithmique est utilisée dans le modèles linéaires généralisés, afin que :

$$Log(m_{ij}) = c + \alpha_i + \beta_j$$

On note que $Log(m_{ij})$ est connu comme le prédicteur linéaire, et est noté par η_{ij} . Cette structure de prédicteur est encore du type Chain-Ladder, car il y a un paramètre pour chaque ligne i et un paramètre pour chaque colonne j.
L'avantage de cette forme de modèle est que comme un modèle linéaire généralisé, il est facile à estimer. Les logiciels standards peuvent être utilisés pour avoir des estimations qui en général sont de bonne qualité. Pourtant, l'inconvénient est que les valeurs des paramètres seront plus difficiles à interpréter, donc il est nécessaire de les convertir sous une forme plus familière. Les contraintes sont appliquées à l'ensemble des paramètres, par exemple,

$$\alpha_1 = \beta_1 = 0$$

Malgré qu'on s'approche d'une loi de Poisson au modèle dans la section, ça n'implicite pas qu'il est capable d'appliquer aux données positives. Les contraintes peuvent être résolus en utilisant l'approche de quasi-vraisemblance qui permets appliquer aux données positives. En termes de modélisation, l'hypothèse la plus importante est que la variance doit être proportionnelle à la moyenne et que les données ne sont pas nécessaires en forme de positifs. En utilisant l'approche de quasi-vraisemblance avec le support de plusieurs logiciels, nous pouvons calculer les paramètres des modèles linéaires généralisés.

Les tableaux suivants sont des illustrations du modèle de Poisson sur-dispersion, obtenus en utilisant le logiciel Excel. Les X_{ij}, sinistres non cumulés, ont été modélisés comme étant des variables aléatoires de Poisson avec une fonction de lien logarithmique et un prédicteur linéaire $\eta_{ij} = c + \alpha_i + \beta_j$. La sur-dispersion a été considérée en estimant le paramètre de l'échelle inconnu ϕ comme faisant partie de la procédure. Le logiciel standard de statistique qui fait l'adéquation des modèles linéaires généralisés peut être utilisé pour obtenir le paramètre estimé avec la fonction du maximum de vraisemblance. Aussi, on note que les valeurs adéquates seront toujours positives et que les équivalents links ratios seront toujours par conséquent plus grands que 1. Dans le premier pas de l'exemple ci-dessous, les valeurs initiales ont été simulées, et des valeurs des paramètres ont été choisies comme point de départ pour l'optimisation. Ensuite, le prédicteur linéaire et les moyennes sont calculés. Puis, la log-vraisemblance pour chaque point dans le triangle a été calculée, la Log-Vraisemblance totale est la somme de tous les éléments du triangle. La maximisation des estimations Log-vraisemblance des paramètres a été faite en utilisant le Solveur dans Excel. Les estimations des paiements futurs ont été obtenues à partir des estimations des paramètres $\alpha_i et \beta_j$ en les insérant dans l'équation $Log(m_{ij}) = c + \alpha_i + \beta_j$ dont on prend ensuite la forme exponentielle : $\hat{X}_{ij} = m_{ij} = \exp(\eta_{ij})$

La provision globale d'une année d'origine est égale à l'addition des termes d'une diagonale. On note que la table de poids est prise en compte dans le calcul de la Log-vraisemblance. En général, on observe que l'estimation de la réserve obtenue en utilisant ce modèle est la même que celle obtenue en utilisant le modèle de Chain Ladder standard.

Tableau 2.1 : les données décumulées dans Exemple England et Verral.

Données décumulées

	Dvlpt									
Surv	J=1	2	3	4	5	6	7	8	9	10
I=1	5012	3257	2638	898	1734	2642	1828	599	54	172
2	106	4179	1111	5270	3116	1817	-103	673	535	
3	3410	5582	4881	2268	2594	3479	649	603		
4	5655	5900	4211	5500	2159	2658	984			
5	1092	8473	6271	6333	3786	225				
6	1513	4932	5257	1233	2917					
7	557	3463	6926	1368						
8	1351	5596	6165							
9	3133	2262								
10	2063									

Tableau 2.2 : Prédicteur linéaire $\eta_{ij} = c + \alpha_i + \beta_j$.

Prédicteur linéaire

	Dvlpt									
Surv	1	2	3	4	5	6	7	8	9	10
1	7.66	8.35	8.28	7.93	7.72	7.46	6.57	6.38	5.74	5.15
2	7.54	8.24	8.17	7.82	7.60	7.35	6.46	6.27	5.63	5.04
3	7.90	8.59	8.53	8.18	7.96	7.71	6.82	6.63	5.98	5.39
4	8.08	8.77	8.70	8.35	8.14	7.88	6.99	6.80	6.16	5.57
5	8.08	8.78	8.71	8.36	8.14	7.89	7.00	6.81	6.17	5.58
6	7.69	8.38	8.32	7.97	7.75	7.49	6.61	6.42	5.77	5.18
7	7.60	8.29	8.22	7.87	7.66	7.40	6.51	6.32	5.68	5.09
8	7.90	8.59	8.52	8.18	7.96	7.70	6.82	6.62	5.98	5.39
9	7.49	8.19	8.12	7.77	7.56	7.30	6.41	6.22	5.58	4.99
10	7.63	8.32	8.26	7.91	7.69	7.44	6.55	6.36	5.72	5.12

Tableau 2.3 : L'espérance du modèle Poisson sur-dispersion : $\hat{X}_{ij} = m_{ij} = \exp(\eta_{ij})$

Espérance

	Dvlpt									
Surv	J=1	2	3	4	5	6	7	8	9	10
I=1	2111.2	4221.3	3948.5	2785.0	2243.1	1735.8	714.8	590.8	310.8	172.0
2	1889.8	3778.5	3534.3	2492.9	2007.8	1553.8	639.8	528.8	278.2	154.0
3	2699.8	5398.0	5049.2	3561.4	2868.4	2219.7	914.0	755.5	397.4	220.0
4	3217.7	6433.5	6017.7	4244.5	3418.7	2645.5	1089.4	900.4	473.6	262.2
5	3242.7	6483.6	6064.6	4277.6	3445.3	2666.1	1097.9	907.4	477.3	264.2
6	2186.1	4370.9	4088.5	2883.8	2322.7	1797.4	740.1	611.7	321.8	178.1
7	1989.7	3978.3	3721.2	2624.7	2114.0	1635.9	673.6	556.8	292.9	162.1
8	2692.6	5383.6	5035.7	3551.9	2860.8	2213.8	911.6	753.4	396.3	219.4
9	1798.7	3596.4	3364.0	2372.7	1911.1	1478.9	609.0	503.3	264.8	146.6
10	2063.0	4124.9	3858.3	2721.4	2191.9	1696.2	698.5	577.3	303.7	168.1

En maximisant la Log vraisemblance, on obtient les paramètres α et β

Tableau 2.4 : Log-vraisemblance

Log-vraisemblance

Sur v \ Dvlpt	j=1	2	3	4	5	6	7	8	9	10
I=1	36255.8	22967.8	17897.0	4337.9	11135.8	17971.5	11298.8	3231.7	-0.9	713.4
2	-1090.1	30644.3	5542.8	38724.8	21688.8	11798.3	-1305.3	3691.3	2732.9	
3	24242.4	42572.5	36571.0	14986.1	17783.8	24586.5	3510.8	3240.8		
4	42454.4	45305.2	30628.3	41699.1	14149.1	18301.2	5792.1			
5	5585.2	67884.2	48557.2	48673.6	27390.8	-891.2				
6	9448.7	36972.7	39628.4	6939.4	20285.5					
7	2241.1	24725.2	53223.0	8145.2						
8	7978.0	42692.3	47516.7							
9	21682.6	14924.2								
10	13681.6									

Tableau 2.5 : Les paramètres et les réserves du modèle de Poisson sur-dispersion

	Paramètre du modèle	Ultimes	Réserves
Mu	**7.6550**	18833	0
alpha1		16858	154
alpha2	**-0.1108**	24083	617
alpha3	**0.2459**	28703	1636
alpha4	**0.4214**	28927	2747
alpha5	**0.4291**	19501	3649
alpha6	**0.0348**	17749	5435
alpha7	**-0.0593**	24019	10907
alpha8	**0.2432**	16045	10650
alpha9	**-0.1602**	18403	16340
alpha10	**-0.0231**	TOTAL :	52137
beta1			
beta2	**0.6929**		
beta3	**0.6261**		
beta4	**0.2770**		
beta5	**0.0606**		
beta6	**-0.1958**		
beta7	**-1.0831**		
beta8	**-1.2736**		
beta9	**-1.9160**		
beta10	**-2.5074**		

§3. Modèle binomial négatif :

Le modèle binomial négatif est un dérivé du modèle de Poisson sur-dispersion en utilisant un paramètre différent. Ce modèle paraît être plus intuitif pour la technique de Chain-Ladder, car les manières d'estimer les paramètres se ressemblent. Le modèle binomial peut être exprimé comme un modèle pour sinistres non cumulés ou cumulés. X_{ij} les sinistres non cumulés ont une distribution binomiale négative sur-dispersion où

$$Moyenne = (\lambda_j - 1)C_{i,j-1} \text{ et } Variance = \phi\lambda_j(\lambda_j - 1)C_{i,j-1}$$

λ_j ressemble au facteur de développement de Chain Ladder standard ; ϕ est un paramètre inconnu de la dispersion de la variance qui caractérise la 'sur-dispersion' de la distribution.

On va noter que, $C_{ij} = C_{i,j-1} + X_{ij}$, ce qui suppose que C_{ij} est connu. Cela implique que ce modèle peut aussi être écrit en terme de sinistres cumulés où C_{ij} a une distribution binomiale négative sur-dispersion, avec

$$Moyenne = \lambda_j C_{i,j-1} \text{ et } Variance = \phi\lambda_j(\lambda_j - 1)C_{i,j-1}$$

Il importe de noter que le modèle binomial négatif a été transformé en premier par Verrall (2000), en intégrant les paramètres de la ligne du modèle Poisson. Ce modèle étant dérivé du modèle Poisson, il est essentiel que leurs distributions estimées soient les mêmes, donnant des valeurs prédites identiques.

