



UNIVERSITÉ PARIS DAUPHINE

Département MIDO^(*)

MASTER MIDO

MENTION MMD^(**)

SPÉCIALITÉ ACTUARIAT

Année Universitaire : 2007-2008

Mémoire d'Actuariat présenté en **novembre 2008** devant l'Université Paris Dauphine
et l'Institut des Actuaire

Par : ***Romain BOYER CHAMMARD***

Tuteur : ***François FUGERE***
Frédéric EICHEL

Sujet : **Processus de surveillance et de majoration des contrats flottes d'entreprise d'AXA France**

Entreprise d'accueil : ***AXA France***

NON CONFIDENTIEL

JURY

Membres du Jury

Christian HESS

Fonctions / Entreprise

Président du jury, Professeur à l'Université Paris-Dauphine

(*) MIDO : Mathématiques, Informatique, Décision, Organisation

(**) MMD : Mathématiques, Modélisation, Décision

Résumé

Le but de ce mémoire est de développer un processus de surveillance des contrats flottes automobiles d'AXA France. On cherche à déterminer quels contrats sont sous-tarifés et quel niveau de majoration permettrait de les ramener à l'équilibre. L'outil utilisé est le ratio S/C : le rapport entre les sinistres d'un contrat et ses primes. Nous chercherons donc à optimiser le calcul de cet estimateur et à tester la validité des résultats obtenus sur le portefeuille.

La première partie présente brièvement le marché des flottes d'entreprise, les différents types de contrats et de garanties. On donne aussi les classifications les plus utilisées.

Dans un second temps, nous avons modélisé la sinistralité des flottes. Nous étudions la fréquence des sinistres en fonction de la taille de la flotte. Ensuite, nous modélisons leurs coûts. Nous distinguons les sinistres ordinaires (traités par des méthodes d'adéquations classiques) des graves (traités par des méthodes de type valeurs extrêmes). Pour le calcul des lois de coûts, nous avons segmenté le portefeuille suivant les types de véhicules composant les flottes et suivant leur usage. Enfin, nous présentons un utilitaire Monte Carlo permettant de simuler la loi de la charge totale d'une flotte (en fonction de sa taille et du type de véhicule).

Dans la troisième partie, nous décrivons le processus de calcul des majorations. Le problème essentiel est que les S/C sont extrêmement volatiles et donc peu fiables. Deux procédés permettent de gérer cette difficulté. Tout d'abord, nous écrêtons les sinistres : nous leur fixons une valeur maximale, le montant en surplus étant mutualisé. Reste à choisir cette valeur limite (qui sera prise comme un quantile de la loi de coût pour les flottes de même taille) et comment mutualiser (la mutualisation se fait proportionnellement à la prime, une partie de l'excès est mutualisé sur tous le portefeuille, le reste uniquement sur les flottes similaires à celle ayant eu le sinistre). Un deuxième outil permet de rendre plus stable le S/C : la crédibilité. Le S/C d'un contrat est obtenu comme moyenne de son ratio propre et de celui de l'ensemble des flottes de même type. Nous présentons plusieurs modèles. Celui finalement retenu est le modèle de Jewell à trois niveaux. Au final, le S/C est obtenu comme la moyenne des S/C écrêtés et mutualisés des 3 dernières années puis crédibilisé. Des ajustements sont effectués pour tenir compte de la plus grande stabilité de certains contrats, et des tendances (contrats en amélioration ou en dégradation). Il est ensuite comparé à un niveau d'équilibre. Les contrats au dessus de ce niveau sont majorés de façon à être ramené à l'équilibre. Pour tenir compte du poids commercial des gros contrats, le seuil d'équilibre est choisi croissant avec la taille de la flotte.

La dernière partie est consacrée aux tests du modèle et à l'étude des résultats obtenus. Nous vérifions que les niveaux de majoration que nous avons retenus captent bien un accroissement de la sinistralité. Nous effectuons ensuite une étude pour voir le lien entre ces majorations et divers segmentations (par des tris à plat et une ACM). Enfin, nous avons testé la stabilité dans le temps des résultats.

Mots clés :

flottes d'entreprises, ratio S/C , majorations QP, théorie des valeurs extrêmes, modélisation de la sinistralité, écrêtement, crédibilité

Abstract

In this paper, we describe a monitoring process for the portfolio of fleets of car contracts of the AXA company. We want to determine which contracts have insufficient premiums and how to make them acceptable. The main tool will be the loss ratio : the total amount of claims for a contract divided by its premium. We seek to optimize the calculation of this estimator and to test its use on our portfolio.

Our first paragraph deals with the fleet of automobile market, the different types of contracts and guarantees. We also present the most widely used classifications.

In a second time, we will model the total claim of our contracts. We first study the frequency of claims for different size of fleets. Then we model the amounts. We will distinguish ordinary claims (treated by classical adequation methods) and large claims (treated through the extreme value theory). We obtain laws of severity for the different kind of vehicle and of use. Finally, we use a Monte Carlo method to simulate the total claim amount (for the different size of fleets and type of vehicles).

The third paragraph describe the monitoring process. Our main problem is the high volatility of the S/C, which makes it unreliable. We use two tools to deal with this. First, we have trimmed our data : the claims have received a maximum value. The amount beyond that value is mutualised. We have to choose that threshold (it will be a quantile of the fleet total claim law) and the way to deal with the excess (it will be mutualised proportionately to the premium, for part between the whole portfolio and for part between the fleets similar to the one responsible for the claim). Another tool makes the S/C more stable : credibility. The S/C of a contract is defined as a weighted average of its own ratio and that of the equivalent fleets. We tried several models. We finally chose a Jewell model with three levels. Finally, the S/C is the average of the trimmed-mutualised S/C of the last 4 years and credibilized. Some modifications are made to take into account the fact that some contracts are more stable than others and that they can exhibit tendencies (improving or getting worse). It is then compared to the equilibrium level. For the contracts above it, we increase the premium so as to bring them back to balance. Bigger contracts having more negotiation capacity, we use a threshold increasing with the fleet's size.

The last part studies and tests the results we have obtained. We get sure that the increase of the premium corresponds to an increase in claims. Then we study the links between these increases and several variables (through frequencies and an MCA). Lastly, we study the stability of the model over time.

Keywords :

fleets of vehicles, loss ratio, experience rating, extreme value theory, claim and frequency modelisation, data trimming, credibility

Avertissement

Pour des raisons de confidentialité, les valeurs numériques présentes dans ce mémoire ont été modifiées. Les ordres de grandeur ont néanmoins été conservés.

Remerciements

Je tiens à remercier mon responsable de stage, François Fugere pour son aide, ses conseils et pour m'avoir fait profiter de son expérience. Je remercie également toute l'équipe auto entreprise pour leur accueil. Enfin, pour l'aide qu'ils m'ont apporté pour la rédaction de ce mémoire, je remercie Philippe Marie-Jeanne, Jérôme Vignancour, Frédéric Eichel et Solange Hamel.

Table des matières

| | |
|--|-----------|
| Introduction | 9 |
| 1 L'assurance des flottes automobiles | 11 |
| 1.1 Le marché des flottes automobiles | 12 |
| 1.2 Les différents types de contrats automobile entreprise | 14 |
| 1.3 Les différentes garanties | 17 |
| 2 Modélisation de la sinistralité | 20 |
| 2.1 Modélisation de la fréquence des sinistres | 22 |
| 2.1.1 Répartition des contrats par classe de cotisations | 23 |
| 2.1.1.1 Méthode de segmentation | 23 |
| 2.1.1.2 Subdivision du portefeuille | 25 |
| 2.1.2 Méthodes d'ajustements des lois de fréquences : | 29 |
| 2.1.3 Lois des fréquences | 31 |
| 2.1.3.1 Fréquences des sinistres classiques | 31 |
| 2.1.3.2 Cas des sinistres IDA forfaitaires | 33 |
| 2.2 Modélisation de la sévérité des sinistres | 36 |
| 2.2.1 Distinction entre sinistres graves et ordinaires | 37 |
| 2.2.1.1 Les coûts suivent-ils une loi extrême ? | 37 |
| 2.2.1.2 Choix du seuil pour les sinistres graves | 39 |
| 2.2.2 Modélisation des sinistres ordinaires | 41 |
| 2.2.2.1 Lois étudiées et tests d'adéquation | 41 |
| 2.2.2.2 Loi des sinistres ordinaires | 43 |
| 2.2.3 Modélisation des sinistres graves | 45 |
| 2.2.3.1 Rappel de théorie des valeurs extrêmes | 45 |
| 2.2.3.2 Loi de coût des sinistres extrêmes | 47 |
| 2.2.4 Prise en compte de l'usage | 51 |
| 2.2.4.1 Importance de l'usage | 51 |
| 2.2.4.2 Loi des coûts par usage | 53 |
| 2.2.4.3 Comparaison des modèles | 59 |
| 2.2.4.4 Modèle à double critère | 61 |
| 2.2.5 Résumé | 63 |
| 2.3 Simulation de la charge totale | 64 |
| 2.3.1 Obtention de la loi de la charge | 65 |
| 2.3.2 Estimation des moments | 67 |
| Synthèse de la deuxième partie | 70 |
| 3 Processus de majoration QP | 72 |
| 3.1 Ecrêtement des sinistres | 74 |
| 3.1.1 Détermination du seuil d'écèlement | 75 |
| 3.1.2 Utilisation des coefficients de variation : | 77 |
| 3.1.3 Mutualisation de la surcrête | 79 |

| | | |
|----------|--|------------|
| 3.1.4 | Deuxième niveau d'écrêtement | 81 |
| 3.1.5 | Résumé et exemples | 85 |
| 3.2 | Crédibilité | 86 |
| 3.2.1 | Description des différents modèles | 89 |
| 3.2.1.1 | Modèle des fluctuations limités | 89 |
| 3.2.1.2 | Modèle de Buhlmann Straub | 90 |
| 3.2.1.3 | Modèle de Jewell | 92 |
| 3.2.1.4 | Modèle de Buhlmann à trois estimateurs | 96 |
| 3.2.2 | Choix des poids | 98 |
| 3.2.3 | Application des modèles de crédibilité au portefeuille | 101 |
| 3.2.3.1 | Application de la crédibilité américaine | 101 |
| 3.2.3.2 | Utilisation du modèle de Buhlmann Straub | 103 |
| 3.2.3.3 | Modèle de Jewell à un niveau intermédiaire | 105 |
| 3.2.3.4 | Modèle de Jewell à deux niveaux intermédiaires | 107 |
| 3.2.3.5 | Modèle à trois estimateurs | 109 |
| 3.2.4 | Choix du modèle | 111 |
| 3.3 | Description du processus de majoration QP | 112 |
| 3.3.1 | Calcul du S/C par années | 113 |
| 3.3.2 | Calcul du S/C historique et du S/C crédibilisé | 115 |
| 3.3.3 | Un exemple | 117 |
| 3.3.4 | Choix du seuil d'équilibre | 119 |
| 3.3.5 | Pilotage de l'algorithme | 121 |
| 3.3.6 | Niveau de majoration | 123 |
| | Synthèse de la troisième partie | 126 |
| 4 | Etude des résultats fournis par le processus QP | 128 |
| 4.1 | Détection de la sinistralité | 129 |
| 4.2 | Etudes statistiques des contrats retenus | 132 |
| 4.2.1 | Variables de l'étude | 133 |
| 4.2.2 | Liens entre les variables et les majorations | 135 |
| 4.2.3 | Analyse des correspondances multiples | 137 |
| 4.2.3.1 | Rappel sur l'ACM | 137 |
| 4.2.3.2 | Application aux données | 139 |
| 4.3 | Comparaison avec le S/C dossier-dossier | 146 |
| 4.4 | Etude de la stabilité dans le temps | 152 |
| 4.4.1 | Evolution des résultats au cours du temps | 153 |
| 4.4.2 | Variation du S/C d'un contrat au cours du temps | 155 |
| 4.4.3 | Bilan | 157 |
| | Synthèse du quatrième chapitre | 158 |
| | Conclusion | 159 |
| | Pistes de développement | 161 |
| | Bibliographie | 164 |
| | Liste des abréviations | 165 |

Introduction

L'une des problématiques essentielles de l'assurance (et des actuaires) est de déterminer la prime juste pour chaque contrat. Il convient de distinguer deux situations différentes : pour les nouveaux contrats, nous ne disposons pas de données précises sur la sinistralité du contrat (il n'y a pas de bonus-malus pour les flottes d'entreprise). La prime est donc obtenue à partir de divers segmentations (nombre et types de véhicules, zone géographique...). Par la suite, nous pouvons observer les performances de ces contrats et éventuellement ajuster leurs cotisations. En effet, il arrive que des contrats se révèlent moins bons que ce que nous attendions ou bien se détériorent. L'objet de ce mémoire est d'optimiser ce processus de surveillance et de calcul de majorations.

Le procédé le plus classique consiste à déterminer le ratio S/C (sinistres à cotisations). Une valeur trop élevée de cet indice nous permet de conclure que le contrat n'est pas rentable. Néanmoins, il faut se méfier d'une approche trop simpliste. En effet, le S/C possède une variance très élevée et n'est guère utilisable en l'état. Il faut donc recourir à divers techniques pour le rendre plus stable et plus fiable. On aura recours à des méthodes d'écrêtement et de crédibilité. Ces techniques s'appliquent à des portefeuilles homogènes, ce qui n'est clairement pas le cas du notre. Il faudra donc le segmenter de façon pertinente.

Dans un premier temps, on présente brièvement le marché des flottes d'entreprises, les différents contrats et les garanties offertes. On décrit aussi les différents types de flottes que nous avons en portefeuille.

Dans la suite du mémoire, nous aurons souvent besoin d'une loi pour la charge totale du contrat. L'objet de la deuxième partie est de déterminer cette loi en modélisant la sinistralité. On commence par définir une loi pour la fréquence des sinistres. Celle-ci dépend évidemment du nombre de véhicules de la flotte. On segmentera donc le portefeuille en classe de taille. Ensuite, on s'intéresse au coût des sinistres. Le poids des sinistres graves en automobile est très élevé. On les modélisera donc à part. Les sinistres ordinaires seront traités par des outils d'adéquations classiques. L'étude des sinistres graves (déterminer à partir de quel seuil un sinistre n'est plus ordinaire et modéliser sa loi) use fera à l'aide de la théorie des valeurs extrêmes et plus spécialement la méthode "Peak over threshold". On testera deux modèles, selon que l'on segmente le portefeuille par l'usage ou par le type de véhicules. Enfin, on utilisera une méthode de type Monte Carlo pour simuler la charge totale d'un contrat.

On décrit ensuite le processus de surveillance (ou processus QP). On insiste sur les deux outils visant à rendre plus stable le résultat. Dans le cas de l'écrêtement, on a présenté plusieurs choix de seuils et de méthodes de mutualisation. Pour la crédibilité, on a testé cinq modèles. On décrit ensuite en détail le calcul du S/C puis de l'éventuelle majoration. On présente les différentes difficultés techniques comme le choix du niveau d'équilibre ou la prise en compte de l'évolution de certains contrats dans le temps.

La dernière partie vise à étudier les résultats obtenus. On vérifie en particulier qu'il n'y ait pas de biais dans l'algorithme. On s'assure que des niveaux de majorations plus élevés correspondent à des sinistralités plus problématiques. Ensuite, on étudie la proportion de contrats majorés selon divers variables, pour voir les facteurs pouvant rendre un contrat problématique. Enfin, on évalue la stabilité des résultats de l'algorithme dans le temps.

Chapitre 1

L'assurance des flottes automobiles

L'assurance automobile d'entreprise couvre tous les véhicules terrestres à moteur utilisés dans le cadre professionnel. C'est un domaine extrêmement varié, où les contrats sont bien plus hétérogènes qu'en assurance des véhicules particuliers. Ces véhicules assurés sont de type divers, comprenant des voitures (véhicules de VRP, de dirigeants..), des deux roues, des camionnettes mais aussi des véhicules lourds voire très lourds (camions, engins de chantier, moissonneuses...).

Les contrats varient aussi par la taille, du petit commerçant assurant sa voiture à la grande entreprise de transport et ses cents poids lourds. L'usage des véhicules est également très varié : transport de l'assuré, de marchandises, de passagers (avec risques de sinistres corporels en responsabilité civile). Selon les cas, les conducteurs peuvent être connus ou non (entreprise de location de voiture, auto école...). Le domaine couvre aussi les véhicules particuliers s'ils sont utilisés ponctuellement dans un cadre professionnel ("contrat mission").

Dans cette partie, on donnera un bref aperçu du marché de l'assurance automobile d'entreprise en général et des flottes automobiles en particuliers. On détaillera les différentes classifications des contrats, qui seront utilisés dans la suite de l'étude. Enfin, on rappellera les divers types de garanties existantes (celles ci sont assez similaires à l'automobile particulier).

1.1 Le marché des flottes automobiles

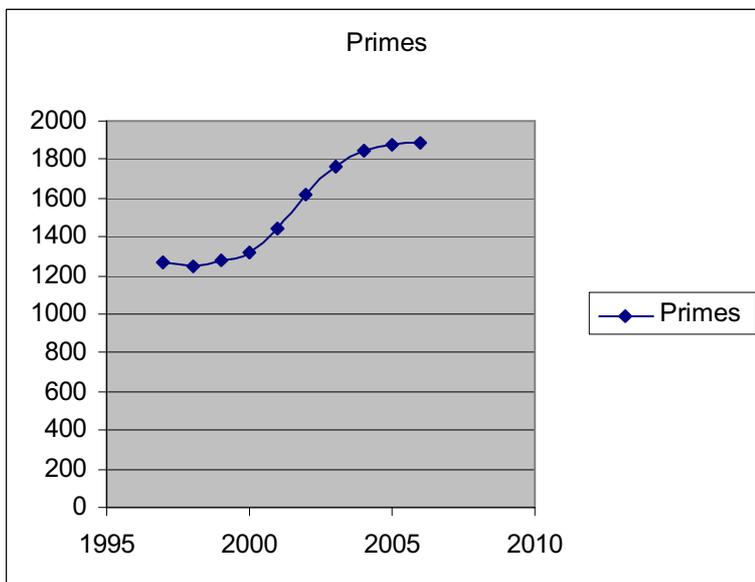
L'assurance automobile est la première branche de l'assurance non vie, avec près de 18 milliards d'euros de cotisation soit environ 45% du total des cotisations non vie. On distingue deux domaines : l'assurance automobile de particuliers (80% du total) et l'assurance automobile des professionnels (20%).

Parmi les contrats professionnels, il faut distinguer les contrats individuels des flottes. Les contrats individuels sont des contrats visant à assurer un unique véhicule, utilisé dans le cadre professionnel. Typiquement, il peut s'agir de la camionnette d'un artisan ou d'un vendeur ambulancier, ou d'un taxi indépendant. Ces contrats sont gérés par les mêmes méthodes que les véhicules automobiles particuliers. L'autre type de contrat automobile professionnel est la flotte. Dans ce cas, on assure plusieurs véhicules appartenant à une même entreprise. Le nombre peut être très variable, souvent inférieur à 10 mais pouvant dépasser 1000 (flotte de collaborateurs de très grosses entreprises, grosses entreprises de transports, flottes de collectivités locales..). Ces contrats sont traités de façon assez différente des contrats automobiles classiques. En particulier, comme il n'y a en général pas de conducteurs attitrés, on ne peut pas utiliser de système de type bonus-malus. Par contre, on peut utiliser les performances historiques des flottes de façon plus fiable, compte tenu de leur taille.

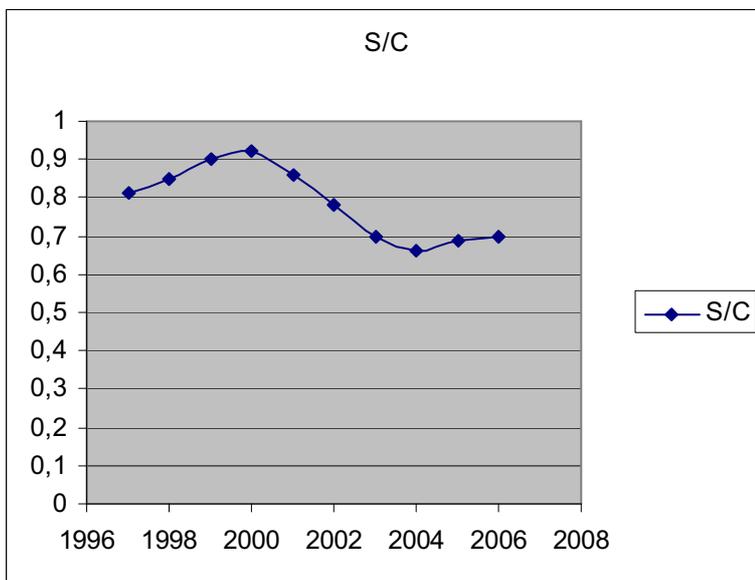
On donne ici quelques résultats concernant le marché des flottes automobiles en 2006. Les données sont issues des études de la FFSA. En 2006, les flottes automobiles ont totalisés 1.8 milliards d'euros de primes (environ 10% des cotisations automobile). Il représente 3 millions de véhicules (9% du parc automobile français). Le marché est globalement en expansion mais stagne depuis quelques années, d'où une concurrence plus rude entre les différents acteurs. Le S/C, qui s'était amélioré depuis 2000 est donc légèrement en hausse.

On donne ci-contre la répartition du marché par taille de flotte. On constate que les très gros contrats, bien que peu nombreux représentent une part essentielle du volume des primes : les 6% plus grosses flottes représentent 40% des primes. On constate aussi que la rentabilité des flottes (matérialisée par leur S/C) diminue nettement avec la taille. Les plus grosses flottes représentent des primes de plusieurs centaines de milliers d'euros. Elles disposent donc d'un pouvoir de négociation important et peuvent obtenir de gros rabais.

AXA est un des leaders français du marché, se positionnant deuxième sur l'automobile entreprise et premier sur les flottes. La compagnie est présente sur tous les secteurs du marché. La commercialisation est gérée par les directions régionales et le réseau d'agents généraux et de courtiers. Les questions techniques (établissement des tarifs, des conditions générales...) sont prise en charge par l'équipe auto entreprise basée à Nanterre au sein de la direction technique non vie.



Cotisation totale des flottes d'entreprises (en million d'euros)



S/C global du marché des flottes d'entreprises

| Nombre de véhicules | Part des contrats | Part des primes | S/C |
|---------------------|-------------------|-----------------|------|
| <10 | 60% | 22% | 0,62 |
| 10 à 50 | 34% | 38% | 0,63 |
| 50 à 200 | 5% | 21% | 0,7 |
| 200 à 999 | 1% | 12% | 0,87 |
| 1000 et + | quasi nulle | 8% | 0,99 |

Répartition des contrats flottes par nombre de véhicules

Source : enquête FFSA 2007

1.2 Les différents types de contrats automobile entreprise

Chez AXA, il existe deux types principaux de contrats auto entreprise : les contrats mono véhicules et les contrats flottes. Les contrats mono véhicules (ou mono), comme leur nom l'indique, ne couvrent qu'un seul véhicule. Ces contrats sont aussi utilisés pour les entreprises ayant peu de véhicules (moins de 5). On groupe alors les différents contrats mono. Au total, ces contrats représentent 58% du portefeuille mais seulement 5% des sinistres, 7% des primes et 12% des véhicules assurés. Nous ne nous y intéresseront plus par la suite. Par opposition, un contrat flottes permet de couvrir tout ou partie des véhicules d'une entreprise. Les contrats flottes sont subdivisés en plusieurs familles, selon divers critères (mode de gestion, poids, usage...). On donne ici les principales subdivisions :

○ par le mode de gestion :

On distingue les flottes fermées (ou à véhicules dénommés) des flottes ouvertes.

→ Dans les flottes fermées, on connaît le nombre de véhicule de la flotte ainsi que leur caractéristiques. En particulier, à chaque nouveau véhicule entrant dans la flotte, une prime pure est calculée qui vient s'ajouter à la prime du contrat.

→ Dans les flottes ouvertes, l'assureur ne connaît de façon précise ni le volume ni la composition de la flotte. La prime est en général calculée chaque année à partir du chiffre d'affaire. Ce type de flotte est plus dangereux car on ne peut pas ajuster le tarif de façon fine, faute d'information.

A AXA, la politique est de privilégier la gestion en flotte fermée. Cependant, pour les très grosses flottes (qui ont des flux de véhicules importants), il est plus commode d'avoir des flottes ouvertes. En pratique, les flottes de plus de 50 véhicules sont essentiellement de ce type. On voit ci contre que les flottes ouvertes sont assez minoritaires (15% des contrats) mais beaucoup plus grosses en moyenne que les flottes fermées. Pour être exhaustif, signalons un autre type de contrat : les contrats garages. Les garages s'assurent contre des dégâts causés aux véhicules sous leur garde. Ces contrats sont assez nombreux (12% du portefeuille) mais ont un fonctionnement spécifique. On ne les étudiera donc pas dans ce mémoire.

○ par poids :

Il est d'usage de distinguer les véhicules de plus de 3.5 tonnes (camions, bennes...) des autres. Une flotte sera dite plus de 3.5 tonnes (ou P3T5) si la majorité de ses cotisations vient des véhicules P3T5 et moins de 3.5 tonnes (M3T5) sinon. Il y a à peu près autant de contrats de chacun des poids. Néanmoins, une part importante des contrats P3T5 est composée de garages. En contrats flottes, les P3T5 sont assez minoritaires. La distinction entre M3T5 et P3T5 vient d'une différence au niveau des coûts de sinistres : les P3T5 ont des sinistres généralement plus élevés.

○ par usage :

Les flottes diffèrent par leur usage. La classification AXA comporte plusieurs dizaines de catégories. Par soucis de lisibilité, on donne la subdivision en 3 groupes qui est la plus commune :

→ les TPM (transports public de marchandises) regroupent les véhicules servant à transporter des marchandises n'appartenant pas à l'assuré (typiquement, entreprise de transport routier). Ces contrats ont en moyenne des coûts de sinistres élevés et une rentabilité plus faible.

→ les TPV (transports publics de voyageurs) regroupent les véhicules de transports de voyageurs (cars, autobus, taxis...). Ces contrats ont en général peu de sinistres mais plus de sinistres extrêmes (sinistres corporels multiples).

→ les AUT (autres) qui ne rentrent pas dans les deux catégories précédentes. Ils ont en général une sinistralité moindre. Ces contrats constituent l'essentiel du portefeuille et sont très divers. Citons les TPC (transport pour propre compte) -qui servent à transporter l'assuré ou des marchandises lui appartenant- et les engins de chantier.

| Type de gestion | % des contrats | % des primes | prime moyenne | % des sinistres |
|-----------------|----------------|--------------|---------------|-----------------|
| Mono véhicule | 58,11% | 7,24% | 645 | 5,05% |
| Flotte fermée | 23,12% | 50,36% | 11216 | 50,02% |
| Flotte ouverte | 6,04% | 27,79% | 23308 | 35,46% |
| garage | 12,57% | 14,54% | 5951 | 9,43% |

Répartition des contrats entre les différents types en 2007

| Type de flotte | % des contrats | % des primes | prime moyenne | % des sinistres | cout moyen | S/C moyen |
|----------------|----------------|--------------|---------------|-----------------|------------|-----------|
| M3T5 | 53% | 49% | 6945 | 55% | 1183 | 0,57 |
| P3T5 | 47% | 51% | 7642 | 45% | 1785 | 0,71 |
| AUT | 82% | 68% | 5647 | 67% | 1380 | 0,56 |
| TPM | 13% | 22% | 15963 | 21% | 1829 | 0,78 |
| TPV | 5% | 10% | 7476 | 12% | 1442 | 0,63 |

Répartition des flottes entre les différents poids et usages en 2007

1.3 Les différentes garanties

On donne brièvement une liste des garanties auxquelles peuvent souscrire les assurés.

- La garantie responsabilité civile (RC) :

On rappelle qu'en France, cette garantie est obligatoire.

- La garantie dommage :

Cette garantie peut prendre de nombreuses formes.

→ garantie contre le vol

→ garantie incendie

Cette garantie incluse généralement aussi les dégâts liés aux explosions, attentats, à la grêle et aux tempêtes.

→ garantie catastrophe naturelle

→ garantie contre la perte d'effets personnels laissés dans le véhicule

- La garantie bris de glace (BDG)

Couvre les dégâts aux vitres et aux phares.

Et deux garanties en nature :

- la protection juridique
- l'assistance (dépannage, remorquage du véhicule.)

Dans la suite de l'étude, les sinistres liés à ces différentes garanties ont été regroupés en trois groupes :

◇ la RC corporelle ou RCCORP. Les sinistres corporels sont peu nombreux mais leur coût est beaucoup plus élevé que ceux des autres catégories. De plus, il est plus incertain car le paiement peut s'effectuer plusieurs années après la survenance (recours en justice).

◇ la RC matérielle ou RCMAT.

◇ les dommages ou DOMM (garanties dommage et bris de glace). Ces sinistres sont les plus courants et les moins coûteux. Ils sont de plus à développement court.

Remarque :

Un même sinistre peut faire jouer plusieurs garanties à la fois. Par exemple, si un véhicule assuré en percute un autre, il fera jouer les trois grands types : DOM pour les dommages à son véhicule, RCMAT pour ceux au véhicule victime et RCCORP pour ceux au conducteur du véhicule percuté.

| Garantie | Proportion du portefeuille |
|-----------------------|----------------------------|
| Responsabilité civile | 99,4% |
| Domage | 63,8% |
| Incendie | 66,7% |
| Vol | 66,0% |
| Bris de glace | 62,9% |

Proportion des contrats ayant souscrit les différentes garanties en 2007

| Garantie | Nombre de sinistres (en milliers) | Cout moyen |
|---------------------|-----------------------------------|------------|
| RCMAT | 4498 | 760 |
| RCCORP | 263 | 6700 |
| DOM | 3864 | 456 |
| dont dommages | 2090 | 1812 |
| dont incendie | 46 | 6120 |
| dont vol | 220 | 3320 |
| dont bris de glace | 1508 | 312 |
| dont cat. naturelle | <1 | 8120 |

Fréquence et cout moyen des sinistres par garantie en 2007

Chapitre 2

Modélisation de la sinistralité

La sinistralité d'une compagnie d'assurance est un phénomène complexe. Néanmoins, à plusieurs reprises dans ce mémoire, nous devrons utiliser la loi de la charge totale d'un contrat. Par exemple, les sinistres seront écrêtés, c'est-à-dire qu'on fixera un montant maximum au-delà duquel un sinistre sera jugé aberrant. Ce seuil sera défini comme un quantile de la loi de la charge. Cette loi servira aussi pour évaluer la fiabilité de la moyenne empirique des S/C d'un contrat comme estimateur de sa rentabilité, et pour évaluer la stabilité dans le temps de nos modèles. Une solution serait d'utiliser la loi empirique. Néanmoins, elle présente plusieurs défauts : il est rare d'obtenir des formules fermées avec la loi empirique, elle est généralement moins lisse et elle risque d'être moins précise pour les valeurs de charges les plus élevées (fréquence des sinistres des très grosses flottes, coûts des gros sinistres corporels...) faute de données en nombre suffisant. Le recours à une loi paramétrique permet de traiter de façon plus stable ces queues de loi. C'est par exemple essentiel dans le cas de l'écrêtement.

En pratique, les flottes sont très diverses. Vouloir imposer une loi unique à l'ensemble du portefeuille n'a guère de sens. On se doute qu'une flotte de 1000 véhicules aura plus de sinistres qu'une de 10, qu'un camion coûte plus cher à réparer qu'une moto et est susceptible de causer des dégâts plus importants (matériels comme corporels). On va donc segmenter le portefeuille en plusieurs classes. Les facteurs pouvant intervenir sont nombreux. Pour la fréquence des sinistres on peut penser au nombre de véhicules, à la zone géographique ou à l'expérience des conducteurs. Le coût moyen dépend lui de la masse des véhicules, de leur valeur marchande, de leur utilisation. Il convient néanmoins de ne pas multiplier les segmentations car chaque classe doit contenir assez de données. Au final, on s'intéresse à 3 facteurs : la taille de la flotte (matérialisée par sa prime) pour la fréquence, le type des véhicules et leur usage pour le coût moyen.

Un facteur notable, qui revient très fréquemment dans ce mémoire est le poids très important des événements extrêmes. Pour presque toutes les grandeurs que nous étudieront (nombre de sinistres, coûts, S/C..), l'écart type est supérieur à la moyenne, qui dépasse largement la médiane et l'asymétrie est positive. Toutes les lois que nous allons utiliser auront donc des queues épaisses voire très épaisses.