Dans l'appendice 1 nous avons utilisé le même exemple avec le modèle binomial négatif dont les paramètres ont été obtenus en utilisant le logiciel Excel. Les Link-ratios ont été créés premièrement avec les valeurs initiales. Alors que la Log-vraisemblance pour chaque point dans le triangle a été calculée et sommée afin d'obtenir la Log-vraisemblance totale, comme pour le modèle de Poisson sur-dispersion. Encore une fois, le maximum de la Log-vraisemblance a été optimisé en utilisant l'outil Solveur d'Excel. Comme il peut être observé, les facteurs du développement obtenus par le modèle sont exactement les mêmes que les facteurs de développement de Chain Ladder. On peut en déduire que le modèle binomial négatif produit les mêmes évaluations de réserve (les différences observées s'expliquent par les erreurs d'arrondis) que le modèle de Chain Ladder standard.

§4. Approximation Normale à un modèle binomial négatif :

Le modèle Normale pris comme une approximation du modèle binomial négatif a l'avantage de produire des estimations pour une large gamme de données, et d'être moins affecté par la présence de valeurs négatives.

En utilisant une approximation Normale pour la distribution de sinistres non cumulés, X_{ij} suit approximativement une loi Normale, avec

$$Moyenne = (\lambda_j - 1)C_{i,j-1} \text{ et } Variance = \phi_j C_{i,j-1}$$

où C_{ij} suit approximativement une loi Normale avec

$$Moyenne = \lambda_j C_{i,j-1} \text{ et } Variance = \phi_j C_{i,j-1}$$

Quand les C_{ij} sont connus, il est possible d'estimer la variance des sinistres cumulés par une approche récursive.

Les paramètres inconnus de la variance peuvent être estimés en utilisant une procédure d'estimation itérative "Joint-Modelling", ce qui implique d'estimer l'ensemble des paramètres. Une description de la procédure "Joint Modelling" est donnée par Renshaw (1994).

Le modèle binomial négatif ainsi que le modèle de l'approximation Normale sont des modèles récursifs où les paramètres sont obtenus par maximisation de la vraisemblance.

§5. Le modèle de MACK :

Le modèle de Mack (1993) est un des premiers modèles stochastiques qui a reproduit les estimations de Chain Ladder. Le modèle a une approche récursive semblable à celle du modèle

binomial négatif et celle du modèle Normal. Le modèle de Mack a la même structure que l'approximation normale du modèle binomial négatif.

Sous le modèle de Mack, des hypothèses limitées sont faites comme par exemple sur la distribution des données sous-jacentes. Seuls les premiers deux moments sont spécifiés. La moyenne et la variance de C_{ij} sont

$Moyenne = \lambda_j C_{i,j-1}$ et $Variance = \sigma_j^2 C_{i,j-1}$ respectivement.

Mack a produit des estimateurs des paramètres inconnus λ_j et σ_j^2. Pour les calculer, il n'est pas nécessaire de faire des hypothèses. Mack a aussi fourni une formule pour les erreurs de prédiction de paiements prévus et des évaluations de réserve (lesquelles sont illustrées dans la section suivante). Mack considère un modèle distribution-libre (free-distribution), parce que la distribution exacte des données sous-jacentes n'est pas spécifiée. Cela simplifie le modèle, mais il limite l'analyse de la distribution des réserves aux deux premiers moments seulement. Si les résultats sont utilisés pour l'analyse financière dynamique, où la distribution des réserves attendues peut être simulée, alors des hypothèses supplémentaires sont nécessaires.

Pour finir, nous notons que la moyenne et la variance de C_{ij} sous le modèle du Mack sont semblables à les celles de C_{ij} dans l'approximation Normale du modèle binomial négatif avec les paramètres d'échelle inconnus ϕ_j de l'approximation Normale qui sont remplacés par σ_j^2 dans le modèle de Mack. Donc, il est essentiel que les valeurs prédites et les erreurs de prédiction des deux modèles soient identiques (voir section suivante).

§6. Modèle Log Normale:

Au début, les oeuvres sur les modèles stochastiques sur la technique de Chain Ladder standard se sont concentrées sur le logarithme des paiements non cumulés $Y_{ij} = \log(X_{ij})$ et la classe de Log-Normale des modèles $Y_{ij} = m_{ij} + \varepsilon_{ij}$, les ε_{ij} étant les erreurs aléatoires indépendantes Normales. Donc,

$Y_{ij} \approx IN(m_{ij}, \sigma_{ij}^2)$ où $\varepsilon_{ij} \approx IN(0, \sigma^2)$

L'usage de la transformation logarithmique impose immédiatement une limitation sur cette classe de modèles, par conséquence, ces montants non cumules doivent être positifs. Les variables réponses Normales Y_{ij} sont supposées se décomposer (additivité) dans un composant non aléatoire déterministe avec la moyenne $m_{ij} = \eta_{ij}$ et des composants d'erreur aléatoire distribués selon une loi Normale centrée. On applique la structure Chain Ladder pour la moyenne comme on a introduit dans le modèle de Poisson sur-dispersion, cela donne:

$\eta_{ij} = c + \alpha_i + \beta_j$

Ce modèle Log-Normale a été présenté par Kremer (1982) et utilisé par Renshaw (1989), Verrall (1989), Zehnwirth (1989) et Christofides (1990) entre autre. L'usage de ce modèle produit normalement des valeurs prévues proches de celles de la technique de Chain Ladder simple. Mais ce n'est pas garanti, et il peut y avoir des différences.

La distribution Log-Normale a l'avantage de pouvoir être appliquée sans avoir recours à un logiciel spécialiste. Par exemple, Christofides (1990) a montré comment on pouvait utiliser une feuille de calcul Excel pour analyser les données d'un modèle Log-Normale.

Un autre avantage est que les autres techniques statistiques peuvent être également utilisées pour permettre aux différentes hypothèses d'être incorporées en concernant la type de run-off et les connections entre les années de l'origine. Puisque Y_{ij} est distribué suivant une loi Normale, il est possible, par exemple, d'appliquer le filtre de Kalman (Verrall, 1989).

Partie III : L'estimation des erreurs de prédiction et bootstrap

Chapitre 1: Estimation des erreurs de prédiction:

§1. Introduction :

Deux mesures de l'incertitude peuvent être utilisées pour estimer la variabilité. La première est basée sur l'estimation de la Moyenne des Carrés des Ecarts à la Prédiction (MCEP en anglais MSEP-Mean Squared Error of Prediction). La deuxième est donnée par la variabilité des projections réalisées à partir de données issues de rééchantillonnage des résidus et des valeurs attendues.

Soit X une variable aléatoire à estimer par un modèle linéaire généralisé. L'espérance du carré de l'écart à la prédiction de la variable X est donnée par :

$$E\left[(X-\hat{X})^2\right] \approx E\left[\left((X-E(X))-(\hat{X}-E(X))\right)^2\right]$$

$$E\left[(X-\hat{X})^2\right] \approx E\left[(X-E(X))^2\right]+E\left[(\hat{X}-E(X))^2\right]-2.E\left[(X-E(X)).(\hat{X}-E(X))\right]$$

L'estimateur obtenu par la méthode du maximum de vraisemblance est un estimateur sans biais. L'espérance de X peut être remplacée par l'espérance de la valeur attendue $E(\hat{X})$ dans les écarts $\hat{X}-E(X)$. Il vient :

$$E\left[(X-\hat{X})^2\right] \approx E\left[(X-E(X))^2\right]+E\left[(\hat{X}-E(X))^2\right]-2.E\left[(X-E(X)).(\hat{X}-E(\hat{X}))\right]$$

Par hypothèse, les observations sont issues de tirages indépendants. Les erreurs des valeurs passées et futures $\hat{X}-E(\hat{X})$ et $X-E(X)$ sont supposées non corrélées. La formule se simplifie.

$$E\left[(X-\hat{X})^2\right] \approx E\left[(X-E(X))^2\right]+E\left[(\hat{X}-E(\hat{X}))^2\right]$$

L'estimation de la moyenne des carrés à la prédiction d'un règlement X est donnée par :

$$MCEP \approx s^2(X) + s^2(\hat{X})$$

Cette technique est utilisée sur les modèles Chain Ladder et les modèles linéaires généralisés comme Poisson sur-dispesion, Log-normale…que nous allons introduire dans la section suivant.

§2. Estimation des erreurs de prédiction

2.1. Loi poisson sur-dispersion :

On a obtenu les estimations de réserve globale et les estimations d'année d'origine, maintenant il faut estimer les erreurs de prédiction.

On considère un paiement non cumulé X_{ij}. De plus, on considère l'année d'origine i et les paiements dans l'année de développement j. La moyenne des carrés des écarts à la prédiction est donnée par :

$$MCEP(\hat{X}_{ij}) = E\left[(X_{ij} - \hat{X}_{ij})^2\right] \approx Var(X_{ij}) - Var(\hat{X}_{ij}) \qquad \textbf{(1)}$$

On a,

$$Var(X_{ij}) = \phi m_{ij} \quad \textbf{(2)} \text{ et } \hat{X}_{ij} = \phi \hat{m}_{ij} = \exp(\hat{\eta}_{ij}) \qquad \textbf{(3)}$$

Ensuite, en utilisant l'équation 3 et les développement de Taylor

$$Var(\hat{X}_{ij}) = \left|\frac{\partial m_{ij}}{\partial \eta_{ij}}\right|^2 Var(\hat{\eta}_{ij}) \qquad \textbf{(4)}$$

Combinant 2 et 4 donne :

$$MCEP(\hat{X}_{ij}) \approx \phi \hat{m}_{ij} + \hat{m}_{ij}^{\,2} Var(\hat{\eta}_{ij}) \qquad \textbf{(5)}$$

On note que la variance du prédicteur linéaire est toujours directement disponible à partir des logiciels statistiques qui permettent de calculer l'erreur de prédiction sans difficulté.

Jusqu'à maintenant nous avons considéré un paiement non cumulé X_{ij}. De plus, on peut calculer les erreurs de prédiction pour les estimations de réserve d'année d'origine et les estimations de réserve globale. La variance de la somme des valeurs prévues est considérée, prenant compte de toutes covariances entre les valeurs prédites.