Dans un premier temps, nous allons modéliser la fréquence des sinistres. Les contrats seront regroupés en classes de prime. Comme il s'agit d'une variable continue, on définira d'abord des seuils pour pouvoir segmenter le portefeuille. Ensuite, on étudiera la fréquence au sein de chaque classe. Après cela, nous modéliserons ensuite le coût de ces sinistres. Deux modèles seront étudiés : en regroupant les flottes par type de véhicule ou par usage. De plus, dans le cas des coûts, les sinistres exceptionnels jouent un rôle clé car ils représentent une part importante de la charge totale. La modélisation se fera donc en trois temps. On cherchera tout d'abord un seuil pour distinguer les sinistres ordinaires des graves. Ensuite, chacun des deux sera modélisé séparément. Enfin, on s'intéressera à la charge totale (ordinaire et écrêtée). Comme il n'y a pas d'expressions simples pour cette loi, on aura recours à des méthodes de type Monte Carlo.

2.1 Modélisation de la fréquence des sinistres

Le but de cette partie est d'estimer la loi du nombre de sinistres d'un contrat pendant une année. En pratique, les contrats sont loin d'être uniformes. La plupart ont un nombre moyen de sinistres inférieur à trois par ans quand certains peuvent dépasser les 100. Utiliser une loi unique est manifestement inapproprié. Dans le domaine des flottes d'entreprises, le facteur essentiel pour expliquer la fréquence de la sinistralité est la taille de la flotte : plus il y a de véhicules, plus il y aura d'accidents. On pourrait donc classer les contrats en fonction du nombre de véhicules. Deux problèmes se posent : tout d'abord nous ne le connaissons pas pour les flottes ouvertes. Ensuite, certains types de véhicules ont plus de sinistres que d'autres. On regroupera donc les contrats par classe de cotisations. Cette donnée est une bonne approximation de la taille de la flotte et est disponible pour tous les contrats.

Pour pouvoir estimer une loi, il faut disposer d'un nombre suffisant d'observations. On va donc regrouper les contrats en classes. Vu ce qui précède elles seront déterminées par le niveau de cotisation. On va tout d'abord déterminer la meilleure segmentation possible (meilleure au sens où les classes sont le plus homogènes possibles). Ensuite, on modélisera une loi du nombre de sinistres pour chacune de ces classes. Tout d'abord, on décrit les différentes lois envisagées et les outils pour évaluer leurs paramètres et tester l'adéquation aux données. On appliquera ensuite ces méthodes au portefeuille. Enfin, on modélisera séparément la fréquence d'un type particulier de sinistres : les remboursements forfaitaires IDA.

2.1.1 Répartition des contrats par classe de cotisations

Dans ce paragraphe, on cherche à répartir les contrats en classe de cotisations. On les choisit de façon à ce que le nombre de sinistres soit le plus homogène possible au sein de chaque classe.

2.1.1.1 Méthode de segmentation

Soit n le nombre de contrats du portefeuille, S_i la charge du sinistre i , C_i sa cotisation et N_i son nombre de sinistres. $N = \{N_1..N_n\}$. On suppose les contrats répartis en k classes $\mathcal{C}_1..C_k$ et on note \bar{N}_k le nombre moyen de sinistres des contrats de la classe k et \bar{N} celui du portefeuille. La variance totale de N se décompose en :

$$V(N) = V_1(N) + V_2(N)$$

$$\text{avec } V_1(N) = \frac{1}{n} \sum_k \sum_i (\bar{N}_k - \bar{N})^2 \text{ et } V_2(N) = \frac{1}{n} \sum_k \sum_{i \in \mathcal{C}_k} (N_i - \bar{N}_k)^2.$$

Le premier terme est la variance interclasse ou variance expliquée. Elle mesure l'hétérogénéité entre les classes. Le second est la variance intraclasse ou variance résiduelle. Elle mesure l'hétérogénéité à l'intérieur des classes. Comme on cherche à avoir des classes les plus homogènes possibles, on veut minimiser la variance intraclasse ou, ce qui revient au même, maximiser la variance interclasse.

En pratique, pour augmenter cette quantité, il suffit d'augmenter le nombre de classe. A l'extrême, si chacune ne contient qu'un contrat, la variance intraclasse est nulle. Néanmoins, on souhaite limiter le nombre de classes. En effet, chacune doit contenir assez de données pour permettre d'appliquer de façon robuste les outils statistiques.

On veut donc maximiser la variance expliquée à nombre de classe fixée.

La qualité de la segmentation est d'autant plus grande que la variance interclasse et la variance totale sont proches donc d'autant plus grande que le rapport $R^2 = \frac{V_1(N)}{V(N)}$ est proche de 1. En pratique, c'est cette statistique que l'on utilisera pour comparer les diverses segmentations.

On décrit maintenant l'algorithme de choix des classes. C'est une méthode de type pas à pas ascendante. Les cotisations des contrats sont comprises entre 0 et un maximum de M euros.

★ Etape 1 : on introduit un seuil s_1 ce qui nous donne deux classes $[0; s_1]$ et $[s_1; M]$. Le seuil s_1 est choisi de façon à maximiser le R^2 , par exemple par une méthode de dichotomie ou de Newton-Raphson.

★ Etape 2 : on introduit le seuil s_2 de façon à obtenir trois classes (si $s_2 > s_1$, il s'agira de $[0; s_1]$, $[s_1; s_2]$ et $[s_2; M]$). A nouveau, le seuil est choisi pour maximiser le R^2 .

★ Etape n : On cherche dans chacune des n classes existantes la partition $C = C' \cup C''$ qui augmente le plus le R^2 . On garde ensuite celle qui conduit au R^2 le plus élevé.

On poursuit le processus jusqu'à ce que le R^2 ne puisse plus être significativement augmenté.

Cette méthode ne donne pas nécessairement la meilleure solution à nombre de classes fixées mais fournit en général un résultat proche de l'optimal.

Remarque :

Ce problème peut être vu comme un cas particulier de modèle linéaire.

On veut expliquer la variables nombre de sinistres à l'aide des indicatrices 1_{C_k} : on approxime N_i par $\tilde{N}_i = \sum_k \lambda_k 1_{C_k}(i)$. Les coefficients λ_k sont choisis de façon à minimiser :

$\sum_i (N_i - \tilde{N}_i)^2$. La solution optimale est $\lambda_k^* = \bar{N}_k$. La régression est d'autant meilleure que les écarts

entre N et \tilde{N} sont petits. On les mesure par $R^2 = \frac{\sum (N_i - \tilde{N}_i)^2}{\sum N_i^2}$. Pour la solution optimale, le R^2

vaut $\sum_k \sum_{i \in C_k} (N_i - \bar{N}_k)^2$ c'est - à dire la variance résiduelle.

L'intérêt de ce point de vue est que l'on peut donc déterminer la segmentation en classes du portefeuille à l'aide de SAS en utilisant la procédure REG (régression linéaire) avec l'option FORWARD.

Une contrainte supplémentaire est que chaque classe doit contenir suffisamment de contrats. En effet, les S/C très élevés de certains contrats seront en partie redistribués sur les membres de leur classe via deux mécanismes : l'écrêtement-mutualisation qui traite les sinistres de montants anormalement élevés (cf. paragraphe 3.1) et la crédibilité qui amortit les années anormalement mauvaises (cf. paragraphe 3.2). Pour éviter que ces mécanismes n'impactent trop des contrats à priori bons, on doit avoir des classes assez importantes. Les plus gros contrats ayant une capacité d'absorption plus grande, cela joue moins pour eux.

Pour des raisons de souplesse et de lisibilité, il est également souhaitable que le nombre de classes ne soit pas trop important. Par comparaison avec les modèles utilisés au paravent, on vise une segmentation comprise entre trois et huit classes. Comme on distinguera ensuite les M3T5 des P3T5, cela donne au finale une classification comprise entre six et seize classes (contre huit et quatorze pour les deux modèles actuels).

2.1.1.2 Subdivision du portefeuille

On a appliqué l'algorithme ci dessus au portefeuille. Les résultats sont représentés par l'arbre ci contre. Chaque ligne de l'arbre représente une étape de l'algorithme et chaque cellule une des classes. On donne aussi la variance résiduelle venant de chaque classe (comme pourcentage de la variance totale). Le R^2 obtenu à une étape vaut un moins la somme des pourcentages de la ligne. On constate que très rapidement le R^2 se stabilise. Après l'étape 5 il n'évolue pratiquement plus. A ce stade, la segmentation n'explique que 46 % de la variance totale ce qui reste assez faible. Ce n'est pas surprenant : la fréquence des sinistres ne dépend pas que de la taille d'une flotte mais aussi d'autres facteurs (types de véhicules, zones géographiques, usage des véhicules...) et du hasard. On ne pouvait donc pas espérer obtenir un R^2 très élevé.

Au delà de l'étape 5, les gains de R^2 sont très faibles. On obtient donc un modèle à 6 classes : $[0; 10000]$ $[10000; 25000]$ $[25000; 75000]$ $[75000; 250000]$ $[250000; 550000]$ et $[550000; +\infty[$.

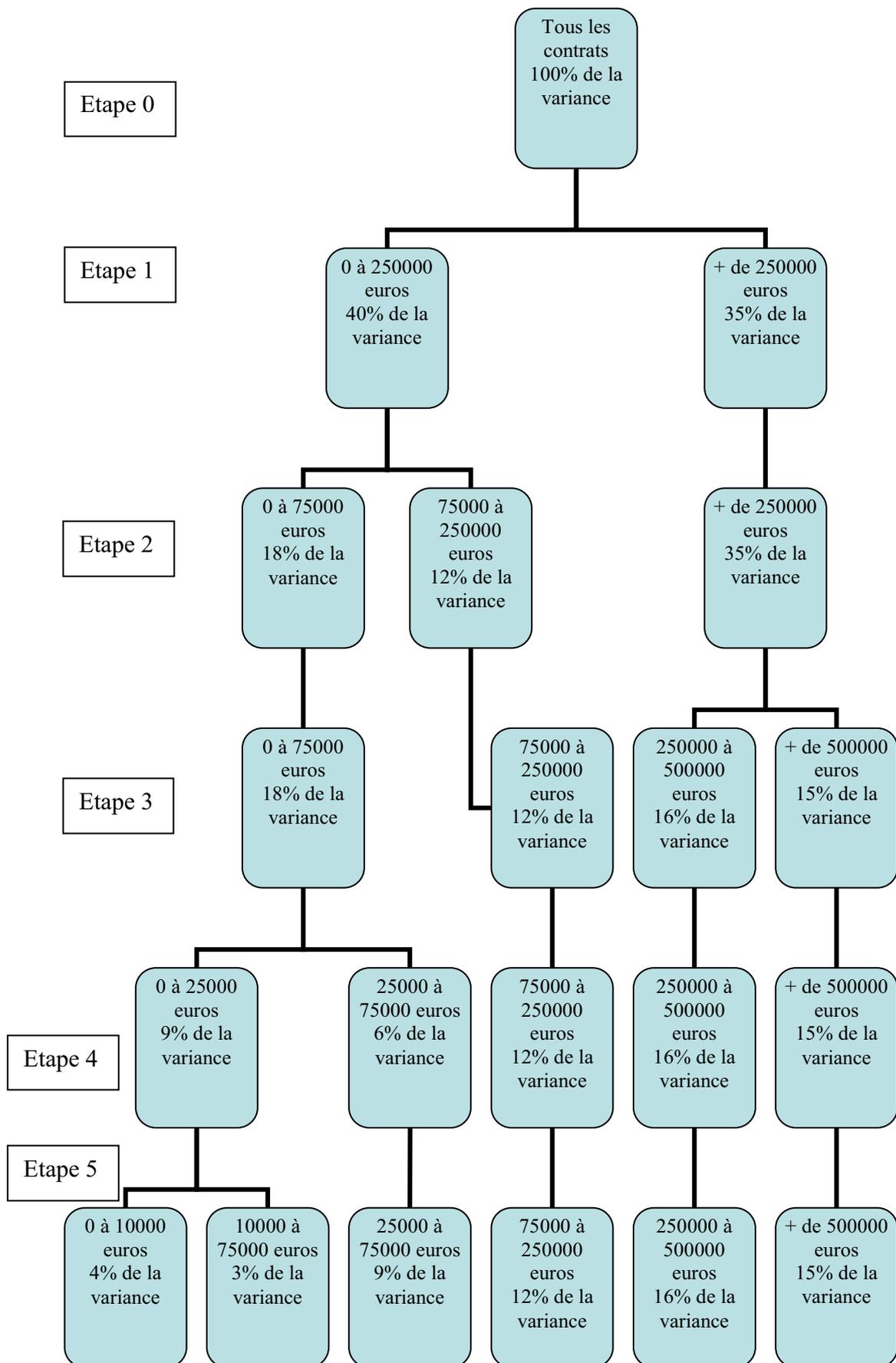
Remarque :

Pour des raisons de lisibilité, on a arrondi les seuils au multiple de 5000 euros le plus proche.

On constate que les dernières classes qui comptent très peu de contrats représentent néanmoins une part importante de la variance résiduelle. Cela vient du fait que pour les contrats petit à moyen (moins de 10000 euros) le nombre de sinistres ne dépasse presque jamais 5 (rarement 2 pour les moins de 5000). Par contre, pour les très grosses flottes (classes 5 et 6), le nombre de sinistres peut prendre des valeurs beaucoup plus variées (typiquement de 200 à 600 pour la plus grosse classe). Elles sont donc bien plus hétérogènes.

En pratique, la subdivision obtenue n'est pas satisfaisante :

- les deux dernières classes $[250000; 550000]$ et $[550000; +\infty[$ comptent très peu de contrats (moins de 100 pour l'avant dernière et moins de 20 pour la dernière). On les regroupe donc en une classe $[250000; +\infty[$.
- à l'inverse, la première classe $[0; 10000]$ contient les deux tiers des contrats. On l'a donc subdivisé en deux en suivant toujours le critère de R^2 maximum pour obtenir les classes $[0; 5000]$ et $[5000; 10000]$.



Algorithme de segmentation du portefeuille par la cotisation

On donne ci contre un résumé de l'algorithme de segmentation, la première subdivision et celle finalement retenue.

On constate un phénomène important en automobile d'entreprise (qui n'existe pas en automobile de particulier) : la part prépondérante des très gros contrats. Les contrats des deux dernières classes ne représentent que 3% du portefeuille mais un tiers des sinistres.

Ces contrats concentrent de plus l'essentiel de la variance non expliquée.

En fait, une part importante de l'activité provient de quelques contrats énormes (grosses entreprises de transport routier, flotte de dirigeants de grandes entreprises...). Ces contrats possèdent donc un fort pouvoir de négociation, ce qui leur permet d'obtenir des rabais importants. Ils sont donc en moyenne moins bons que le reste du portefeuille et il convient de les surveiller attentivement. On conserve donc la classe 6 malgré son nombre très faible de contrats. Cela aurait d'ailleurs pu nous inciter à conserver une classe pour les contrats de plus de 500000 euros de cotisations. Néanmoins cela n'aurait pas apporté grand chose. La segmentation intervient essentiellement à deux étapes :

- Pour l'écrêtement des sinistres. On fixe une valeur maximale M aux sinistres, qui varie avec la taille de la flotte. En pratique, pour les contrats de la classe 6, ce niveau est déjà très élevé et l'augmenter encore pour les plus de 500000 euros aurait un impact faible.

- Pour la crédibilité. En pratique, les contrats de la classe 6 sont pratiquement crédibilisés à 100% donc ne sont pas affectés.