On note le triangle des règlements prévus qui contribuent aux estimations de réserve par Δ. Puis, l'estimation de réserve dans l'année de l'origine i est donné en sommant les valeurs prévues dans la ligne i de Δ , c'est

$$\hat{X}_{i+} = \sum_{j \in \Delta_i} \hat{X}_{ij} \qquad \textbf{(6)}$$

De Renshaw(1994b), la moyenne des carrés des écarts à la prédiction de l'année d'origine est donnée par:

$$MCEP(\hat{X}_{i+}) \approx \sum_{j \in \Delta_i} \phi \hat{m}_{ij} + \sum_{j \in \Delta_i} \hat{m}_{ij}^{2} Var(\hat{\eta}_{ij}) + 2 \sum_{\substack{j_1, j_2 \in \Delta_i \\ j_2 \rangle j_1}} \hat{m}_{ij_1} \hat{m}_{ij_2} Cov(\hat{\eta}_{ij_1} \hat{\eta}_{ij_2}) \quad \textbf{(7)}$$

L'estimation de la réserve globale est donnée par :

$$\hat{X}_{++} = \sum_{i, j \in \Delta_i} \hat{X}_{ij} \quad \textbf{(8)}$$

Et la moyenne des carrés des écarts à la prédiction de la provision globale est donnée par:

$$MCEP(\hat{X}_{++}) \approx \sum_{i, j \in \Delta} \phi \hat{m}_{ij} + \sum_{i, j \in \Delta} \hat{m}_{ij}^{2} Var(\hat{\eta}_{ij}) + 2 \sum_{\substack{i_1, j_2 \in \Delta \\ i_2, j_2 \in \Delta \\ i_1, j_1 \neq i_2, j_2}} \hat{m}_{i_1 j_1} \hat{m}_{i_2 j_2} Cov(\hat{\eta}_{i_1 j_1} \hat{\eta}_{i_2 j_2})$$

On note que le premier terme dans l'année d'origine et les erreurs de prédiction globale sont simplement la somme de la variance des processus. Les autres termes se rapportent à la variance d'estimation.

Plus loin, la terme covariance dans les équations (7) et (9) n'est pas disponible à partir de logiciels statistiques. Cependant, la matrice des variances-covariances des évaluations de paramètre peut être extraite du logiciel statistique. Une matrice des termes de covariance peut être calculée en formant la matrice explicative des paiements futurs et exécutant quelques opérations matricielles simples. En fait, les variances des prédicteurs linéaires sont simplement la diagonale de cette matrice. Nous avons réussi dans les calculs des erreurs de prédiction en deux moyens : Excel et R. Dans R, il s'agit de créer une nouvelle famille pour la fonction Quasi- Poisson. Les codes et les étapes sont écrits dans l'appendice

Dans l'appendice, les données (lesquelles représentent des montants de règlements non cumulé, données dans le Tableau 1 de la documentation England and Verral) sont utilisées pour illustrer la méthodologie permettant de calculer la réserve ainsi que les erreurs de prédiction pour le modèle de Poisson sur-dispersion, comme expliqué ci-dessus. Le tableau 2 présente les résultats des erreurs de prédiction, en utilisant les équations 7 et 9. Comme on peut l'observer, il y a une incertitude considérable dans la provision, en particulier dans les premières années de développement où les provisions mathématiques sont petites. Les erreurs de prédiction, en pourcentage de la moyenne, sont grandes les premières années, et tout particulièrement pour l'année la plus récente. A cause de cette erreur d'estimation, il est nécessaire d'utiliser un plus grand nombre des paramètres que dans le modèle poisson sur-dispersion. L'explication détaillée de l'exemple peut être trouvée dans l'article original de England et Verrall

2.2 <u>Le modèle binomial négatif</u> :

Dans le modèle binomial négatif, au contraire du modèle de Poisson sur-dispesion, les paramètres différents correspondent au délai de règlement j, il n'y a aucun paramètre de l'année d'origine. Pour calculer la variabilité de MCEP, il faut prendre en compte les estimateurs pour la variance du processus et la variance de l'estimation.

Pour les évaluations de réserve de l'année d'origine, les ultimes, U_{ij}, sont les paiements cumulés dans le dernier délai de règlement. Donc,

$$U_i = C_{in} \quad \textbf{(10)}$$

Les évaluations des réserves de l'année d'origine, R_i , sont $U_i - C_{i,n-i+1}$, où C_{i-n+1} est payé à date connu. D'où :

$$Var(U_i) = Var(C_{in}) = Var(R_i) \qquad \textbf{(11)}$$

$$Var(\hat{U}_i) = Var(\hat{C}_{in}) = Var(\hat{R}_i) \qquad \textbf{(12)}$$

Les variances d'estimation et les variances du processus de l'année d'origine peuvent être estimé en considérant $Var(C_{in})$ et $Var(\hat{C}_{in})$

Premièrement, on considère la variance du processus, $Var(C_{in})$. Les procédures récursives sont nécessaires pour obtenir la variance de l'estimation et la variance du processus car le modèle binomial négatif est formulé comme un modèle récursif. Le calcul de la variance du processus implique d'estimer la variance de la prévision des k premiers pas où k= i-1.

Pour un modèle binomial négatif,

$$Var(C_{in}) = C_{i,n-i+1} \prod_{k=n-i+2}^{n} \hat{\lambda}_k \left(\prod_{k=n-i+2}^{n} \hat{\lambda}_k - 1 \right) \qquad \textbf{(13)}$$

La variance d'estimation est calculé par

$$Var(C_{in}) = Var\left[C_{i,n-i+1} \prod_{k=n-i+2}^{n} \hat{\lambda}_k \right] = C_{i,n-i+1}^2 Var\left[\prod_{k=n-i+2}^{n} \hat{\lambda}_k \right] \qquad \textbf{(14)}$$

On note que la variance du produit des facteurs de développement dans l'équation (14) peut être calculée car les paramètres dans le modèle binomial négatif se rapportent aux facteurs de développement λ_i , et la matrice de covariance des évaluations du paramètre est disponible.
L'évaluation de la réserve globale et la variance du processus peuvent être estimés en considérant $Var(R_+)$ et $Var(\hat{R}_+)$ respectivement, où

$$R_+ = \sum_{i=2}^{n} R_i \qquad \textbf{(15)}$$

La variance du processus de la réserve globale est la somme des variances du processus d'estimation de chaque année d'origine, en supposant qu'il y a indépendance entre les années. La variance d'estimation de la réserve globale est donné par :

$$Var(R_+) = \sum_{i=2}^{n} Var(\hat{C}_{in}) + 2 \sum_{\substack{i=2 \\ j \rangle i}}^{n} Cov(\hat{C}_{in}, \hat{C}_{jn}) \qquad \textbf{(16)}$$

Donc, la variance de l'estimation de la réserve globale est la somme des variances des estimations des réserves de chaque année d'origine, avec un composant supplémentaire qui est la covariance entre les années.

L'appendice, Tableau 6, montre les évaluations de la réserve qui sont calculées à partir des facteurs de développement. Les résultats de MCEP des réserves sont calculés en utilisant l'équation (13), (14) et (16). On peut observer que l'évaluation de la réserve pour le modèle Poisson sur-dispersion et pour le modèle binomial négatif sont identiques et leurs variabilités de MCEP sont très proches. La Table 7 montre une comparaison entre les variances des processus et les variances des estimations pour le modèle de Poisson sur-dispersion et pour le modèle binomial négatif, pondérées par les paramètres de la dispersion pour éviter toute distorsion. Il est essentiel que la variabilité de MCEP soit la même pour les deux modèles (les différences sont dues aux arrondis des calculs numériques). On note pourtant que la variance d'estimation du modèle Poisson sur-dispersion est plus grande que le modèle binomial négatif du fait du plus grand nombre des paramètres utilisés dans le modèle Poison sur-dispersion. En conséquence, la variance du processus du modèle binomial négatif est plus grande que celle du modèle Poisson sur-dispersion. Mais, les évaluations de la réserve obtenues en utilisant les deux modèles sont identiques. Ce qui implique que le choix du modèle n'est pas important car le résultat sera le même.

2.3)***Approximation Loi de Normale à Modèle binomial négatif***

On note que les paramètres inconnus dans la variance du modèle Normal peuvent être estimés en utilisant la procédure itérative « Joint Modelling ». La table 8, dans l'Appendice, montre les évaluations des paramètres et les erreurs standards obtenues par l'adéquation du « Joint Modelling » au triangle des données de la Table 1. Plus loin, les facteurs du développement et leurs erreurs standard, obtenus par des évaluations et de la matrice de covariance, sont montrés dans Table 9. La Table 10 montre les évaluations des composants de variance ϕ_j . L'information est insuffisante pour estimer la variance ϕ_{10} parce qu'il y a seulement un résidu pour ce point.

La théorie principale pour calculer des erreurs de prédiction pour le modèle Normal est la même que pour le modèle binomial négatif. Car le modèle Normal est aussi formulé comme un modèle récursif, les procédures récursives sont nécessaires pour obtenir la variance de l'estimation et la variance du processus.

Pour calculer la variance du processus, il est demandé estimer la variance d'une prévision k - pas en avant où k=i-1

Pour la variance du processus, nous savons que la variance d'une prévision 1- pas en avant est donnée par :

$$Var(C_{j+1}\left|X_1, X_2,...,X_{j-1}\right.) = \phi_j C_{j-1} \qquad \textbf{(17)}$$

Puis, la variance de la prévision 2-pas-en avant est donnée par :

$$Var(C_{j+1}\left|X_1, X_2,...,X_{j-1}\right.) = E(\phi_{j+1}C_j\left|X_1, X_2,...,X_{j-1}\right.) + \lambda_{j+1}^2\phi_j C_{j-1} \qquad (18)$$

$$= (\phi_{j+1}\lambda_j + \lambda_{j+1}^2\phi_j)C_{j-1}$$

Et la variance de la prévision 3-pas-en avant :

$$Var(C_{j+2}\big|X_1,X_2,...,X_{j-1}) = E(\phi_{j+2}\lambda_{j+1}C_j + \lambda_{j+1}^2\phi_{j+1}C_j\big|X_1,X_2,...,X_{j-1}) + \lambda_{j+2}^2\lambda_{j+1}^2\phi_j C_{j-1}$$
$$= (\lambda_{j+1}\lambda_j\phi_{j+2} + \lambda_{j+1}^2\lambda_j\phi_{j+1} + \lambda_{j+2}^2\lambda_{j+1}^2\phi_j)C_{j-1} \quad (19)$$

D'où, il est facile de continuer le processus pour obtenir k-pas en avant.