Au final, séparer en deux la classe 6 permettrait certes d'augmenter le R^2 mais compliquerait le modèle tout en ne changeant presque rien aux résultats. On conserve donc le découpage ci-contre.

| Etape | Seuil ajouté | R2 |
|-------|--------------|------|
| 1 | 250000 | 0,25 |
| 2 | 75000 | 0,35 |
| 3 | 550000 | 0,39 |
| 4 | 25000 | 0,42 |
| 5 | 10000 | 0,45 |

Étapes successives de la segmentation

| Classe | Cotisation | Pourcentage de la variance | Pourcentage des contrats | Pourcentage des sinistres | Nombre moyen de sinistres |
|--------|-----------------|----------------------------|--------------------------|---------------------------|---------------------------|
| 1 | 0 à 10000 | 4% | 67% | 17% | 0,5 |
| 2 | 10000 à 25000 | 3% | 20% | 20% | 1,6 |
| 3 | 25000 à 75000 | 9% | 10% | 27% | 4,0 |
| 4 | 75000 à 250000 | 12% | 2,55% | 24% | 13,3 |
| 5 | 250000 à 550000 | 16% | 0,29% | 8% | 45,0 |
| 6 | plus de 550000 | 15% | 0,05% | 5% | 132,0 |

Première segmentation obtenue

| Classe | Cotisation | Pourcentage de la variance | Pourcentage des contrats | Pourcentage des sinistres | Nombre moyen de sinistres |
|--------|----------------|----------------------------|--------------------------|---------------------------|---------------------------|
| 1 | 0 à 5000 | 2% | 44% | 7% | 0,3 |
| 2 | 5000 à 10000 | 2% | 23% | 10% | 0,7 |
| 3 | 10000 à 25000 | 3% | 20% | 20% | 1,6 |
| 4 | 25000 à 75000 | 9% | 10% | 27% | 4,0 |
| 5 | 75000 à 250000 | 12% | 2,55% | 24% | 13,3 |
| 6 | plus de 250000 | 31% | 0,34% | 13% | 61,0 |

Segmentation finalement retenue

2.1.2 Méthodes d'ajustements des lois de fréquences :

Au sein de chacune des classes de cotisations définies plus haut, il reste à estimer la loi du nombre de sinistre par an. Plusieurs lois discrètes vont être testées.

◇ La loi de Poisson $\mathcal{P}(\lambda)$ donnée par $P(N = n) = \frac{\lambda^n}{n!} e^{-\lambda}$.

Le paramètre λ est estimé par le nombre moyen de sinistres de la classe. Cette valeur est à la fois l'estimateur des moments et celui du maximum de vraisemblance.

La loi de Poisson a été retenue car il s'agit de la loi discrète la plus classique. Néanmoins, elle a le défaut d'être à queue fine. En particulier, $E(N) = V(N)$. Or dans toutes classes, l'espérance du nombre de sinistres est très inférieure à la variance (l'espérance est de l'ordre de l'écart type).

◇ La loi binomiale négative $\mathcal{B}(m, p)$ donnée par $P(N = n) = C_{m+n-1}^n p^m (1-p)^n$. Cette loi a le mérite d'être à queue épaisse et de vérifier $E(N) < V(N)$.

Les estimateurs des moments sont donnés par $\hat{p} = \frac{\bar{N}}{\hat{\sigma}^2}$ et $\hat{m} = \bar{N} \frac{\hat{p}}{1 - \hat{p}}$.

Il n'y a pas de formules fermées pour le maximum de vraisemblances mais les EMV vérifient :

$\tilde{p} = \frac{\tilde{m}}{\tilde{m} + \bar{N}}$ et $\sum_{j \geq 0} \frac{1}{\tilde{m} + j} (1 - \sum_{k \leq j} f_k) = \ln(1 + \frac{\bar{N}}{\tilde{p}})$ où f_k est la proportion de contrats ayant eu k

sinistres. En pratique, on utilisera la méthode du maximum de vraisemblance. La formule ci dessus sera résolue par une méthode de Newton en prenant comme valeurs initiales les estimateurs des moments.

◇ La loi logarithmique $L(b)$ donnée par $P(N = n) = \frac{-1}{\ln(1-b)} \frac{b^{n+1}}{n+1}$.

Cette loi est à décroissance lente ce qui correspond bien aux données. Elle a aussi le mérite de ne dépendre que d'un seul paramètre. Néanmoins, la suite $P(N = n)$ est nécessairement décroissante. Pour nos données, cela n'est vrai que pour les trois premières classes (et presque vrai pour la quatrième).

L'estimateur des moments coïncide avec l'EMV : \hat{b} vérifie $\frac{-\hat{b}}{(1-\hat{b})\ln(1-\hat{b})} = \bar{N}$.

◇ On a aussi testé les versions "doubles" des lois ci dessus. Par exemple, une variable X suit une loi de Poisson double si $X \sim \mathcal{P}(\lambda_1)$ avec une probabilité p et $X \sim \mathcal{P}(\lambda_2)$ avec une probabilité $1-p$. Ces lois permettent une meilleure adéquation que leurs variantes simples mais avec un nombre de paramètres plus élevés.

Il n'y a pas de formule naturelle pour évaluer ces paramètres.

On les a choisis de façon à minimiser le khi deux d'adéquation :

$\sum_{n \geq 0} \frac{(freq^{th}(n) - freq^{obs}(n))^2}{freq^{th}(n)} \sqrt{n}$. Où $freq^{th}(n)$ et $freq^{obs}(n)$ désignent les fréquences théoriques et observées de n .

Remarque :

On n'a pas retenu la loi binomiale car elle vérifie $E(N) > V(N)$ et ne s'ajusterait donc pas aux données.

Une fois qu'on s'est fixé les lois candidates, reste à déterminer celle qui représente le mieux les données. Dans le cas des lois discrètes, l'outil classique est le χ^2 d'adéquation. On rappelle brièvement son principe.

Soit $N_1..N_n$ n réalisations iid d'une variable entière N . On répartie \mathbb{N} en k classes. On note $p_i = P(N \in C_i)$ et P_i^{obs} la proportion d'observations appartenant à la classe i . On note $S = \sum_i \frac{(p_i - p_i^{obs})^2}{p_i}$. Alors S tend en loi vers $\chi^2(k - l - 1)$ où l est le nombre de paramètres estimés.

S mesure l'écart entre les observations et la distribution empirique. En pratique, on choisira donc la loi qui minimise la valeur de la statistique S .

S permet de tester l'adéquation des observations à la loi envisagée : on rejette cette loi si S est supérieur au quantile à 5% de la loi $\chi^2(k - l)$. Néanmoins, nous travaillons avec un ensemble de données très important (plus de 10000 contrats dans certaines classes). Dans ce cas, les tests d'adéquation sont rarement positifs (modéliser 10000 données par au plus trois paramètres reste assez rare). On utilisera donc plutôt S comme outils de comparaison entre les modèles que comme test de validité.

Un autre test permet de vérifier l'ajustement à ces lois. Il consiste à utiliser la relation de Panjer : $P(N = n) = (a + \frac{b}{n-1}) * P(N = n - 1)$ valable pour les lois de Poisson, binomiales et

binomiales négatives On peut la réécrire $n \frac{P(N = n + 1)}{P(N = n)} = an + b$. L'idée est de tracer les points

$(n, n \frac{P(N = n + 1)}{P(N = n)})$. Sous l'hypothèse binomiale négative, ils doivent être alignés suivant une droite de pente positive, pour une loi binomiale la pente serait négative et nulle pour une loi de Poisson).

2.1.3 Lois des fréquences

2.1.3.1 Fréquences des sinistres classiques

On a constitué une base de données regroupant les nombres des sinistres par ans des contrats du portefeuille AXA depuis 2005. Pour les contrats ayant existés pendant plusieurs années, chaque année compte comme une observation (par exemple, un contrat dans le portefeuille depuis 2005 fournit 3 observations : 2005, 2006 et 2007). On n'a gardé que les observations ayant existé pendant toute l'année : un contrat résilié en milieu d'année aurait deux fois moins de sinistres sans que cela soit lié à sa taille. Le défaut de ce choix est que les sinistres survenus en 2008 n'ont pas été pris en compte, l'année 2008 n'étant pas terminée. Néanmoins, inclure les contrats n'ayant pas une année complète augmente artificiellement l'hétérogénéité au sein des classes et l'ajustement des lois se fait mal. Enfin, les sinistres forfaitaires IDA qui sont traités à part ont été supprimés de la base.

On donne ci dessous les valeurs du khi deux d'adéquation des différentes lois aux classes de cotisation.

○ La loi de Poisson a une très mauvaise adéquation, ce qui était prévisible étant donné que dans toutes les classes la variance est très supérieure à l'espérance.

○ La loi logarithmique s'ajuste également très mal (en fait, à peine mieux que la loi de Poisson). Pour les grandes classes, on pouvait s'en douter : la suite des $P(N = n)$ ne décroît pas tout de suite avec n . Pour les petites classes, c'est plus étonnant. Cela vient peut être du fait que cette loi n'est pas encore à queue assez épaisse.

○ La loi binomiale négative s'ajuste plus ou moins bien suivant les classes, mais dans tous les cas c'est de loin la loi simple qui modélise le mieux les données.

○ Les lois Poisson et logarithmiques doubles donnent des résultats bien meilleurs que leur équivalent simple mais très en dessous de la loi binomiale négative.

○ La loi binomiale négative double par contre n'est guère meilleure que la loi BN simple alors que le nombre de paramètres est plus élevé (5 contre 2).

Au final, on a donc retenu la loi binomiale négative pour modéliser la fréquence des sinistres. On donne ci contre les p-valeurs correspondantes pour les différentes classes. L'ajustement est très bon pour la classe 6, acceptable pour les classes 2 et 3 et plus médiocre pour les autres. Néanmoins, vu le nombre de données et le fait que cette loi est de loin la meilleure disponible, on la conserve.

Remarque :

Si l'on cherche à modéliser la loi de la fréquence des sinistres sur l'ensemble du portefeuille sans le subdiviser en classe, le résultat est extrêmement mauvais : aucune loi ne s'ajuste. La p-valeur de la loi binomiale négative est de l'ordre de 10^{-8} . L'hétérogénéité entre contrats est trop forte.

Le graphe correspondant au test graphique pour les différentes classes est donné ci contre.

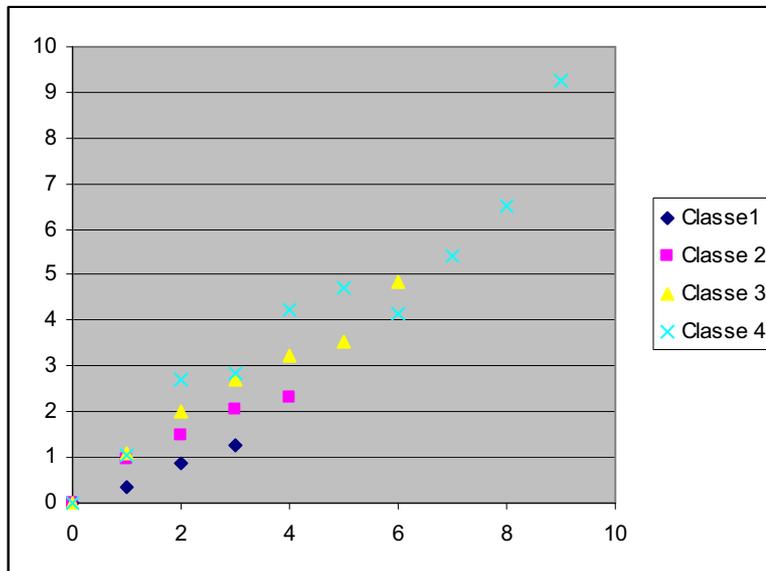
On n'a gardé que les points représentant au moins 5% des observations (les autres sont peu fiables). En particulier pour les classes 5 et 6, les valeurs possibles sont beaucoup plus nombreuses et aucun point n'est jugé fiable donc le test n'a pas été utilisé. On constate que les points sont bien alignés ce qui tend à valider l'hypothèse que les fréquences de sinistres sont approximativement binomiales négatives.

| Classe | Khi 2 Binomiale négative | Khi 2 Poisson | Khi 2 logarithmique | Khi 2 BN double | Khi2 Poisson double | Khi2 logarithmique double | Nombre de classes |
|--------|--------------------------|---------------|---------------------|-----------------|---------------------|---------------------------|-------------------|
| 1 | 10 | 2122 | 178 | 7,5 | 19 | 20 | 6 |
| 2 | 8 | 1361 | 2004 | 7 | 110 | 122 | 7 |
| 3 | 11,7 | 4350 | 2514 | 11,6 | 402 | 495 | 9 |
| 4 | 21,6 | 18475 | 1796 | 21,5 | 1200 | 1370 | 14 |
| 5 | 25 | 8600 | 5300 | 24,3 | 1312 | 2815 | 13 |
| 6 | 6,3 | 22517 | 15815 | 5,5 | 13221 | 9615 | 11 |

Adéquation des lois aux données

| Classe | P-valeur |
|--------|----------|
| 1 | 1,8% |
| 2 | 9,1% |
| 3 | 6,9% |
| 4 | 2,7% |
| 5 | 0,5% |
| 6 | 61,0% |

P-valeur de la loi binomiale négative



Test graphique d'adéquation de la loi binomiale négative

2.1.3.2 Cas des sinistres IDA forfaitaires

En assurance automobile, un grand nombre de sinistres concernent des dommages de faibles montants liés à un accident entre deux conducteurs. Les assureurs ont cherchés à simplifier la gestion de ces sinistres (ce qui permet d'économiser des frais généraux et d'indemniser plus rapidement l'assuré). A cette fin, la plupart des assureurs français ont mis en place un accord d'indemnisation : la convention IDA (indemnisation directe de l'assuré).

Cette convention s'applique pour les sinistres RC n'impliquant que deux véhicules (hors véhicules de transport) et dont les dommages ne dépassent pas 6500 euros par assuré. Dans ce cas, chaque assureur rembourse directement son assuré. Si l'un d'entre eux est jugé responsable du sinistre, son assureur reverse à celui de la victime un montant forfaitaire (1204 euros en 2008 pour un sinistre pleinement responsable). Cette convention permet un traitement très rapide de la plupart des sinistres automobiles.

Les indemnités IDA représente une proportion importante des sinistres du portefeuille (de l'ordre de 15 à 20 % chaque année). Il s'agit de sinistres spéciaux car leurs coûts sont tous identiques. En particulier, une telle proportion de sinistres ayant tous le même montant pose des problèmes importants pour modéliser une loi de coûts. Il a donc été décidé de les traiter à part. On a réalisé une base de données similaire à celle des sinistres ordinaires et on a modélisé la loi des fréquences en reprenant les méthodes ci dessus. Les résultats sont donnés dans le tableau ci contre. A nouveau, la loi binomiale négative est de loin la meilleure et l'ajustement est excellent.

Remarque :

1) L'ajustement est bien meilleur que pour les sinistres ordinaires. Cela vient sans doute du fait que les sinistres IDA sont moins nombreux. Le nombre de classe est donc plus faible ce qui induit en général une meilleure adéquation.