La variance d'estimation peut être calculé récursivement. On sais que :

$$Var(\hat{C}_n\big|C_{n-i+1}) = Var(\hat{\lambda}_{n-i+2}..\hat{\lambda}_n C_{n-i+1}\big|C_{n-i+1}) = C_{n-i+1}^2 Var(\hat{\lambda}_{n-i+2}..\hat{\lambda}_n\big|C_{n-i+1}) \quad \textbf{(20)}$$

Et utilisant

$$Var\left[\hat{\lambda}_{n-2}(\hat{\lambda}_{n-1}\hat{\lambda}_n)\right] = E(\hat{\lambda}_{n-2})^2 Var(\hat{\lambda}_{n-1}\hat{\lambda}_n) + E(\hat{\lambda}_{n-1}\hat{\lambda}_n)^2 Var(\hat{\lambda}_{n-2}) + Var(\hat{\lambda}_{n-1}\hat{\lambda}_n)Var(\hat{\lambda}_{n-2})$$
(21)

Puis, pour la provision globale, la variance de l'estimation est donnée par :

$$Var(\hat{R}_+) \approx \sum_{i=2}^{n} Var(\hat{C}_{in}) + 2\sum_{\substack{i=2 \\ j>i}}^{n} Cov(\hat{C}_{in},\hat{C}_{jn}) \qquad \textbf{(22)}$$

La Table 8, dans l'appendice, montre les évaluations de provisionnement et leurs erreurs de prédiction calculées en utilisant les équations ci-dessus. On note que les variabilités MCEP pour le modèle Normal sont inférieures dans les premières années et plus haut dans les années suivantes par rapport aux résultats du modèle binomial négatif dans la Table 6. Aussi, la variabilité MCEP totale est plus haute pour l'approximation Normale que pour le modèle binomial négatif à cause de la variabilité MCEP totale des années récentes.

2.4) *<u>La modèle de MACK</u>* :

Dans Mack (1993), les formules sont développées pour les erreurs de prédiction de l'année d'origine et la réserve globale en terme de règlements cumulés C_{ij}, les facteurs de développement λ_j, et les variances σ_j^2.

D'après Mack (1993), les variances du processus des réserves de l'année d'origine sont données par :

$$Var(R_i) \approx \hat{C}_{in}^2 \sum_{k=n-i+1}^{n-1} \frac{\hat{\sigma}_{k+1}^2}{\hat{\lambda}_{k+1}^2 \hat{C}_{ik}}$$

(23)

Les variances d'estimation sont données par :

$$Var(\hat{R}_i) \approx \hat{C}_{in}^2 \sum_{k=n-i+1}^{n-1} \frac{\hat{\sigma}_{k+1}^2}{\hat{\lambda}_{k+1}^2 \sum_{q=1}^{n-k} C_{qk}} \qquad \textbf{(24)}$$

Donc,

$$MCEP\ (\hat{R}_i) \approx \hat{D}_{in}^2 \sum_{k=n-i+1}^{n-1} \frac{\hat{\sigma}_{k+1}^2}{\hat{\lambda}_{k+1}^2} \left(\frac{1}{\hat{D}_{ik}} + \frac{1}{\sum_{q=1}^{n-k} D_{qk}} \right) \qquad \textbf{(25)}$$

Pour l'erreur de prédiction de la réserve globale, un ajustement de covariance a besoin de la variance d'estimation, donné par,

$$MCEP(\hat{R}_+) \approx \sum_{i=2}^{n} \left\{ MCEP(\hat{R}_i) + \hat{C}_{in} \left(\sum_{q=i+1}^{n} \hat{C}_{qn} \right) \times \sum_{k=n-i+1}^{n-1} \frac{2\hat{\sigma}_{k+1}^2}{\hat{\lambda}_{k+1}^2 \sum_{q=1}^{n-k} C_{qk}} \right\} \qquad \textbf{(26)}$$

Où :

$$\hat{R}_+ = \sum_{i=2}^{n} \hat{R}_i \qquad \textbf{(27)}$$

Chapitre 2: bootstrap

§1. Introduction.

Les méthodes classiques d'inférence statistique ne permettent pas d'obtenir des réponses correctes à tous les problèmes concrets que se pose l'utilisateur. Elles ne sont en effet valables que sous des conditions d'application particulières. Ainsi, par exemple, le test *t* de Student d'égalité des moyennes suppose que les deux populations-parents sont normales, de même variance et que les deux échantillons sont aléatoires, simples et indépendants. Le calcul de l'intervalle de confiance d'une variance par l'intermédiaire des variables $\chi 2$ suppose que la population-parent est normale et que l'échantillon est aléatoire et simple. L'inférence statistique classique en régression suppose, outre les conditions d'application relatives à la population et à l'échantillon, que le modèle est ajusté au sens des moindres carrés. Que peut faire l'utilisateur, en pratique, lorsque ces conditions d'application ne sont pas remplies? Différentes attitudes sont possibles. Dans certains cas, les méthodes classiques sont utilisées malgré le non-respect des conditions. Cette utilisation est alors justifiée par le caractère robuste des méthodes qui garantit que les résultats de l'inférence restent approximativement valables. C'est par exemple le cas pour la comparaison de moyennes de populations qui s'écartent modérément des distributions normales, lorsque les effectifs sont suffisamment grands.

Le recours à des transformations de variables permet, dans certains cas, de se rapprocher des conditions d'application. Ainsi une transformation logarithmique, par exemple, peut rendre normales des distributions qui, au départ, ne le sont pas. Une troisième attitude consiste à abandonner les méthodes paramétriques d'inférence statistique au profit de méthodes non paramétriques, pour lesquelles les conditions d'application sont bien moins restrictives. Pour le

problème de la comparaison de deux moyennes évoqué ci-dessus, le test *t* de Student sera par exemple remplacé par le test des rangs de Wilcoxon, aussi appelé test de Mann-Whitney (Dagnelie, 1998). Les solutions proposées ci-dessus permettent incontestablement d'élargir l'éventail des problèmes auxquels une solution peut être apportée. Elles ne permettent cependant pas de résoudre tous les problèmes. L'accès généralisé à des moyens de calcul puissants a permis le développement de méthodes d'inférence statistique basées sur l'utilisation intensive de l'ordinateur. Le *bootstrap* fait partie de ces méthodes. Le mot *bootstrap* provient de l'expression anglaise "*to pull oneself up by one's bootstrap*" (Efron, Tibshirani, 1993), qui signifie littéralement "se soulever en tirant sur les languettes de ses bottes". Le mot *bootstrap* fait penser à des traductions telles que "à la force du poignet" ou "par soi-même" ou "passe partout" (Dagnelie, 1998), mais en fait il n'est jamais traduit dans la littérature scientifique d'expression française.

Le principe général de la méthode est de rééchantillonner un grand nombre de fois l'échantillon initial qui a été réellement prélevé dans la population, l'inférence statistique étant basée sur les résultats des échantillons ainsi obtenus. L'objectif de cette section est de décrire comment le *bootstrap* peut être utilisé pour résoudre les problèmes d'inférence statistique en relation avec l'estimation des paramètres, tout particulièrement dans le provisionnement. Je vais présenter d'abord le principe de rééchantillonnage des résidus dans deux modèle de Poisson et Chain Ladder. Ensuite, j'examine l'estimation de l'erreur-standard. Je donne alors quelques informations dans l'échantillon avant de conclure. Des informations plus détaillées concernant le *bootstrap* et d'autres méthodes de rééchantillonnage sont données, notamment, par Chernick (1999), Efron et Tibshirani (1993) et Manly (1997).

§2. Le principe du Bootstrap.

Le principe est de s'écarter du modèle de référence en réalisent des tirages aléatoires et d'observer la sensibilité de ceux-ci. Les tirages aléatoires peuvent aussi être réalisés sur les paramètres. C'est le cas du tirage aléatoire réalisé sur les modèles Chain Ladder ou sur les résidus dans le modèle de Poisson présenté dans l'application à la section suivant. Dans ce cas, le tirage aléatoire est un rééchantillonnage des résidus. Le rééchantillonnage des résidus d'un modèle linéaire généralisé permet d'estimer la variabilité de la provision. Les paragraphes suivants montrent comment le rééchantillongage des résidus permet de créer de nouveaux jeux de données. Une analyse par jeu de données sera réalisée. Ces analyses produiront une distribution empirique des règlements futurs. Il suffira alors de choisir un quantile de cette distribution.

L'hypothèse sous-jacente des modèles linéaires généralisés est que les résidus standardisés suivent une même loi asymptotiquement gaussienne. Par conséquent les résidus observés peuvent être considérés comme des tirages aléatoires d'une même loi de distribution. Cette loi est inconnue et ne permet pas de réaliser des tirages aléatoires sans celle-ci. Cependant, il est possible de réaliser des tirages dans un échantillon de cette loi, constitué par les résidus standardisés.

Le principe du Bootstrap est de réaliser autant de tirages avec remise qu'il existe d'éléments dans l'échantillon. Dans notre cas, un nouveau vecteur r_P^{boot} est constitué à partir du vecteur des résidus standardisés de Pearson r_P **.** Le nombre d'éléments de ces vecteurs est identique. Un rééchantillonage aurait pu être réalisé de la même façon à partir des résidus de la déviance.

A partir des résidus standardisés et des valeurs attendues, les valeurs observées peuvent être calculées. Les résidus de Pearson s'écrivent :

$$r_P = \frac{X_{ij} - \hat{\mu}_{ij}}{\sqrt{\hat{\mu}_{ij}}}$$

Les résidus de Pearson ajustés sont donnés par :

$$r_P^{adt} = \sqrt{\frac{n}{1/2n(n+1) - 2n + 1}} \times r_P$$

L'équation peut être inversée. Alors les règlements non cumulés sont donnés par :

$$\hat{X}_{ij} = r_P^{boot} \sqrt{\hat{\mu}_{ij}} + \hat{\mu}_{ij}$$

Le modèle réalisé avec les donnes sources est utilisé avec chaque jeu de données. Autant de projections que de jeux de données sont obtenues. A partir de ces projections, une distribution des règlements futurs est estimée. Il suffira alors de calculer les moments de cette loi et de choisir un quantile en fonction du risque choisi pour obtenir le montant de la provision.