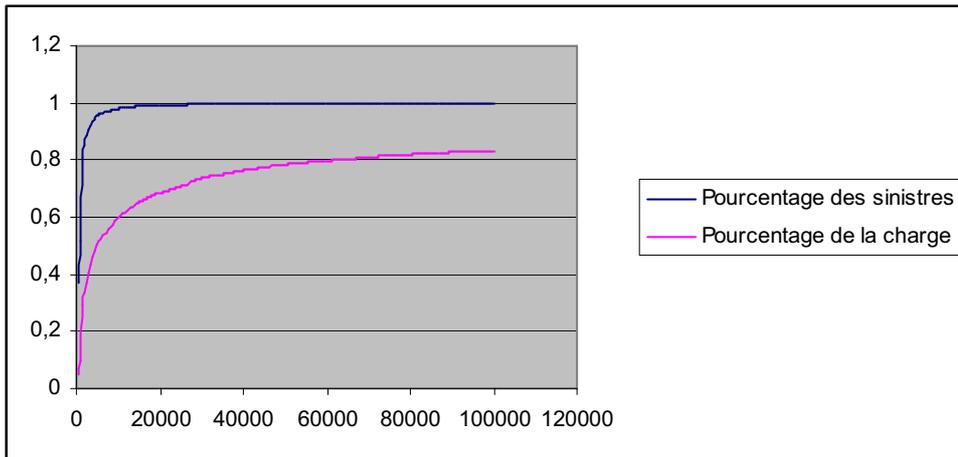
2) On n'a pas utilisé le test graphique donné en 2.1.3 car les seuls valeurs ayant une probabilité supérieures à 5% sont 0, 1, 2 (et 3 pour les plus grandes classes). Il n'y a donc pas assez de points fiables pour pouvoir appliquer ce test.

| Classe | Khi 2 Binomiale négative | Khi 2 Poisson | Khi 2 logarithmique | Nombre de classes |
|--------|--------------------------------|------------------|------------------------|----------------------|
| 1 | 0,378 | 211,15 | 2,36 | 4 |
| 2 | 0,68 | 241,52 | 25,34 | 4 |
| 3 | 6,97 | 1119,81 | 119,19 | 5 |
| 4 | 4,03 | 1364,35 | 378,38 | 7 |
| 5 | 6,29 | 884,66 | 651,23 | 6 |
| 6 | 0,099 | 115,94 | 1024,57 | 6 |

Adéquation des lois aux données

| Classe | P-valeur |
|--------|----------|
| 1 | 53,8% |
| 2 | 40,9% |
| 3 | 3,0% |
| 4 | 25,8% |
| 5 | 9,8% |
| 6 | 99,1% |

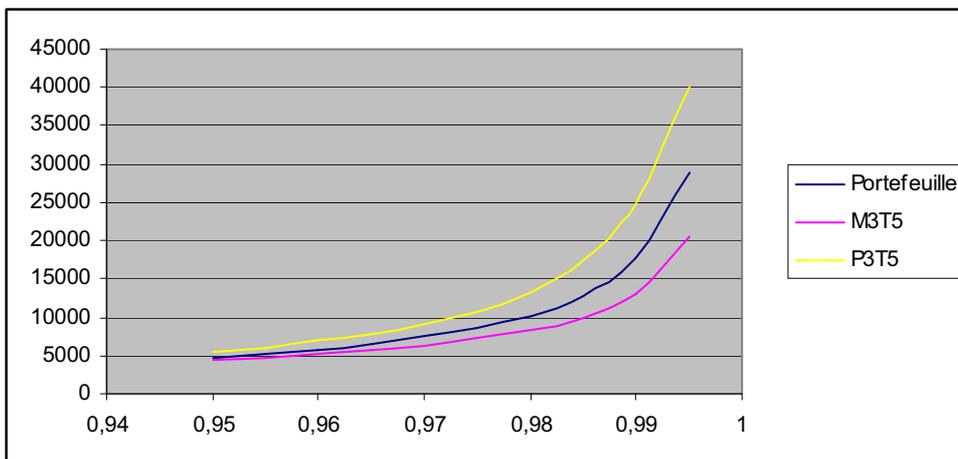
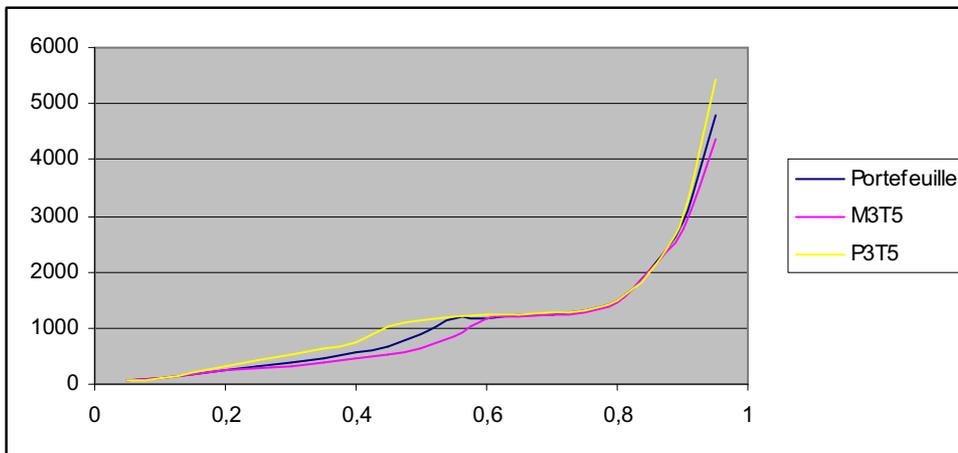
P-valeur de la loi binomiale négative



Proportion du nombre de sinistre et de la charge totale venant des sinistres inférieurs à un seuil donné.

| | Cout moyen | Cout médian | Quantile à 75% | Ecart type |
|--------------|------------|-------------|----------------|------------|
| Portefeuille | 1885 | 890 | 1300 | 20810 |
| M3T5 | 1558 | 633 | 1267 | 17821 |
| P3T5 | 2299 | 1155 | 1328 | 24302 |

Coûts des sinistres en fonction du type de véhicules



Loi du coût de sinistre pour les différents type de véhicules.

2.2 Modélisation de la sévérité des sinistres

Dans cette partie, on cherche une loi permettant de modéliser le montant des sinistres. En assurance automobile, ils peuvent avoir des valeurs très diverses. Alors que la plupart des sinistres ont un coût inférieur à 1000 euros (bris de glace, dégâts de carrosserie..), certains atteignent des niveaux supérieurs (vol du véhicule, dégâts importants..) voir très supérieur dans les cas des dommages corporels. A titre d'illustration, le coût moyen est de 1885 euros, bien au dessus du coût médian qui est de 890 euros. Cet écart mesure l'importance des gros sinistres. Un autre élément allant dans ce sens est la valeur très élevée de l'écart type : 21 000 euros. Tout ceci reflète le fait que les plus gros coûts jouent un rôle prépondérant. On donne la courbe représentant pour chaque seuil le pourcentage de sinistres de coût inférieur au seuil et la proportion de la charge totale représentée par ces sinistres. Alors que la proportion du nombre atteint très rapidement des valeurs proches de 1, la proportion des coûts croît bien plus lentement. En somme, l'essentiel de la charge provient des sinistres graves.

Il y a plusieurs façons de tenir compte de ce qui précède. On pourrait chercher à modéliser le coût des sinistres par une loi à queue très épaisse type Pareto ou log-gamma. Néanmoins, le résultat est assez médiocre : selon la valeur des paramètres, la courbe s'ajustera mal soit aux faibles coûts (qui représentent l'essentiel des données), soit à la queue de distribution (qui représente l'essentiel de la charge). En fait, les deux portions de la courbe semblent obéir à des lois relativement différentes. On a donc choisi de les modéliser séparément. Les approches permettant cela relèvent de la théorie des valeurs extrêmes (EVT en anglais). La méthode retenue est la méthode « Peak over threshold » ou POT. Elle consiste à se fixer un niveau T au delà duquel les sinistres seront considérés comme extrêmes et à modéliser séparément la loi des excès par rapport à ce seuil, c'est-à-dire la loi de $(C - T)_+$. Au final, un seuil de 15000 euros a été retenu soit environ 1% des sinistres.

Le portefeuille AXA contient des contrats et des types de véhicules très divers. Or tous n'ont pas la même gravité de sinistres. Typiquement, les gros véhicules (camions, engins..) ont des sinistres nettement plus lourds que les autres. Il existe de nombreux critères pour segmenter les flottes par type de véhicules et d'utilisation. La classification AXA comporte plusieurs dizaines de catégories (taxis, transporteurs de marchandises, tracteurs, bennes à ordures..). Néanmoins, pour disposer d'estimations robustes, il est nécessaire d'avoir assez d'observations dans chaque classe. Au final, dans le domaine des flottes d'entreprise, le critère le plus utilisé est le poids : plus ou moins de 3.5 tonnes. On donne à la page suivante les lois empiriques de sinistres pour ces deux familles de contrats. On constate que les P3T5 ont des sinistres supérieurs aux M3T5, en particulier dans la queue de distribution. Deux causes à cela : la valeur des véhicules (un poids lourd coûte plus cher à réparer ou à remplacer qu'une camionnette) et la gravité des sinistres corporels (à vitesse égale, il est plus dangereux d'être renversé par un camion que par une voiture). On modélisera donc séparément les coûts pour les M3T5 et les P3T5.

On va d'abord modéliser le gros de la courbe des coûts. On commence par donner les différentes lois et les outils pour choisir la meilleure. On applique ensuite ces méthodes au portefeuille pour déterminer une loi pour les sinistres attritionnels. Dans un deuxième temps, on rappelle les éléments d'EVT utilisés. On s'en sert pour modéliser les coûts des sinistres les plus graves. Enfin, on va étudier une autre possibilité de segmentation du portefeuille tenant compte de l'usage des véhicules (transport de marchandise, de voyageurs ou autre).

2.2.1 Distinction entre sinistres graves et ordinaires

2.2.1.1 Les coûts suivent-ils une loi extrême ?

Cette partie vise à déterminer si la théorie des valeurs extrêmes s'applique bien aux coûts et si oui à choisir le seuil T utilisé. On vérifie d'abord que nos intuitions sont justes et que la loi est bien à queue épaisse. Un premier critère consiste à tracer un QQ plot avec la loi exponentielle : on donne le graphe des $(q_{\text{empirique}}; q_{\text{exponentiel}})$ où les q sont les quantiles empiriques et de la loi $\exp(1)$. Si les coûts croissent suivant une loi exponentielle, le graphe sera à peu près une droite. Si la queue est moins épaisse (par exemple normale), le graphe sera convexe. Si la queue est plus lourde, il sera concave. On observe des courbes fortement concaves, aussi bien pour les M3T5 que pour les P3T5. La loi est donc à queue assez épaisse.

On n'a tracé que la partie de la courbe correspondant aux sinistres de plus de 3000 euros. En effet, on s'intéresse ici à la loi des très gros sinistres. Pour le reste de la courbe, la concavité est moins marquée mais reste présente.

Un deuxième outil utile pour évaluer l'épaisseur de queue est l'espérance résiduelle $e(u) = E(C - u | C > u)$. Cette quantité peut être vue comme une T-Var. Elle mesure le risque restant sachant que l'on est au dessus d'un seuil. Plus une loi est à queue épaisse, plus son espérance résiduelle croît rapidement. A titre d'exemple, on donne le comportement de $e(u)$ pour certaines lois :

- Normale : $e(u)$ tend vers 0
- Exponentielle (λ) : constante égale à $\frac{1}{\lambda}$.
- Gamma (n, λ) : tend en décroissant vers $\frac{1}{\lambda}$
- Log normale : croît approximativement comme $\frac{u}{\ln u}$
- Log-gamma (n, λ) : croît approximativement comme $\frac{u}{\lambda - 1}$ si $\lambda > 1$, infini sinon.
- Pareto (m, α) : croît linéairement avec coefficient directeur $\frac{\alpha}{\alpha - 1}$ si $\alpha > 1$, infini sinon .

En pratique, on a estimé $e(u)$ par $\widehat{e}(u) = \frac{1}{N_u} \sum_{C \leq u} C - u$ où N_u est le nombre de sinistre supérieurs à

u . Les courbes sont données ci contre. On constate que $\widehat{e}(u)$ croît linéairement avec u . Cela paraît cohérent avec une hypothèse de loi à queue épaisse mais à espérance finie. En particulier, c'est compatibles avec une loi de Pareto (de paramètre $\alpha > 1$), loi utilisée en EVT.

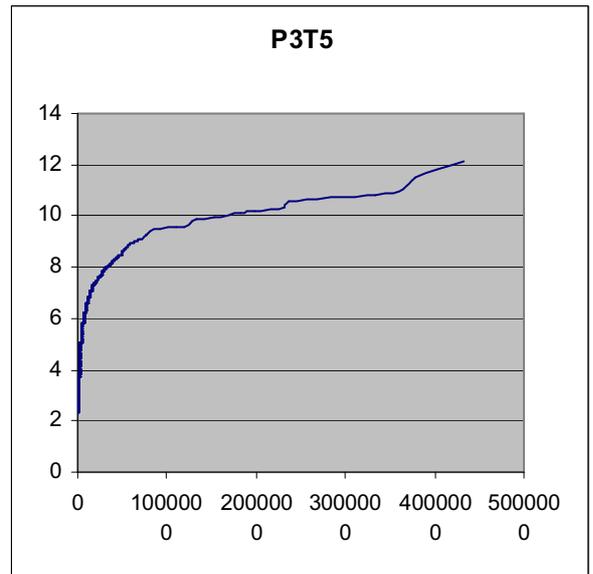
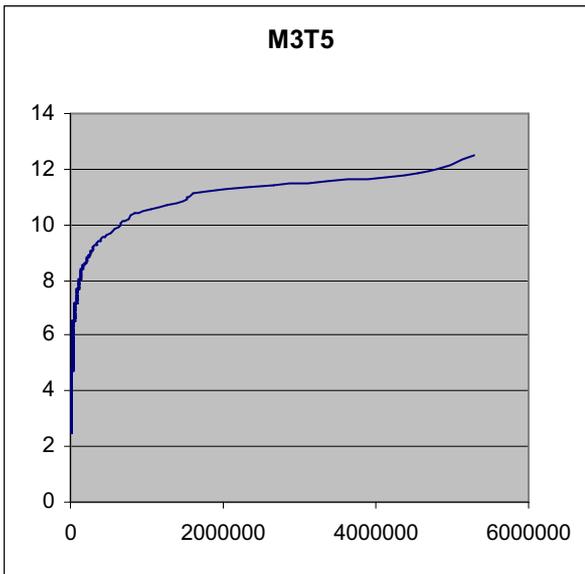
Une particularité des lois à queues très épaisses est que leurs moments ne sont pas nécessairement finis. Par exemple pour une loi de Pareto de paramètre α , les seuls moments finis sont ceux d'ordre inférieurs à $E(\alpha)$. Un critère de choix de lois peut donc être de déterminer quels sont les moments des coûts qui sont finis. Malheureusement, on ne peut pas utiliser les moments empiriques qui sont nécessairement finis. On va donc se servir du théorème suivant :

Soit $C_i, i \geq 0$ des réalisations iid d'une même variable et $U_n^r = \frac{\max_{i \leq n} C_i}{\sum_{i \leq n} C_i}$. Alors $U_n^r \rightarrow 0$ presque

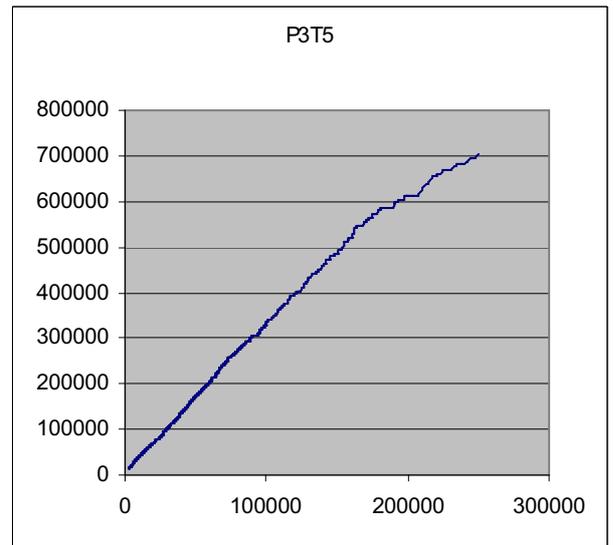
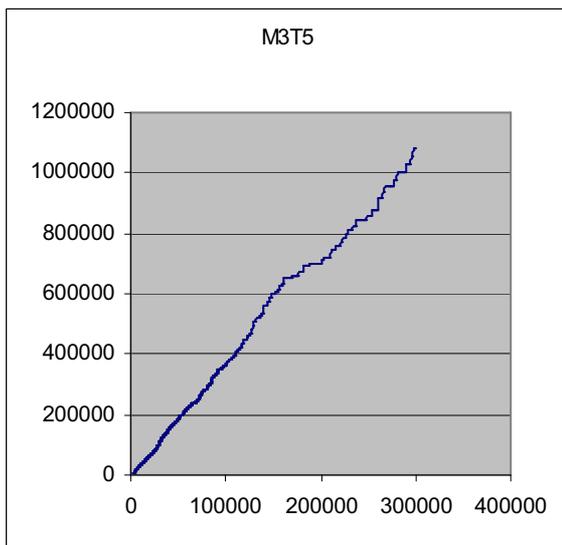
surement si et seulement si $E(C^r) < +\infty$.

On a appliqué cette méthode aux coûts en excès de 3000. On donne ci contre les graphes obtenus pour $r = 1, 2, 3$. A la vue des courbes, on peut supposer que l'espérance des coûts est finie aussi bien pour les M3T5 que pour les P3T5. Par contre pour la variance et (de façon encore plus nette) pour le moment d'ordre 3, la suite ne semble pas tendre vers 0. Ces moments seraient donc infinis. On remarque que la décroissance est plus lente pour les P3T5 que pour les M3T5.