§3. Application de la méthode du Boostrap :

À partir du modèle linéaire généralisé réalisé sur le triangle de données de MAAF SA Auto 2005 hors IBNR (KM). On va appliquer la méthode du Bootstrap sur le modèle de Poisson :

Données d'entrée cumulées

	Dvlpt						Ultimes	Réserves
Surv	j=1	2	3	4	5	6		
i=1	275 923	328 048	328 292	328 970	329 348	329 667	329 667	
2	311 162	369 368	369 580	370 355	370 424	.	370 783	359
3	330 934	377 788	378 617	379 264	.	.	379 874	610
4	318 204	364 370	365 250	.	.	.	366 552	1 302
5	312 855	357 096	.	.	.	.	358 908	1 812
6	336 247	.	.	.	.	.	391 968	55 722

Facteurs de développements Chain-Ladder

1.160	1.002	1.002	1.001	1.001

Données décumulées

	Dvlpt					
Surv	j=1	2	3	4	5	6
I=1	275 923	52 125	244	678	378	319
2	311 162	58 206	211	776	69	

3	330 934	46 854	829	647
4	318 204	46 165	881	
5	312 855	44 242		
6	336 247			

En maximisant la Log vraisemblance, on obtient les paramètres α et β

	Paramètre du modèle
Mu	19.4602
alpha1	
alpha2	0.1176
alpha3	0.1418
alpha4	0.1061
alpha5	0.0850
alpha6	0.1731
beta1	
beta2	-1.8336
beta3	-6.3513
beta4	-6.0892
beta5	-7.2032
beta6	-6.7875

Puis, on calcule les valeurs prédites du modèle de Poisson

Espérance

Surv	Dvlpt j=1	2	3	4	5	6
i=1	282 791	45 199	493	641	210	319
2	318 077	50 839	555	721	237	359
3	325 873	52 085	568	739	243	367
4	314 443	50 258	549	713	234	355
5	307 885	49 210	537	698	229	347
6	336 244	53 743	587	762	250	379

Alors, on obtient le triangle des résidus de Pearson non standardisés avec la formule

$$r_P = \frac{X_{ij} - \hat{\mu}_{ij}}{\sqrt{\hat{\mu}_{ij}}}$$

	j=1	2	3	4	5	6
i=1	-408.43	1030.23	-355.40	46.55	364.85	0.38
2	-387.72	1033.21	-461.07	64.01	-344.02	
3	280.33	-724.81	346.02	-106.88		
4	212.13	-577.27	448.48			
5	283.22	-708.19				
6	0.13					

A l'étape suivante, on applique la méthode du Bootstrap au triangle des résidus de Pearson. On obtient ainsi 250 triangles de résidus par tirage uniforme des résidus initiaux

Rééchantillon

1033.21	283.22	-708.19	-106.88	280.33	283.22
1030.23	1030.23	-724.81	1030.23	0.13	
-344.02	0.13	-708.19	-408.43		
-577.27	-387.72	-577.27			
0.38	212.13				
0.38					

Puis, on reconstitue pour chaque triangle de résidus le triangle des paiements avec l'équation :

$$\hat{X}_{ij} = r_P^{boot} \sqrt{\hat{\mu}_{ij}} + \hat{\mu}_{ij}$$

En connaissant le triangle des $\hat{\mu}_{ij}$ et r_P^{boot}, on en déduit celui des paiements annuels.

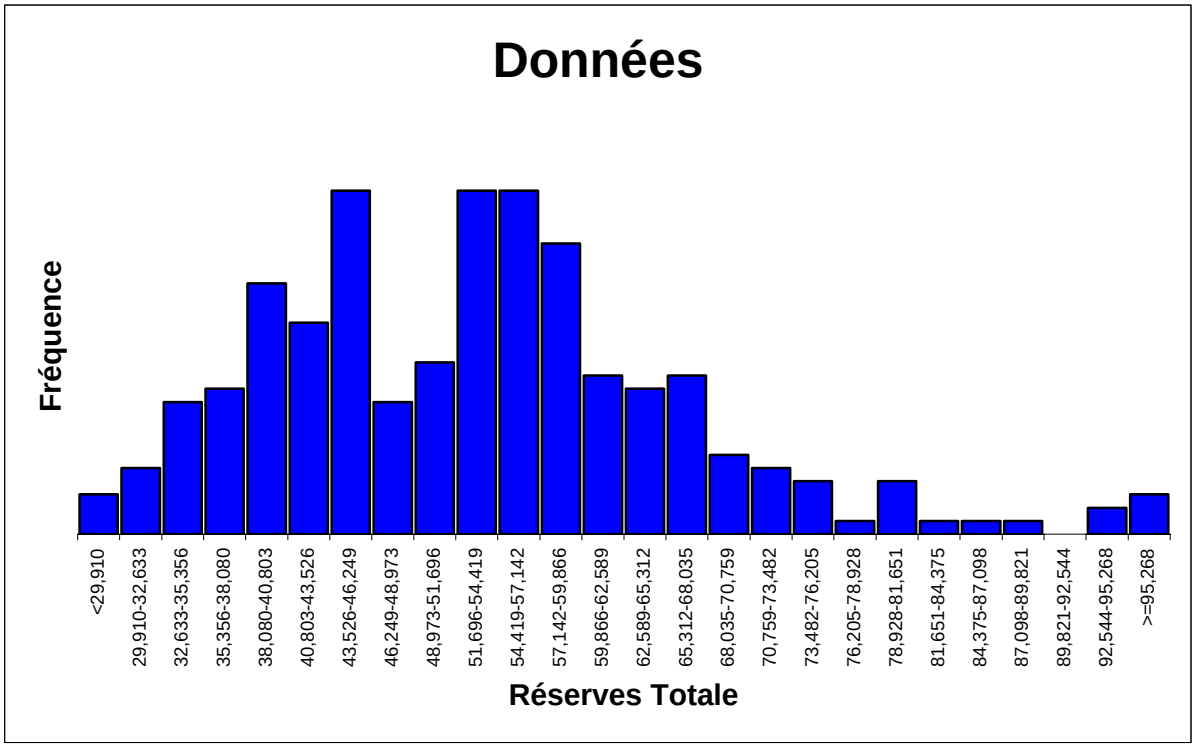
	j = 1	2	3	4	5	6
i = 1	300 166	347 269	347 265	347 821	348 160	348 639
2	336 451	394 636	394 651	396 247	396 483	
3	319 663	371 749	371 783	372 171		
4	304 206	351 716	351 837			
5	307 892	358 590				
6	336 252					

En procédant de même pour chaque triangle bootstrapé, on obtient un 250-échantillon de réalisations de notre variable « provision ».

Résultats de Simulations Bootstrap

Moyen	60 025	350	590	1 288	1 808	55 990
Std Dev	3 812	362	416	492	541	3 107
Proc SD	3 383	258	335	496	587	3 267
Total SD	5 096	444	534	698	798	4 509
C.V.:	**0.08**	**1.27**	**0.91**	**0.54**	**0.44**	**0.08**

q60	60 312
q70	61 156
q80	62 405
q90	63 975
q95	64 947
q99	65 974



"Le *bootstrap* est une méthode d'inférence statistique basée sur l'utilisation de l'ordinateur qui peut répondre sans formule à beaucoup de questions statistiques réelles". C'est en ces termes qu'Efron et Tibshirani (1993) présentent la méthode du Bootstrap dans la préface de leur ouvrage. Il est incontestable que l'utilisation des techniques de rééchantillonnage a été rendue possible grâce à la généralisation des moyens de calculs performants. Ces techniques reposent, au départ, sur des idées simples. Toutefois, il faut bien admettre que les développements apportés aux méthodes de base leur ont fait perdre une partie de cette simplicité.

Dans cette section, nous nous sommes limités au problème de l'estimation de la provision des règlements. Il ne s'agit cependant pas là des seules applications des méthodes de rééchantillonnage. Celles-ci peuvent, en effet, aussi être utilisées pour la réalisation de différents tests d'hypothèses, pour le choix des variables et l'estimation de l'erreur de prédiction en régression, pour l'estimation du taux d'erreur en analyse discriminante, notamment. Le caractère relativement général des problèmes qui peuvent être résolus par la méthode du bootstrap ne doit pas faire perdre de vue que la qualité de l'inférence dépend de la nature de la question posée et de la disponibilité des données. Comme le suggère Manly (1997), la méthode du Bootstrap doit être utilisée avec prudence dans les situations où elle n'a pas encore été testée de manière approfondie.

Une discussion assez générale sur l'intérêt et les limites de la méthode du bootstrap est donnée dans la synthèse proposée par Young (1994) et dans les commentaires suscités par l'article en question.

Enfin, l'absence dans les logiciels classiques de fonctions permettant d'utiliser la méthode du bootstrap, sans obligation pour l'utilisateur de programmer lui-même les procédures de calcul, est certainement un frein à l'utilisation du bootstrap dans les cas où il pourrait être utilisé.

CONCLUSION :

L'évaluation de la force financière d'une compagnie d'assurance non-vie inclut une analyse des règlements attendus et une estimation de la variabilité possible du provisionnement. Toute erreur de calcul dans le provisionnement pourra causer l'insolvabilité de certains assureurs. Il est important de choisir la meilleure méthode pour l'estimation des réserves afin d'éviter tous effets négatifs dans la position financière future de l'entreprise.

Dans les premières sections, les modèles déterministes traditionnels avec la technique de Chain-Ladder ont été examinés. Ensuite, nous nous sommes intéressés aux modèles stochastiques qui utilisent la même technique que le Chain-Ladder, comme l'a démontré England & Verrall (2002).

Ces dernières années, le sujet du provisionnement a été à l'origine de plusieurs controverses, principalement au sujet du choix du modèle le plus performant à utiliser pour obtenir la meilleure estimation des réserves. Mais, il n'y a pas de réponse particulière à une telle question. Certains modèles s'adapteront à certains problèmes et type de données. Lorsqu'un calcul de provision est réalisé, les données devraient être examinées en détail pour trouver un modèle approprié, plutôt que d'utiliser la même approche de modélisation dans toutes les circonstances, que le modèle soit déterministe ou stochastique. Calculer des réserves est un exercice pratique d'analyse des données pour lequel il est vital de s'efforcer de comprendre et apprendre des données plutôt que d'imposer la même approche dans toutes les situations.

Parmi les modèles stochastiques que nous avons exposés, il est difficile de choisir entre un Poisson sur-dispersion et un modèle binomial négatif. En effet, les prédictions et les erreurs de prédiction sont les mêmes ainsi que les restrictions des données négatives bien que le modèle binomial négatif paramétré est plus logique avec le modèle de Chain Ladder traditionnel. Mais, la mise en oeuvre du modèle binomial négatif est complexe parce que la disponibilité de logiciel pour adapter ce modèle est limitée. Par exemple, dans le logiciel R et SAS, on ne trouve pas le modèle Binomial Négatif et il faut écrire les codes. L'approximation normale du modèle binomial négatif offre plus grande flexibilité en acceptant la présence de quelques données négatives. Il est important de noter que le modèle Mack et l'Approximation Normale sont semblables. Notez pourtant, que le modèle de Mack n'est pas recommandé pour être utilisé dans toutes les situations bien qu'il est mois pénible en practice.

Pour toute affaire de l'assurance, un actuaire est exigé pour certifier que les réserves calculées sont ''at least as large as those implied by a "best estimate" basis without precautionary margins'(Guidance Notes 20 and 33, Manual of Actuarial Practice, Institute of Actuaries). Par le terme "best estimate" on entend la valeur estimée de la distribution des résultats possibles des passifs non réglés. Il est essentiel que cette valeur soit la moyenne de la distribution prédite.

Les valeurs de l'erreur de prédiction illustrées dans Partie III dépendront principalement du nombre de données disponibles et des limitations des modèles pour l'estimation. Par ailleurs, l'erreur de la prédiction reflète les hypothèses sous-jacentes du modèle. Si les hypothèses sous-jacentes sont inexactes, les erreurs de prédiction seront aussi inexactes. Notez que les modèles stochastiques considèrent les variations statistiques d'après une distribution statistique particulière (Poisson sur-dispersion, Normale, Log Normale…). Les hypothèses sont validées par l'observation des données de l'année suivante.

En conclusion, durant les vingt dernières années, il y a eu un intérêt croissant pour les modèles stochastiques, malgré cela leur utilisation est toujours limitée, due à un certain nombre de raisons dont un manque de compréhension générale des méthodes ainsi que leur manque de flexibilité. Cependant, la raison principale est probablement le manque de besoin pour les méthodes, quand les méthodes déterministes traditionnelles suffisent pour le calcul d'une meilleure évaluation des réserves. Depuis ces dernières années, il y a un intérêt grandissant pour l'estimation du risque d'erreur de prédiction dans le calcul des réserves et l'impact de cette erreur sur le bilan. Pour cette raison, il est essentiel d'être capable d'estimer la variabilité de réserves et idéalement, d'être

capable d'estimer une distribution pour les résultats possibles, et comment obtenir les centiles de cette distribution. Cela peut être accompli par l'usage des modèles stochastiques. Donc, il est important que les caractéristiques des méthodes stochastiques soient explorées, et leurs hypothèses comprises, et nous espérons que ce mémoire contribue à ce processus.

APPENDICE :

Les erreurs de prédiction dans les modèles MACK, Poisson sur-dispersion et Binomial négatif

Poisson sur-dispersion :

Tableau 1 : Les paramètres estimés

	Paramètre du Modèle	Erreur
Mu	**7.6550**	0.3193
alpha2	**-0.1108**	0.3450
alpha3	**0.2459**	0.3185
alpha4	**0.4214**	0.3100
alpha5	**0.4291**	0.3130
alpha6	**0.0348**	0.3538
alpha7	**-0.0593**	0.3819
alpha8	**0.2432**	0.3786
alpha9	**-0.1602**	0.5143
alpha10	**-0.0231**	0.7816
beta2	**0.6929**	0.2685
beta3	**0.6261**	0.2784
beta4	**0.2770**	0.3115
beta5	**0.0606**	0.3417
beta6	**-0.1958**	0.3885
beta7	**-1.0831**	0.6079
beta8	**-1.2736**	0.7893
beta9	**-1.9160**	1.3617
beta10	**-2.5074**	2.4911

Tableau 2 : Les résultats de réserves

	Réserves	Erreur de prédiction	%Erreur de prédiction
Ligne 2	153.95	556.0	361%
Ligne 3	617.37	1119.9	181%
Ligne 4	1636.14	1775.4	109%
Ligne 5	2746.74	2231.2	81%
Ligne 6	3649.10	2439.9	67%
Ligne 7	5435.30	3124.2	57%
Ligne 8	10907.19	5031.6	46%
Ligne 9	10649.98	6075.3	57%
Ligne 10	16339.43	12987.3	79%
Totale	52135.22	18193.1	35%

Le modèle binomial négatif :

Tableau 3 : Les paramètres estimés

	Paramètre du modèle	Erreur
Mu	0.6928	0.2733
alpha2	-1.1652	0.3490
alpha3	-1.9989	0.3675
alpha4	-2.4550	0.3909
alpha5	-2.8698	0.4318
alpha6	-3.8645	0.6416
alpha7	-4.0691	0.8207
alpha8	-4.7711	1.3945
alpha9	-5.3796	2.5399

Tableau 4: Les résultats du Modèle Binomial Négatif.

		Pearson				Déviance			
Année	**Réserve**	**Variance du processus**	**Variance d'estimation**	**Erreur**	**% Erreur**	**Variance du processus**	**Variance d'estimation**	**Erreur**	**% Erreur**
1									
2	154	282 870	253 192	732	476%	168 853	151 137	566	367%
3	617	1 153 550	1 021 176	1 475	239%	688 586	609 568	1 139	185%
4	1 636	3 158 803	2 308 667	2 338	143%	1 885 578	1 378 108	1 807	110%
5	2 747	5 525 339	3 112 161	2 939	107%	3 298 229	1 857 736	2 271	83%
6	3 649	8 172 848	2 155 716	3 214	88%	4 878 601	1 286 807	2 483	68%
7	5 435	14 263 235	2 675 602	4 116	76%	8 514 123	1 597 142	3 180	59%
8	10 907	36 375 971	7 573 975	6 629	61%	21 713 833	4 521 117	5 122	47%
9	10 650	57 664 556	6 415 282	8 005	75%	34 421 584	3 829 461	6 185	58%
10	16 339	265 353 995	27 734 300	17 120	105%	158 397 208	16 555 378	13 227	81%
Totale	52 135	391 951 167	183 148 547	23 981	46%	233 966 594	109 326 481	18 528	36%

Le modèle Approximation Normal au modèle binomial négatif :

Tableau 6 : Les paramètres estimés

	Paramètre du modèle	Erreur
Mu	2.9994	1.1781
alpha2	-1.3758	1.1865
alpha3	-1.7285	1.1818
alpha4	-1.8277	1.1784
alpha5	-1.8860	1.1786
alpha6	-1.9574	1.1783
alpha7	-1.9661	1.1781
alpha8	-1.9824	1.1782
alpha9	-1.9901	1.1782

Dispersion 1.2222

Tableau 7 : Les résultats du modèle Approximation Normal au modèle binomial négatif

Année	Réserve	Erreur de Prédiction	% Erreur de Prédiction
1			
2	154	371	241%
3	617	643	104%
4	1636	764	47%
5	2747	1361	50%
6	3649	1822	50%
7	5435	2017	37%
8	10907	5007	46%
9	10650	5939	56%
10	16339	23375	143%
Totale	52135	25706	49%

Le Modèle MACK

Tableau 8 : Les données modifiés de Auto dommage MAAF

Source:Auto dommage MAAF

"i"	k = 1	K = 2	k = 3	k = 4	k = 5	k = 6
1	275 923	328 048	328 292	328 970	329 348	329 667
2	311 162	369 368	369 580	370 355	370 424	.
3	330 934	377 788	378 617	379 264	.	.
4	318 204	364 370	365 250	.	.	.
5	312 855	357 096	.	.	.	.
6	336 247	.	.	.	.	.

Tableau 9 : Les calculs sur variance multiplie :

Processus Variance Multiplie

Annual	162397.188	332.309759	17.3067097	161.059566	17.2958708
Reserve	162925.16	527.971906	195.662146	178.355437	17.2958708

Paramètres Variance Multiplie

Annual	7.3892^{E}-05	1.8403E-07	1.1961E-08	1.4908E-07	2.4692E-08
Reserve	7.4262^{E}-05	3.6976E-07	1.8574E-07	1.7378E-07	2.4692E-08

Tableau 10 : **Matrice du facteur erreur d'estimation**

		2	3	4	5	6
		7.42621^{E}-05	3.69763E-07	1.85738E-07	1.73777E-07	2.46925E-08
2	7.42621E-05	7.42621^{E}-05	7.42621E-05	7.42621E-05	7.42621E-05	7.42621E-05
3	3.69763E-07	7.42621^{E}-05	3.69763E-07	3.69763E-07	3.69763E-07	3.69763E-07
4	1.85738E-07	7.42621^{E}-05	3.69763E-07	1.85738E-07	1.85738E-07	1.85738E-07
5	1.73777E-07	7.42621^{E}-05	3.69763E-07	1.85738E-07	1.73777E-07	1.73777E-07
6	2.46925E-08	7.42621^{E}-05	3.69763E-07	1.85738E-07	1.73777E-07	2.46925E-08

Tableau 11 : **Matrice Covariance – Erreur d'estimation**

		2	3	4	5	6
		370783353	379874128	366552127	358908118	391968305
2	370783353	1.02096E+13	1.04599E+13	1.00931E+13	9.88259E+12	1.07929E+13
3	379874128	1.04599E+13	53358466649	51487211067	50413506566	55057257593
4	366552127	1.00931E+13	51487211067	24955883672	24435458423	26686287474
5	358908118	9.88259E+12	50413506566	24435458423	22385074663	24447036239
6	391968305	1.07929E+13	55057257593	26686287474	24447036239	3793732957