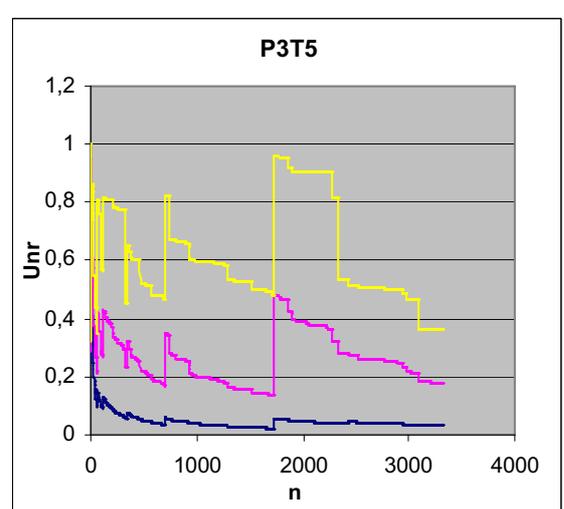
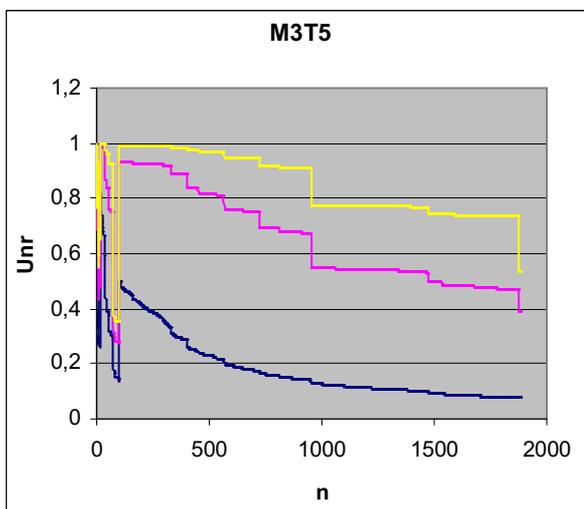
Au final, les observations semblent compatibles avec une loi de Pareto de paramètre compris entre 1 et 2. Elles indiquent aussi que l'exposant des M3T5 est supérieur à celui des P3T5.



QQ plot de la loi observée contre la loi exponentielle(1)



Espérance résiduelle en fonction du seuil



Finitude des moments (en bleu, l'espérance, en rouge la variance, en jaune μ_3)

2.2.1.2 Choix du seuil pour les sinistres graves

Les résultats ci-dessus semblent confirmer que la méthode POT s'applique aux données. Reste à choisir le seuil S le plus approprié. Il y a en effet deux dangers :

- Si on choisit un seuil S trop faible, on va prendre en compte dans l'estimation de la loi de queue des coûts moyennement élevés en plus des sinistres réellement extrêmes. En pratique, cela conduit généralement à sous estimer le risque de très gros sinistres. Or on a vu que les coûts les plus élevés jouent un rôle prépondérant.
- Si on choisit S trop élevé, la loi des coûts va être sur un grand intervalle la loi des sinistres ordinaires : on sous-estime donc les sinistres intermédiaires. De plus, on risque de ne pas avoir assez de données pour modéliser de façon robuste la loi de la queue.

Un critère souvent utilisé consiste à utiliser le graphique de l'espérance résiduelle. A partir du seuil, les observations doivent suivre approximativement une loi de Pareto, donc l'espérance excédentaire doit croître linéairement. On prend comme seuil la première valeur au de là de laquelle la croissance est linéaire. Ce critère s'applique mal ici. En effet, le graphe de $e(u)$ devient très rapidement linéaire. La courbe de l'espérance résiduelle est presque droite pour des valeurs très faibles de u . En fait, en première approximation, on pourrait approcher l'ensemble de la loi des coûts par une loi de Pareto (cela conduirait néanmoins à une grosse sous estimation des risques extrêmes). Si on utilise ce critère, le seuil serait inférieur à 1000 euros, ce qui est beaucoup trop faible (près de la moitié des sinistres seraient extrêmes, et on ne modéliserait pas finement la queue).

Une autre solution consiste à utiliser des estimateurs du paramètre α de la loi de Pareto. On ajuste la loi sur les k plus gros sinistres et on en déduit une valeur $\hat{\alpha}(k)$ du paramètre. En général, $\hat{\alpha}(k)$ dépend fortement de k . On choisit k dans un intervalle où $\hat{\alpha}$ est stable.

On a utilisé ici trois estimateurs différents de α . On va chercher un seuil T aux voisinages duquel ils soient stables et le plus cohérents possibles entre eux. On décrit les estimateurs utilisés :

◇ La pente de l'espérance résiduelle. On a vu que si C suit une loi de Pareto (m, α) , son espérance résiduelle vaut : $e(u) = \frac{\alpha}{1-\alpha}u + cste$.

On conserve donc tous les sinistres au dessus du seuil T et on effectue une régression linéaire. Le coefficient directeur a obtenu est une approximation de $\frac{\alpha}{1-\alpha}$. Finalement, on en déduit l'estimateur :

$$\hat{\alpha}_{pente}(T) = \frac{a}{a+1}.$$

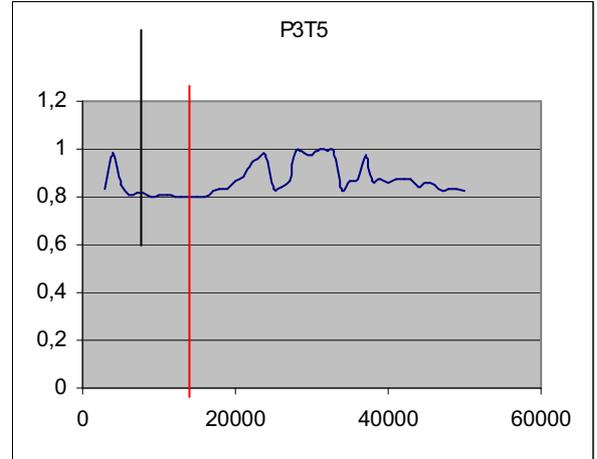
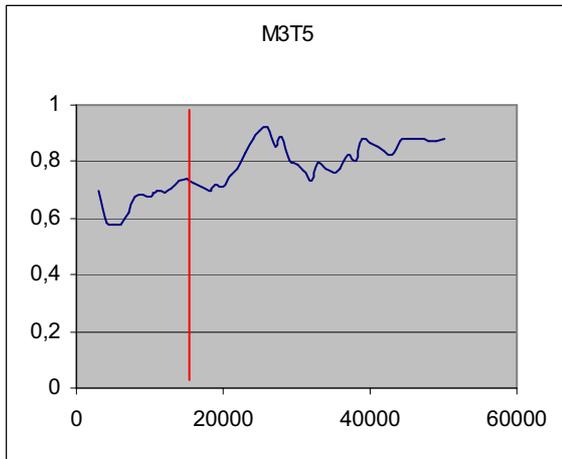
◇ L'estimateur de Hill. Cet estimateur très classique en EVT est l'excès empirique moyen des logarithmes des coûts : on note $C_{(1)}..C_{(n)}$ la liste des coûts par ordre décroissant,

$$\hat{\alpha}_{Hill}(k) = \frac{1}{k} \sum_{i < k} \ln(C_i) - \ln(C_k)$$

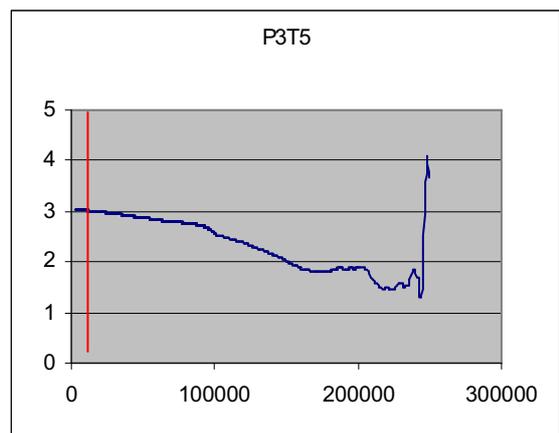
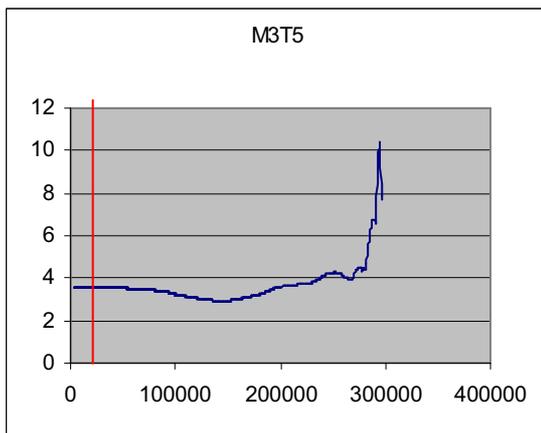
◇ L'estimateur de maximum de vraisemblance. On veut ajuster une loi de Pareto sur les coûts en excédents de T . On peut donc utiliser l'EMV de cette loi. Il est décrit au paragraphe 2.2.4.

Les courbes de ces différents estimateurs sont données ci contre, pour les sinistres M3T5 et P3T5. Au final, on a retenu un seuil de 15000 euros. Sur les graphiques ci contre, ce niveau est indiqué en rouge. Cette valeur correspond à une zone de stabilité pour la plupart des estimateurs considérés. De plus, les différentes valeurs de $\hat{\alpha}$ restent relativement cohérentes entre elles.

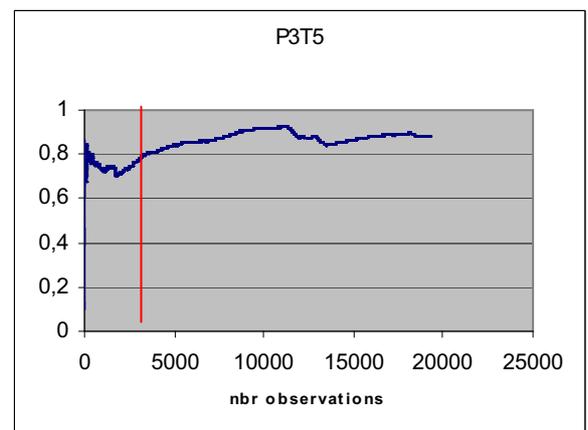
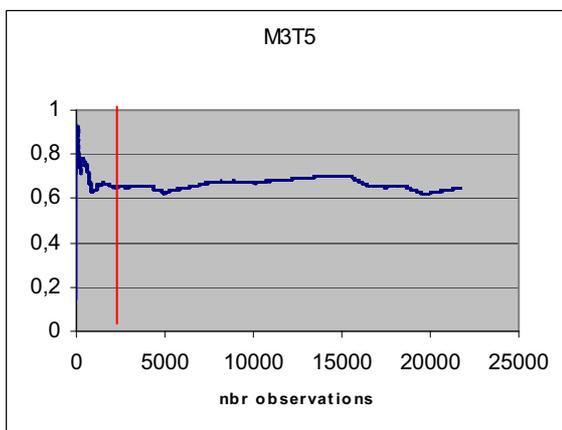
A titre indicatif, en 2007, les sinistres de plus de 15000 euros représentent environ 1% des sinistres (y compris IDA) mais 35% de la charge. Les modéliser finement est donc justifié.



Estimateurs du maximum de vraisemblance (en rouge les estimations à 15000 euros)



Pente de la droite de régression de l'excès moyen au dessus d'un seuil



Estimateur de Hill

| | EMV | Hill | Pente |
|------|------|------|-------|
| M3T5 | 0,74 | 0,77 | 0,66 |
| P3T5 | 0,8 | 0,75 | 0,79 |

Comparaison des différents estimateurs

2.2.2 Modélisation des sinistres ordinaires

2.2.2.1 Loïs étudiées et tests d'adéquation

On cherche à présent à modéliser le coût des sinistres attritionnels, c'est-à-dire les sinistres de moins de 15000 euros. On va tout d'abord donner la liste des loïs étudiées, les méthodes d'estimations de paramètres puis les outils permettant de les comparer (voire de les tester).

On s'intéresse ici aux sinistres de moins de 15000 euros. On est donc en présence d'une variable aléatoire continue bornée. Il y a deux loïs classiques pour ce type de variables :

◇ La loi uniforme $U([0; 15000])$. Cette loi ne convient clairement pas car les coûts faibles sont bien plus fréquents que les coûts élevés.

◇ Une loi dérivée de la loi beta : on peut supposer que le coût divisé par 15000 suit une loi Beta. Cette solution n'est pas non plus satisfaisante car elle sous estime la queue et qu'elle est très malaisée à manipuler numériquement.

Une autre solution consiste à utiliser une loi à support dans \mathbb{R}^+ . Dans ce cas, on cherchera seulement à avoir une bonne adéquation pour les valeurs de $[0; 15000]$ et à ce que la probabilité de dépasser 15000 soit proche de la probabilité empirique. C'est l'approche que nous avons privilégié. Les valeurs de coûts supérieurs à 15000 seront étudiées au paragraphe 2.2.5 par EVT. On donne les différentes loïs envisagées, classée par épaisseur de queue.

◇ La loi exponentielle $\mathcal{E}(\lambda)$ de densité $f(t) = \lambda \exp(-\lambda * t)$. Cette loi est à queue fine. On estime λ par $\frac{1}{\bar{C}}$. C'est à la fois l'estimateur des moments et l'EMV.

◇ La loi gamma $\gamma(n, \lambda)$ de densité $f(t) = \frac{\lambda^n}{\Gamma(n)} t^{n-1} \exp(-\lambda * t)$. Cette loi est à queue assez fine, voire intermédiaire si n est grand. L'EMV est donné par $\hat{\lambda} = \frac{\hat{n}}{\bar{C}}$ et \hat{n} solution de $\ln(u) - \psi(u) = \ln(\frac{\bar{C}}{\bar{C}^g})$ où \bar{C} est la moyenne arithmétique des coûts, \bar{C}^g la moyenne géométrique et $\psi(t) = \frac{\Gamma'(t)}{\Gamma(t)}$.

◇ La loi lognormale $LN(\mu, \sigma^2)$ de densité $f(t) = \frac{1}{\sqrt{2\pi\sigma}} \exp(-\frac{1}{2\sigma^2}(\ln(t) - \mu)^2)$. Cette loi est à queue intermédiaire. Les EMV sont donnés par la moyenne empirique et la variance empirique des logarithmes des coûts. Ces estimateurs sont biaisés mais asymptotiquement sans biais.

◇ La loi log gamma $\log\gamma(n, \lambda)$. $X \sim \log\gamma(n, \lambda)$ ssi $\ln X \sim \gamma(n, \lambda)$. Cette loi est à queue épaisse. Ces moments sont finis jusqu'à l'ordre partie entière de λ . Les EMV sont les mêmes que ceux de la loi gamma en remplaçant les coûts par leurs logarithme.

◇ La loi de Pareto « américaine » $P(m, \alpha)$ de fonction de survie $\bar{F}(t) = (\frac{m}{m+t})^\alpha$. Cette loi est aussi à queue épaisse. Les moments sont finis jusqu'à l'ordre partie entière de α .

L'EMV vérifie les équations : $moyenne(\frac{1}{1 + \frac{C}{\hat{m}}}) * (1 + moyenne(\ln(1 + \frac{C}{\hat{m}}))) = 1$ Et $\hat{\alpha} = (1 +$

$moyenne(\ln(1 + \frac{C}{\hat{m}})))^{-1}$.

Une fois que l'on dispose de loïs candidate pour modéliser les coûts, reste à choisir la meilleure.

Dans le cas des variables continues, l'outil classique est la statistique de Kolmogorov-Smirnov.

On note F_{emp} la fonction de répartition empirique donnée par $F_{emp}(t) = \frac{\text{nombre de sinistres inférieurs à } t}{\text{nombre de sinistres}}$ et $F(t)$ la fdr de la loi à tester. La distance de Kolmogorov est donnée par $d(F_{emp}; F) = D_n = \max_{t \in \mathbb{R}^+} |F_{emp}(t) - F(t)|$. Plus cette distance est faible et plus la loi F est compatible avec les observations.

D'après le théorème de Kolmogorov, si les coûts ont pour fdr F , la statistique $\sqrt{n}D_n$ tend en loi vers une loi canonique. On peut tester l'hypothèse que la loi des coûts a pour fonction de répartition F . On rejettera cette hypothèse avec probabilité α si $\sqrt{n}D_n$ dépasse le quantile à $1 - \alpha$ de la loi de Kolmogorov.

En pratique, cela ne fonctionnera pas sur nos données. En effet, dès que le nombre d'observations dépassent quelques milliers, il est très rare que le test conclue positivement. Dans notre cas aussi bien pour les M3T5 que pour les P3T5, nous disposons de plusieurs centaines de milliers de sinistres. Modéliser autant de valeurs avec seulement deux paramètres (le maximum possibles avec nos lois) est assez irréaliste.

La statistique D_n est donc utilisée comme outils de comparaison entre les modèles : une loi s'ajuste mieux qu'une autre si sa valeur de D_n est plus faible. De plus, comme on ne s'intéresse ici qu'à la partie de la courbe en dessous de 15000 euros, on utilisera plutôt $\tilde{D}_n = \max_{t \in [0; 15000]} |F_{emp}(t) - F(t)|$.

Pour qu'une loi minimise la statistique D_n , elle doit être relativement proche de la loi observée en tout point. Pourtant, on peut penser qu'une loi qui serait la plupart du temps plus proche mais qui s'éloignerait en quelques points serait meilleure. On introduit un autre critère de jugement pour tenir compte de cela. On va utiliser un χ^2 d'indépendance du type de celui employé pour les lois de fréquences. On a divisé l'intervalle $[0; 15000]$ en 20 classes $\mathcal{C}_1.. \mathcal{C}_{20}$ de façon à ce que le nombre d'observations dans chaque classe soit proche (on a employé des découpages différents pour les M3T5 et les P3T5).

On note n_i^{obs} et n_i^{th} le nombre d'observations pour la loi empirique et pour la loi à tester. On pose $\chi = \sum_i \frac{(p_i^{th} - p_i^{obs})^2}{p_i^{th}}$. Si la fdr des coûts est F , alors χ tend vers $\chi^2(17)$.

En pratique, vu le nombre d'observations, on n'a pas utilisé χ comme outils de test mais on s'en sert pour comparer les modèles : un modèle sera meilleur qu'un autre si son χ est plus faible.

2.2.2.2 Loi des sinistres ordinaires

On applique les méthodes décrites au paravent au portefeuille. On a constitué une base regroupant tous les sinistres depuis 2005. On a supprimé les sinistres de coûts nuls (sans suite) et les sinistres IDA forfaitaires (par définition, leur coût vaut 1204 euros et un aussi grand nombre de valeurs toutes identiques rend quasi impossible une bonne adéquation des lois continues). La base comporte 460000 sinistres, 57% en M3T5 et 43% en P3T5. Pour chacune des familles de lois décrites ci-dessus, on a estimé les paramètres et calculé la statistique de Kolmogorov et celle du χ^2 . On donne les résultats.

On constate que les lois à queues fines (exponentielles et gamma) s'ajustent très mal. Les autres lois donnent des résultats assez proches. Dans le cas des M3T5, la loi log normale est clairement celle qui s'ajuste le mieux. Pour les P3T5, les lois à queues très épaisses s'ajustent légèrement mieux. Néanmoins, on préfère retenir la même famille de loi dans les deux cas pour des raisons de cohérence.

Un autre point très important est de comparer la probabilité pour un coût d'être inférieur à 15000 euros, pour les lois candidates avec les probabilités effectivement constatées. Pour bien ajuster notre loi mélange, ces valeurs doivent être proches. A nouveau, la loi log normale est celle qui s'approche le plus des vraies valeurs (les lois à queues fines sous-estiment ce risque, celles à queues épaisses le surestiment).

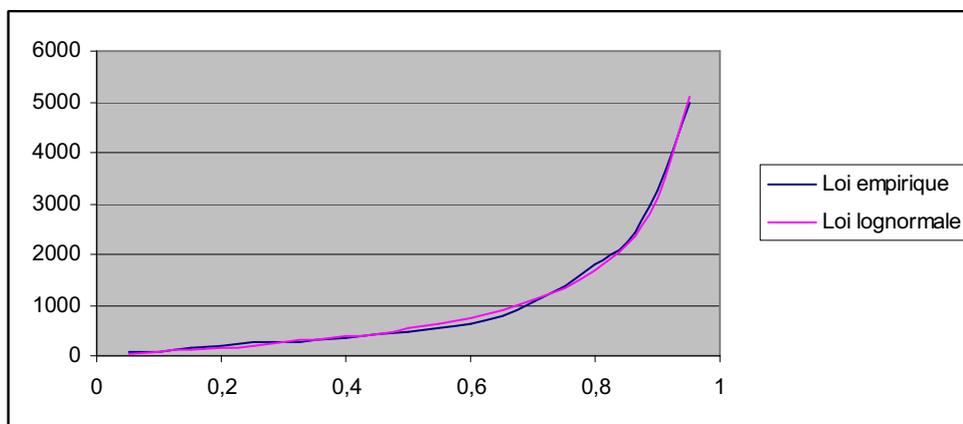
On donne enfin le graphe de la loi empirique et de la loi log normale proposée, pour les M3T5 et les P3T5. Dans les deux cas, l'ajustement est bon. On retient donc le modèle log normal pour modéliser les coûts des sinistres ordinaires.

| | Dn | Khi 2 | Probabilité de dépasser 15000 |
|-------------------|------|-------|-------------------------------|
| Loi exponentielle | 1535 | 1870 | 3,E-06 |
| Loi gamma | 977 | 1050 | 6,E-05 |
| Loi log normale | 178 | 106 | 0,75% |
| Loi log gamma | 337 | 112 | 1,20% |
| Loi de Pareto | 198 | 198 | 0,94% |
| Vraie valeur | | | 0,80% |

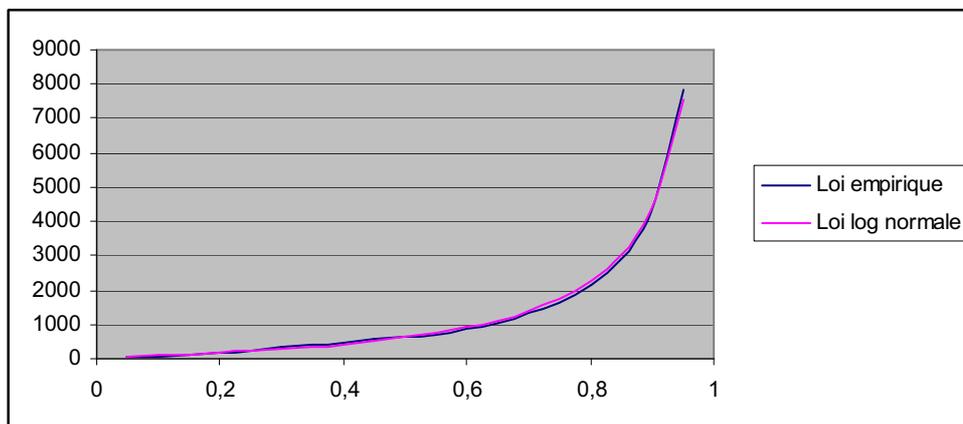
Adéquations des lois aux sinistres M3T5

| | Dn | Khi 2 | Probabilité de dépasser 15000 |
|-------------------|------|-------|-------------------------------|
| Loi exponentielle | 3064 | 4217 | 8,E-05 |
| Loi gamma | 1880 | 1617 | 0,17% |
| Loi log normale | 277 | 79 | 1,80% |
| Loi log gamma | 67 | 63 | 2,30% |
| Loi de Pareto | 82 | 57 | 2,20% |
| Vraie valeur | | | 0,80% |

Adéquation des lois aux sinistres P3T5



Loi empirique et loi log normale pour les sinistres M3T5



Loi empirique et loi log normale pour les sinistres P3T5

2.2.3 Modélisation des sinistres graves

2.2.3.1 Rappel de théorie des valeurs extrêmes

La théorie de valeurs extrêmes ou EVT a été développée à partir des années 20 dans le but d'étudier le comportement de la queue des lois de probabilité. Ce problème est à priori complexe car il s'agit d'étudier des phénomènes rares donc pour lesquels on dispose de peu d'information.

On considère une suite (C_n) de variables aléatoires réelles iid et on s'intéresse à la loi de $M_n = \max_{i \leq n} C_i$. On sait que si F est la fonction de répartition de C , M_n admet comme fdr F^n . En général, pour les grandes valeurs de n , cette expression est néanmoins peu manipulable. On cherche donc une loi canonique pour le comportement du maximum.

Pour illustrer cette idée, considérons le cas de la variable $S_n = \sum_{i \leq n} S_i$. On sait que la fdr de S_n est F^{*n} mais cette quantité est peu aisée à calculer en pratique. Il est souvent plus commode d'utiliser le fait que sous certaines hypothèses ($V(C)$ finie), la suite des $\frac{S_n - nE(S)}{\sqrt{n}\sigma(n)}$ tend vers une loi canonique : la loi $N(0; 1)$. On va procéder de même pour M_n . On cherche une éventuelle loi continue H et des suites (a_n) et (b_n) telles que $\frac{M_n - a_n}{b_n}$ tende en loi vers H . Dans ce cas, on dit que la variable C est dans le domaine d'attraction maximale (DAM) de H .

En pratique, H ne peut pas être n'importe quelle loi. Le théorème de Fisher Tippet Indique qu'il n'y a que trois lois possibles pour le comportement des maxima.

Théorème : Fisher Tippet

Si une suite (C_n) de variables iid est dans le DAM d'une loi H non dégénérée, alors H est du même type qu'une des lois suivantes :

- 1) $G_{\xi}^{-}(t) = \exp(-(1 + \xi t)^{-1/\xi})$ pour $\xi < 0$ et $t \leq -\frac{1}{\xi}$ Domaine de Weibull
- 2) $G_0(t) = \exp(-\exp(-t))$ pour $t \in \mathbb{R}$ Domaine de Gumbel
- 3) $G_{\xi}^{-}(t) = \exp(-(1 + \xi t)^{-1/\xi})$ pour $\xi > 0$ et $t \geq -\frac{1}{\xi}$ Domaine de Fréchet

On rappelle que deux variables X et Y sont de même type s'il existe a et b tels que $aX + b$ et Y aient même loi.

Le premier domaine (domaine de Weibull) regroupe les lois bornées à droite comme les lois uniformes ou beta.

Le second domaine (Gumbel) regroupe des lois qui ne sont pas majorées mais dont la queue n'est pas épaisse. A titre d'exemple, on a les lois normales, exponentielles, gammas, Weibull, log normales. La loi log normale, bien qu'à décroissance plus lente que les précédentes n'est donc pas une loi à queue épaisse. Ces lois sont à décroissance exponentielle et admettent des moments de tous ordres.

Le troisième domaine (Fréchet) regroupe des lois à queue plus épaisses, donc généralement celles qui interviennent dans les modélisations de sinistres graves. Parmi elles, les lois Pareto, Cauchy et log gamma. En fait, il s'agit des lois décroissant à l'infini comme une puissance : une fonction est dans le DAM de Fréchet ssi $1 - F(t) = t^{-\alpha}V(t)$ avec V à variation lente à l'infini. On remarque de plus que les moments d'ordre supérieurs à α sont infinis.

Le théorème de Fisher Tippet fournit le comportement asymptotique de la loi du maximum des sinistres. Vu l'épaisseur de queue, on s'attend à une loi de type Fréchet. Néanmoins, nous voulons modéliser tous les sinistres graves (i.e. : de plus de 15000 euros), pas seulement le plus élevé. Il faut donc approximer la fdr de répartition du coût au-delà de 15000. C'est l'objet du

théorème de Pickands, Balkema, DeHaan.

Dans la suite, on note F la fonction de répartition d'une variable X appartenant au DAM de Fréchet. Si u est un réel positif, on note F_u la fdr des excès de X par rapport à u : $F_u(t) = P(X \leq t | X > u)$. Le théorème suivant permet d'approximer F_u pour u par une loi de Pareto généralisée (loi GPD). On rappelle sa définition :

Une variable suit une loi de Pareto généralisée ou GPD (generalised Pareto distribution) de paramètres (ξ, τ) ssi sa fonction de survie est $\bar{G}_{\tau\xi}(t) = (1 + \tau \times t)^{-1/\xi}$, i.e. : ssi elle suit une loi de Pareto américaine de paramètre $(\frac{1}{\tau}, \frac{1}{\xi})$.

Théorème : Pickands, Balkema, DeHaan.

Soit F une fdr appartenant au DAM de Fréchet de paramètre ξ . Il existe $\tau : \mathbb{R} \rightarrow \mathbb{R}^+$ telle que : $\lim_{u \rightarrow +\infty} \sup_t |\bar{F}_u(t) - \bar{G}_{\xi \tau(u)}(t)| = 0$

Concrètement, on peut donc approximer la loi des excès par rapport à u par une loi GPD si u est assez élevé. Reste ensuite à évaluer les paramètres de cette loi. On a présenté au 2.2.1 trois estimateurs de ξ . On utilisera dans la suite de l'étude les estimateurs du maximum de vraisemblance qui ont été donnés au 2.2.2.

On modélisera les coûts de sinistres par la méthode POT ou « Peak over Threshold ». Les sinistres de montants inférieurs au seuil critique sont modélisés séparément (dans notre cas par des lois log normales). Pour les sinistres au dessus du seuil, la loi de $C - u$ est approximée par une loi GPD. Concrètement, la fdr des coûts F est donc égale à $F_{\text{ordinaires}}(t)$ si $t \leq 15000$ et vaut $F_{\text{ordinaires}} * F_{\text{grave}}(t - 15000)$ si $t \geq 15000$.

2.2.3.2 Loi de coût des sinistres extrêmes

On a appliqué au portefeuille les méthodes décrite au paravent. On donne ici les résultats obtenus pour les sinistres des flottes M3T5. Depuis 2005, il y a eu 1888 sinistres supérieurs à 15000 euros. La plupart dépasse de peu ce montant, la médiane est de 25000 euros (donc 10000 euros de dépassement) mais certains peuvent excéder largement le seuil. A titre d'exemple, 8 sinistres dépassent le million d'euros.

La méthode POT indique un paramètre de queue de 0.74 (pour mémoire, l'estimateur de Hill est de 0.77). Il s'agit donc d'une loi à queue très épaisse : l'espérance est finie mais la variance est infinie, ce qui correspond bien aux suppositions faites au paragraphe 2.2.1. Le paramètre de forme τ vaut 12131. On remarquera que les lois GPD et empiriques ont des espérances très proches : 34986 contre 34156.

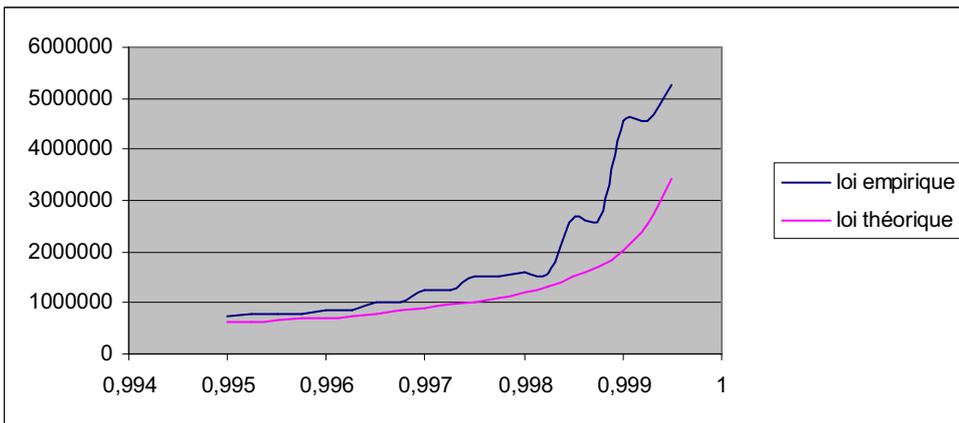
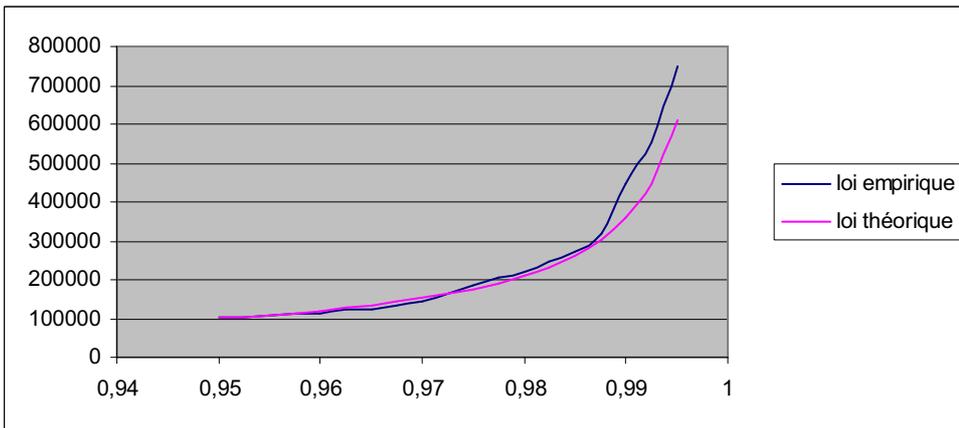
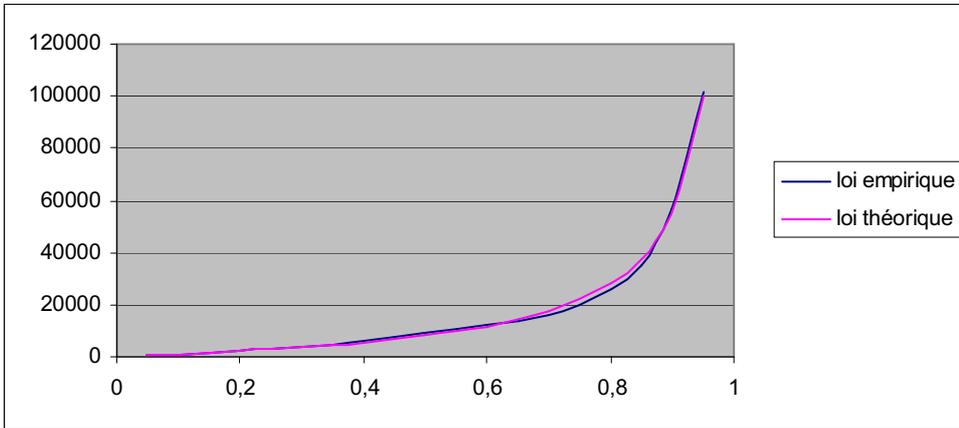
On donne ci contre les graphes des lois empiriques et GPD. Le premier graphe traite les premiers 95% de la courbe soit jusqu'à 100000 euros de dépassement (i.e. : 115000 euros au total). Sur cet intervalle, l'adéquation est bonne, les deux lois ne s'écartent jamais de plus de 2000 euros. En pourcentage de la valeur de la loi GPD, l'écart maximum est de 13%. De plus, la loi GPD n'est ni significativement au dessus, ni significativement en dessous de la loi observée.