Tableau 12 : Matrice des coefficients corrélation d'erreur d'estimation

		2	3	4	5	6
		3195243	230995	157974	149616	61593
2	3195243	1	14.17169288	19.99553153	20.67224742	54.84048138
3	230995	14.17169288	1	1.410948691	1.45869993	3.869719859
4	157974	19.99553153	1.410948691	1	1.033843356	2.742636839
5	149616	20.67224742	1.45869993	1.033843356	1	2.652855312
6	61593	54.84048138	3.869719859	2.742636839	2.652855312	1

Tableau 13 : Les résultats du modèle MACK

Diagonale	LDF	Reserve	Ultime	Proc Mult	SD Proc	CV	SD Param	CV	SD Totale	CV
329 667	1.000		329 667							
370 424	1.001	359	370 783	17	80	22%	3 195	890%	3 196	**890%**
379 264	1.002	610	379 874	178	260	43%	231	38%	348	**57%**
365 250	1.004	1 302	366 552	196	268	21%	158	12%	311	**24%**
357 096	1.005	1 812	358 908	528	435	24%	150	8%	460	**25%**
336 247	1.166	55 722	391 968	162925	7 991	14%	62	0%	7 992	**14%**
2 137 949		59 804			8 012	13%	9 656	16%	12 547	**21%**

Tableau 14 : Niveau de confiance

	Bornes Inf	**Bornes Sup**
Niveau de confiance 90%	43 718	75 890
Niveau de confiance 95%	39 164	80 444
Niveau de confiance 97.5%	35 211	84 397

ANNEXES

A. Les code de logiciel sur R

Poisson sur- dispersion

```
# Load up data
data <- read.csv("Mack.csv", header=F)
data <- as.matrix(data)
data
# Prepare data in correct form
claims <- as.vector(data)
n.origin <- nrow(data)
n.dev <- ncol(data)
origin <- factor(row <- rep(1:n.origin, n.dev))
dev <- factor(col <- rep(1:n.dev, each=n.origin))
# Put into a data frame (no need, but easier to visualise)
mack <- data.frame(claims=claims, origin=origin, dev=dev)
mack[1:5, ] # Print first five rows
# New quasi-poisson family
quasipoisson <- function (link = "log")
## to cope with negative y values##
## Computes Pearson X^2 rather than Poisson deviance##
## Starting values are all equal to the global mean
{
linktemp <- substitute(link)
if (!is.character(linktemp)) {
linktemp <- deparse(linktemp)
if (linktemp == "link")
linktemp <- eval(link)
}
if (any(linktemp == c("log", "identity", "sqrt")))
stats <- make.link(linktemp)
else stop(paste(linktemp, "link not available for poisson",
"family; available links are", "\"identity\", \"log\" and
\"sqrt\""))
variance <- function(mu) mu
validmu <- function(mu) all(mu > 0)
dev.resids <- function(y, mu, wt) wt*(y-mu)^2/mu ###
aic <- function(y, n, mu, wt, dev) NA
initialize <- expression({
n <- rep(1, nobs)
mustart <- rep(mean(y), length(y)) ###
})
structure(list(family = "quasipoisson", link = linktemp,
linkfun = stats$linkfun, linkinv = stats$linkinv, variance =
variance,
dev.resids = dev.resids, aic = aic, mu.eta = stats$mu.eta,
initialize = initialize, validmu = validmu, valideta =
stats$valideta),
class = "family")
}
```

```
# Fit model
model <- glm(claims ~ origin + dev, family = quasipoisson(),
subset=!is.na(claims), data=mack)
summary(model)
# Extract useful info from the model
coef <- model$coefficients                          # Get coefficients
disp <- summary(model)$dispersion                   # Get dispersion parameter
cov.param <- disp * summary(model)$cov.unscaled   # Get covariance matrix of parameters
# To determine future uncertainty, need to create a
# design matrix for future payments. Build up in stages.
# Assume start from bottom left of future triangle.
n.fut.points <- length(claims[is.na(claims)])
fut.design <- matrix(0, nrow = n.fut.points, ncol=length(coef))
fut.points <- claims
fut.points[!is.na(claims)] <- 0
fut.points[is.na(claims)] <- 1:n.fut.points
for(p in 1:n.fut.points){
# All points and a constant in the predictor
fut.design[p, 1] <- 1
# Row factor
fut.design[p, 1 +
as.numeric(origin[match(p, fut.points)]) - 1] <- 1
# Col factor
fut.design[p, 1 + (n.origin-1) +
as.numeric(dev[match(p, fut.points)]) - 1] <- 1}
# Determine fitted future values (as a diagonal matrix)
fitted.values <- diag(as.vector(exp(fut.design %*% coef)))
total.reserve <- sum(fitted.values)
total.reserve
# Determine covariance matrix of linear predictors
cov.pred <- fut.design %*% cov.param %*% t(fut.design)
# Determine covariance matrix of fitted values
cov.fitted <- fitted.values %*% cov.pred %*% fitted.values
# Determine uncertainty statistics
total.rmse <- sqrt(disp*total.reserve+sum(cov.fitted))
total.predictionerror <- round(100*total.rmse/total.reserve)
#Get results
total.rmse
[1] 17612.73
total.predictionerror
[1] 34
```

Negative Binomial model

```
# options (contrasts=c("contr.treatment","contr.treatment"))
library (MASS)
attach("C:/R/hungpham/_data", pos=2)
attach("C:/R/hungpham/_data", pos=3)

# Data preparation
Ex2data.cumulative_Cumulate(Ex2data)
Ex2data.cumulative[is.na(Ex2data)]_0
```

```
data <- Ex2data.cumulative
nc <- ncol(data)
nr <- nrow(data)
nu <- nr * (nc - 1)
d <- rep(as.matrix(data)[cbind(2:nr, (nr - 1):1)], 1:(nr - 1))
data1 <- data[, 2:nr] - data[, 1:(nr - 1)]
data1[cbind(2:nr, (nr - 1):1)] <- 0
data1 <- as.vector(data1)
w2 <- ifelse(data1 == 0, 1, as.vector(data[, 1:(nr - 1)]))
w3 <- ifelse(data1 == 0, 0, 1)
vcol <- 1 + (ceiling(1:nu/nr) - 1) %% (nc - 1)
Col <- ordered(as.factor(vcol))
# Model fitting
res <- glm(data1 ~ Col + offset(log(w2)), family = myneg.bin(w2), weights = w3, epsilon = 1e-005,
maxit = 15)
# Get results
fv <- res$fitted.values/w2
a_matrix(round(data1,3), nc=9)
matrix(round(fv,4),nc=9)              # link ratios - 1
b_summary(res)$coefficients           # note use of log link function
fv2 <- res$fitted.values
matrix(round(fv,3),nc=9)              # link ratios - 1
b
summary(res)$cov.unscaled
summary(res)$dispersion
summary(res)$deviance
summary(res)$df
```

 # Normal Model

```
# Data preparation
data <- Ex2data.cumulative
nc <- ncol(data)
nr <- nrow(data)
nu <- nr * (nc - 1)
d <- as.vector(as.matrix(data)[cbind(2:nr, (nr - 1):1)])
data1 <- data[, 2:nr]
data1[cbind(2:nr, (nr - 1):1)] <- 0
data1 <- as.vector(data1)
w2 <<- ifelse(data1 == 0, 1, as.vector(data[, 1:(nr - 1)]))
w3 <<- ifelse(data1 == 0, 0, 1)
wvar <- w3
wvar[nu - nc + 1] <- 0
vcol <<- 1 + (ceiling(1:nu/nr) - 1) %% (nc - 1)
vcolvar <<- ifelse(vcol == nc - 1, nc - 2, vcol)
Col <<- ordered(as.factor(vcol))
Colvar <<- ordered(as.factor(vcolvar))
phi <- rep(1, length(data1))
data2 <- data1/w2
w4 <- (w3 * w2)/phi
# Model fitting using joint modelling
res1 <- glm(data2 ~ Col, weights = w4, epsilon = 1e-005)
resids <- (w2 * ((data2 - res1$fitted.values))^2)*44/36
resids <- (w2 * ((data2 - res1$fitted.values))^2)
res2 <- glm(resids ~ Colvar, family = Gamma, weights = wvar)
phi <- predict(res2, type = "response")
w4 <- (w3 * w2)/phi
```

```
res3 <- glm(data2 ~ Col-1, weights = w4, epsilon = 1e-005)
fv <- predict(res3, type="response", se.fit=T)
# Results
matrix(round(data2,3), nc=9)
a_matrix(round(phi,1),nc=9)                    # Mack's alphas
matrix(round(fv$fit,4),nc=9)                   # link ratios
matrix(round(fv$se.fit,4),nc=9)                        # standard errors of link ratios
summary(res3)$dispersion
summary(res3)$deviance
summary(res3)$coefficients                     # note use of identity link function, and no constant
summary(res3)$coefficients
```

 # Log Normal Model

```
# Data preparation
# Load up data
data <- read.csv("Mack.csv", header=F)
data <- as.matrix(data)
data
# Prepare data in correct form
claims <- as.vector(data)
logclaims<-log(claims)
n.origin <- nrow(data)
n.dev <- ncol(data)
origin <- factor(row <- rep(1:n.origin, n.dev))
dev <- factor(col <- rep(1:n.dev, each=n.origin))
# Put into a data frame (no need, but easier to visualise)
mack <- data.frame(claims=claims, origin=origin, dev=dev)
mack[1:5, ] # Print first five rows
# Fit model
model <- glm(logclaims ~ origin + dev,subset=!is.na(claims), data=mack)
summary(model)
# Extract useful info from the model
coef <- model$coefficients                     # Get coefficients
disp <- summary(model)$dispersion              # Get dispersion parameter
cov.param <- disp * summary(model)$cov.unscaled  # Get covariance matrix of parameters
# To determine future uncertainty, need to create a
# design matrix for future payments. Build up in stages.
# Assume start from bottom left of future triangle.
n.fut.points <- length(logclaims[is.na(logclaims)])
fut.design <- matrix(0, nrow = n.fut.points, ncol=length(coef))
fut.points <- logclaims
fut.points[!is.na(logclaims)] <- 0
fut.points[is.na(logclaims)] <- 1:n.fut.points
for(p in 1:n.fut.points){
# All points and a constant in the predictor
fut.design[p, 1] <- 1
# Row factor
fut.design[p, 1 +
as.numeric(origin[match(p, fut.points)]) - 1] <- 1
# Col factor
fut.design[p, 1 + (n.origin-1) +
as.numeric(dev[match(p, fut.points)]) - 1] <- 1}
# Determine fitted future values (as a diagonal matrix)
fitted.values <- diag(as.vector(exp(fut.design %*% coef)))
total.reserve <- sum(fitted.values)
```

```
total.reserve
# Determine covariance matrix of linear predictors
cov.pred <- fut.design %*% cov.param %*% t(fut.design)
# Determine covariance matrix of fitted values
cov.fitted <- fitted.values %*% cov.pred %*% fitted.values
# Determine uncertainty statistics
total.rmse <- sqrt(disp*total.reserve+sum(cov.fitted))
total.predictionerror <- round(100*total.rmse/total.reserve)
#Get results
total.rmse
total.predictionerror
```

B.Modèle Linéaire Généralisé

§1. Loi de probabilité de la variable de réponse

Dans les cas des GLM, la variable de réponse suit une loi appartenant à la famille exponentielle. La fonction de densité de la variable de réponse Y peut être exprimé ainsi :

$$f(y) = \exp\left\{\frac{y\theta - b(\theta)}{a(\theta)} + c(y,\phi)\right\}$$

où *a*, *b* et *c* sont des fonctions qui déterminent la distribution spécifique. Si ϕ est fixé, c'est une distribution de la famille exponentielle à un paramètre. Les fonction a et c sont telles que $a(\phi) = \phi/w$ et $c = c(y, \phi/w)$, où *w* est le poids de chaque observation.

Selon la théorie de ce type de distribution, nous pouvons exprimer la moyenne et la variance comme suit :

$E(Y) = b'(\theta)$
$Var(Y) = b''(\theta)\phi/w$

Si μ est la moyenne de Y, nous pouvons exprimer la variance en fonction de la moyenne :

$Var(Y) = V(\mu)\phi/w$, où V est la fonction de variance.

Les fonctions de densité de la variable de réponse Y dans le cas du GLM sont paramétrées en fonction de la moyenne μ et du paramètre de dispersion ϕ au lieu du paramètre θ. Nous énumérons les loi de probabilités qui sont inclues dans la procédure GENMOD :

<u>Loi Normale</u>

$$f(y) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right), \text{ pour } -\infty < y < \infty$$

$\phi = \sigma^2$
$Var(Y) = \sigma^2$

<u>Loi Inverse Gaussienne</u> :

$$f(y)=\frac{1}{\sqrt{2\pi y^3\sigma}}\exp\left(-\frac{(y-\mu)^2}{2(\mu\sigma)^2 y}\right), \text{ pour } 0<y<\infty$$

$$\phi=\sigma^2$$

$$Var(Y)=\sigma^2\mu^3$$

Loi Gamma :

$$f(y)=\frac{1}{\Gamma(\nu)y}\left(\frac{y\nu}{\mu}\right)^{\nu}\exp\left(-\frac{y\nu}{\mu}\right), \text{ pour } 0<y<\infty$$

$$\phi=\nu^{-1}$$

$$Var(Y)=\mu^2/\nu$$

Loi de Poisson

$$f(y)=\frac{\mu^y e^{-\mu}}{y!}, \text{ pour } y=0,1,2...$$

$$Var(Y)=\mu$$

Loi Binomiale

$$f(y)=\binom{n}{r}\mu^r(1-\mu)^{n-r}, \text{ pour } y=r/n,\ r=0,1,2...,n$$

$$Var(Y)=\mu(1-\mu)/n$$

Loi Binomiale Négative

$$f(y)=\frac{\Gamma(y+1/k)}{\Gamma(y+1)\Gamma(1/k)}\frac{(k\mu)^k}{(1+k\mu)^{y+1/k}} \text{ pour } y=0,1,2...$$

$$Var(Y)=\mu(1-\mu)/n$$

§2. Fonction de lien

La moyenne μ_i de la variable de réponse de la $i^{ème}$ observation est liée avec le prédicteur linéaire à l'aide d'un fonction monotone et différentiable g :

$g(\mu)=x_i'\beta$, où x_i est un vecteur des variables explicatives connues et β est un vecteur de paramètres inconnus.

§3. Fonction de Log Vraisemblance

La fonction de log vraisemblance qui sont disponibles dans la procédure GENMOD sont paramétrées à l'aide de la moyenne μ_i et du paramètre de dispersion ϕ. y_i représente la réponse de la $i^{ème}$ observation. La fonction de log vraisemblance est :

$$L(y,\mu,\phi)=\sum_i \log(f(y_i,\mu_i,\phi))$$

§4. Estimation des paramètres

La procédure GENMOD utilise l'algorithme de Newton-Raphson pour maximiser la fonction de log vraisemblance $L(y,\mu,\phi)$ sous les paramètres β .

Le vecteur de paramètres β est calculé par itération. La r[ième] itérations se présente ainsi :

$$\beta_{r+1} = \beta_r - \mathrm{H}^{-1}\mathrm{s}$$

où H est le Hessien (la matrice des dérivées secondes) et s est le gradian (le vecteur des dérivées premières) de la fonction de log vraisemblance L, les deux évaluées avec la valeur courante du paramètre :

$$s = [s_j] = \left[\frac{\partial L}{\partial \beta_j}\right] \text{ et } H = [h_{ij}] = \left[\frac{\partial^2 L}{\partial \beta_i \partial \beta_j}\right]$$

§5. Qualité du modèle

Ils existent deux statistiques qui sont utiles pour mesurer la qualité du modèle. L'un c'est la déviance et l'autres c'est la statistique Khi deux de Pearson. Si *l*(y,μ*)* est la fonction de log vraisemblance exprimée en terme des valeurs estimées μ et de vecteur y des variables de réponse, alors la déviance est égale à :

$$D(\mathrm{y},\mu) = 2(l(\mathrm{y},\mathrm{y}) - l(\mathrm{y},\mu)) * \phi$$

La déviance mesure la différence entre le modèle parfait (celui qui inclut un paramètre par observation) et le modèle estimé.

La statistique de Khi deux de Pearson est définie comme suit :

$$\chi^2 = \sum_i w_i (y_i - \mu_i)^2 / V(\mu_i)$$

§6. Estimation de la moyenne

Si $\hat{\eta}_i = x_i' \hat{\beta}$ est l'estimation linéaire de l'observation i, *g* est la fonction de lien, alors :

$$\hat{\mu}_i = g^{-1}(\hat{\eta}_i) = g^{-1}\left(x_i' \hat{\beta}\right)$$

Donc nous pouvons exprimer la moyenne de l'observation i à l'aide du vecteur des variables explicatives x_i qui sont propre à cette observation et à l'aide des paramètres estimés β

C. Typologie des branches (travaux de la Commission Dommages (FFA 1998)

Cat. 20-21 : Dommages corporels (individuels, collectifs)

Provisionnement des rentes (Invalidité, Incapacité, Accidents du travail) réglementé

Frais de soin : déroulement plus court en individuel qu'en collectif (≈ 3 ans)

Individuel accidents : déroulement plus long.

Cat. 22-23 : Automobile : RC, dommages
Séparer les flottes des contrats de particuliers
Dommages (segmenté par garantie)
RC matériels : déroulement court (≈ 5 ans)
RC corporels : déroulement long (10 à 15 ans)
Isoler les rentes. Sensible à l'évolution de la jurisprudence (montant des indemnisations)

Cat. 24-26 : Dommages aux biens
Séparer
- Dégâts des eaux, vol, incendie, dégâts électriques, tempêtes, autres événements naturels : liquidation rapide (≈ 5 ans)
- Grands sinistres incendies
RC des multirisques, en particulier les corporels : liquidation lente (≥ 10 ans)

Cat; 27 : Catastrophes naturelles
A rapprocher des dommages aux biens.
La proportion de tardifs peut être importante (cf. sinistres sécheresse).

Cat. 28 : RC Générale
Branches à déroulement long ou très long, de 10 à 30 ans (RC médicale ou hospitalière).
Forte évolution de la jurisprudence.
Distinguer
- RC produit, RC exploitation, RC pollution
- RC professionnelle { *RC médicale et hospitalière ; *Services (avocats, conseil,…) }

sinistres matériels : liquidation rapide
sinistres corporels : liquidation très lente (≥ 20 ans)

Cat. 29 : Protection juridique
Déroulement de 3 à 7 ans
Distinguer contrats particuliers, entreprises.

Cat. 34 : Transports
Mélange de garanties à déroulement court et déroulement long.
Isoler les grands sinistres et la RC.
- Corps (maritimes)

Corps de pêche, corps fluviaux : déroulement court
Autres corps : déroulement plus long
- Facultés (terrestres et maritimes) : court
- Aviation (corps et facultés) : problèmes judiciaires
- Spatial : cas par cas

Cat. 35-36 : Assurance construction (dommages ouvrage, RC décennale)
Catégories très spécifiques : gestion en capitalisation et longueur des déroulements (≈ 20 ans)
Scinder RC décennale { artisans ; entreprises ; concepteurs (architectes, bureaux d'études,…) }
Dommages ouvrage { particuliers ; grands chantiers }

A noter : de très nombreux recours en dommages ouvrage

D.BIBLIOGRAPHIE

PARTRAT Christian, BESSON Jean-Luc , *Assurance non-vie, Modélisation, Simulation,* Economica

ENGLAND P.D., VERRALL R.J. (2002), *Stochastic Claims Reserving in General Insurance*, British Actuarial Journal.

Faculty of Actuaries: Institute of Actuaries. (1989). *Claims Reserving Manual, Vol. 1,* The Faculty and Institute of Actuaries.

Faculty of Actuaries: Institute of Actuaries. (1997). *Claims Reserving Manual, Vol. 2*, The Faculty and Institute of Actuaries.

DENUIT Michel, CHARPENTIER Arthur, *Mathématiques de l'assurance non-vie, Tome 2,* Economica

MACK, T. (1994a). *Which Stochastic Model is underlying the Chain-Ladder Model? Vol.15, pages 133-138*, Insurance: Mathematics and Economic.

RENSHAW A.E. (1994a, b). *Claims Reserving by Joint Modelling,* Actuarial Research Paper No. 72, Department of Actuarial Science and Statistics, City University, London.