Le deuxième graphique donne les derniers 5% de la queue, soit les sinistres de plus de 120000 euros. L'adéquation est beaucoup plus discutable. On remarque qu'à partir de 150000 euros, la loi GPD sous estime systématiquement les coûts, l'erreur devenant très nette à partir de 300000 euros. Enfin, le troisième graphe représente les derniers 10% du précédent (à partir de 800000 euros). L'adéquation est encore plus mauvaise. Il ne faut cependant pas y attacher trop d'importance car il ne prend en compte que les 11 plus gros sinistres.

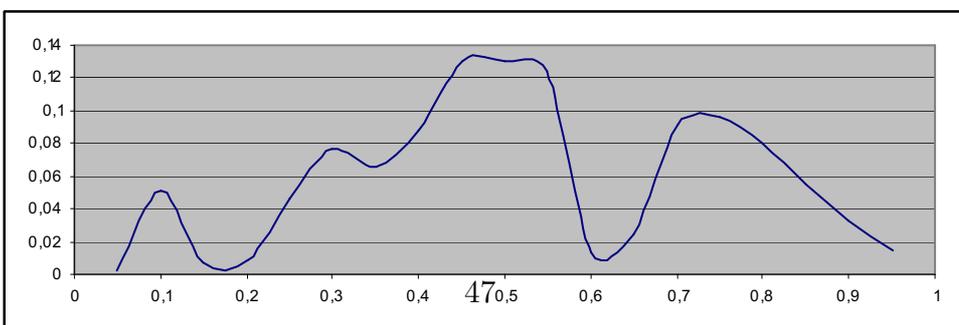
Au total, l'adéquation est globalement assez bonne. Les deux courbes restent proches jusqu'à 300000 euros (soit en dehors des 31 plus gros sinistres). Le graphique de l'écart montre que les valeurs diffèrent généralement de moins de 5%. Enfin, le fait que les lois prédites et observées aient des espérances proches indique qu'il n'y a globalement ni surestimation ni sous-estimation du coût.

| Type de flotte | Nombre d'observations | ksi | tau | Moyenne empirique | Moyenne théorique | Variance empirique | Variance théorique |
|----------------|-----------------------|------|-------|-------------------|-------------------|--------------------|--------------------|
| M3T5 | 1888 | 0,74 | 12131 | 34156 | 34986 | 2,9E+12 | infinie |

Résultat fournis par la méthode POT



Adéquation de la loi GPD à la loi empirique



Ecart entre les lois empiriques et GPD en pourcentage de la valeur prédite

On étudie à présent la loi de queue pour les sinistres des flottes P3T5. Il y a en tout 3328 sinistres de plus de 15000 euros, soit 1.5 fois plus que pour les M3T5. Pour mémoire, il y avait plus de sinistres en M3T5. Cela traduit bien le fait que les flottes P3T5 sont plus sujettes aux sinistres extrêmes que les M3T5. La disparité entre les montants est également très importante. La valeur médiane est de 31000 euros, la plus grande de 3.5 millions (14 sinistres dépassent le million).

Le paramètre de queue vaut 0.8, ce qui correspond bien à ce que nous attendions : espérance finie, variance infinie et queue plus épaisse que pour les M3T5. Le paramètre de forme vaut 14573, ce qui est plus élevé que celui des M3T5 (à nouveau, queue plus épaisse). La moyenne empirique des excès à 15000 euros est de 41488, l'espérance de la loi GPD est de 45223. Il y a donc une erreur de l'ordre de 10%, ce qui n'est pas négligeable. Cela dit, cette erreur tend à nous faire surestimer le coût des sinistres, ce qui est préférable du point de vue de la prudence.

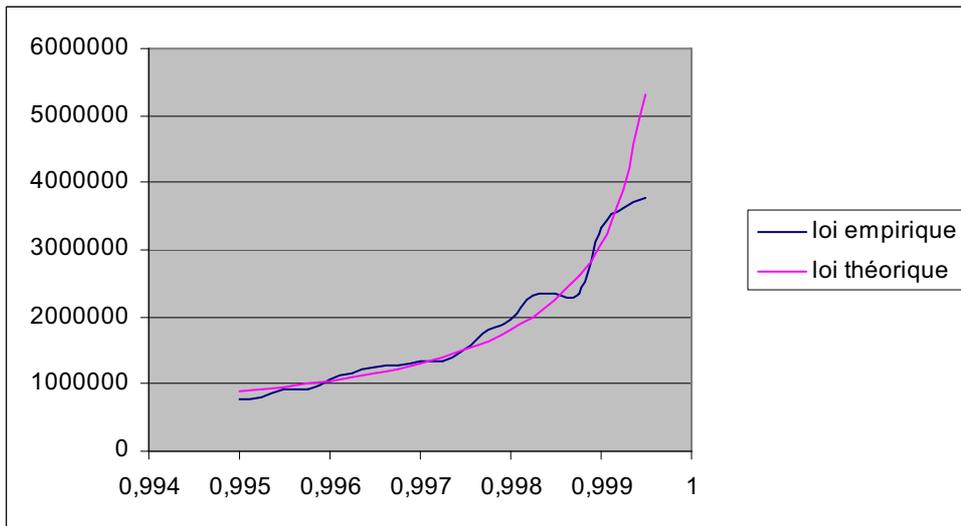
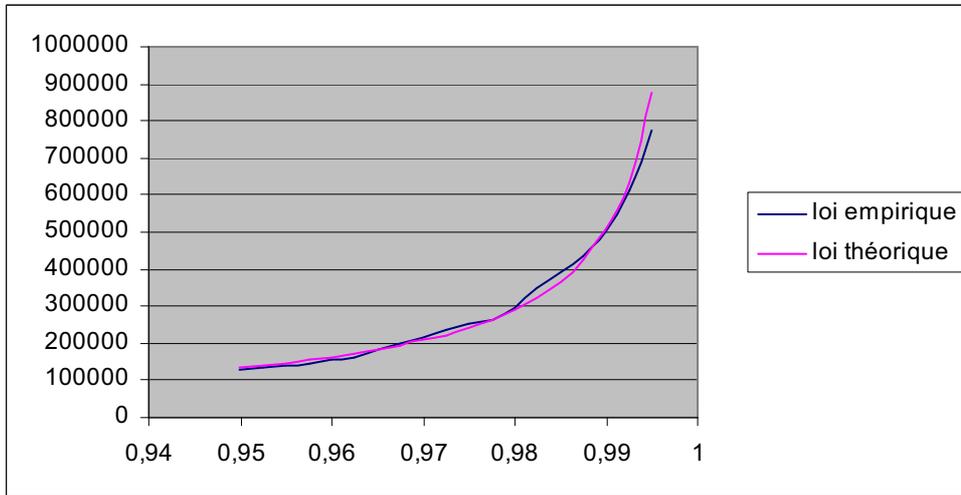
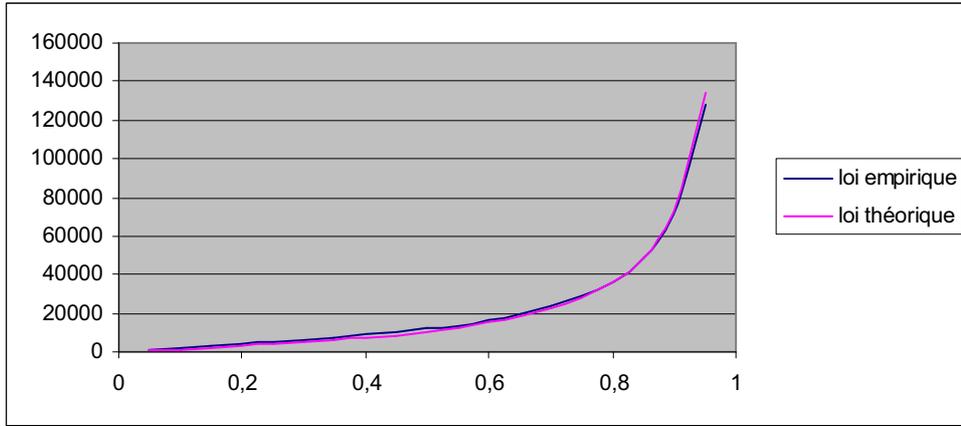
Le premier graphe compare les lois empiriques et GPD sur 95% de la distribution (jusqu'à 145000 euros de coût total). La valeur GPD sous estime légèrement le coût jusqu'à 30000 euros de dépassement puis les surestime. Au total, l'ajustement est assez bon. L'écart maximum est de 2100 euros, et l'écart maximum en pourcentage de 12%.

Le deuxième graphe donne les 5 derniers pourcent de la distribution, soit les 160 plus gros sinistres. L'adéquation paraît acceptable, la valeur observée ne diffère jamais de plus de 9% de la valeur prédite par la loi GPD. Contrairement au cas M3T5, il ne semble pas y avoir de « décrochage » entre les deux lois même pour les très grosses valeurs. Le troisième graphique représente les derniers 10% du second soit les sinistres de plus de 800000 euros. La loi empirique est très irrégulière du fait du peu d'observations (15 en tout). L'adéquation est donc nécessairement mauvaise. Elle paraît néanmoins valable.

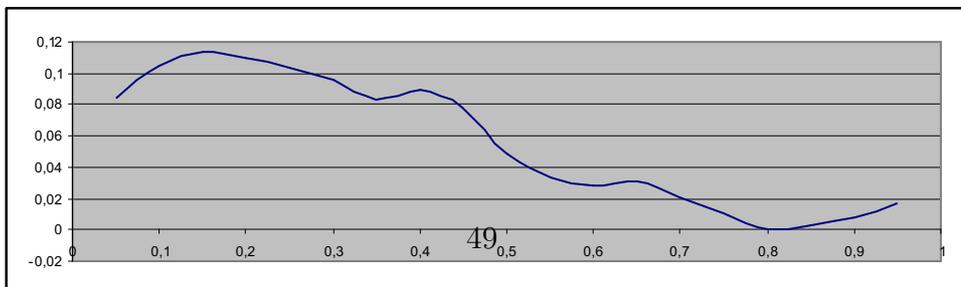
Au final, dans le cas des P3T5, la loi GPD paraît être une bonne modélisation de la queue des sinistres. L'adéquation est assez bonne et il n'y a globalement ni surestimation ni sous-estimation. Les écarts sont la plus part du temps faibles, sauf au début de la courbe où ils dépassent 10%. Néanmoins cette zone correspond à des coûts moindres donc est moins dangereuse.

| Type de flotte | Nombre d'observations | ksi | tau | Moyenne empirique | Moyenne théorique | Variance empirique | Variance théorique |
|----------------|-----------------------|-----|-------|-------------------|-------------------|--------------------|--------------------|
| P3T5 | 3328 | 0,8 | 14573 | 41488 | 45223 | 3,6E+12 | infinie |

Résultats fournis par la méthode POT.



Adéquation de la loi GPD à la loi empirique



Écart entre les lois GPD et empiriques en pourcentage de la valeur prédite

2.2.4 Prise en compte de l'usage

2.2.4.1 Importance de l'usage

On a cherché à différencier les coûts de sinistres selon la division M3T5/P3T5. Il s'agit de la subdivision la plus classique en automobile professionnelle. Néanmoins, il ya une autre segmentation très utilisée : celle par l'usage. On rappelle que les flottes sont divisées en TPM (transport public de marchandises), TPV (transport public de voyageurs) et AUT (autres usages, essentiellement transport pour compte propre et engins de chantiers).

On donne ci contre les lois de coûts pour ces différents usages. Comme toujours, on constate une part prédominante des gros sinistres avec un écart type très élevé, et une moyenne largement supérieure à la médiane (en fait supérieure au quantile à 75%). On peut faire plusieurs constatations

- Les flottes d'usage AUT ont une loi de sinistre assez proches de la loi moyenne du portefeuille. Les quantiles des coûts sont inférieurs a ceux du portefeuille de 5 à 10 %. Pour les sinistres les plus élevés, l'écart devient plus net et atteint les 20% pour les sinistres de plus de 100000 euros. Globalement, les flottes AUT ont donc des sinistres attritionnels légèrement moins élevé que la moyenne et des sinistres extrêmes nettement moins coûteux. On remarque d'ailleurs que le coût moyen est inférieur de 14% à celui du portefeuille. Il faut néanmoins se méfier. En effet, la classe AUT est très vaste (82% des contrats) et très hétérogène. Parmi les différents types de contrats, citons :

- les transports pour propre compte (TPC), qui sont des véhicules servant à transporter l'assuré et des marchandises lui appartenant. Ces contrats sont les plus nombreux, composés essentiellement de véhicules M3T5 (type camionnette ou voiture particulière) et ont des sinistres peu élevés (véhicules de faibles valeurs et de faible poids).

- les deux roues, qui ont des sinistres nettement inférieurs à ceux du portefeuille.

- les engins de chantiers. Ce sont des véhicules lourds (presque toujours P3T5) et à forte valeur marchande. Leurs coûts de sinistres sont élevés. Par contre, ces types de véhicules ont une fréquence de sinistres très faible.

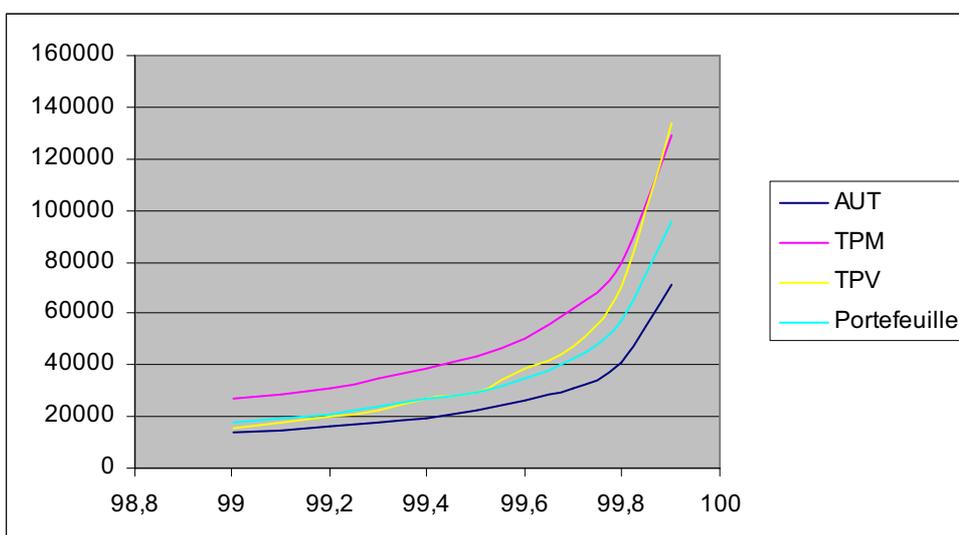
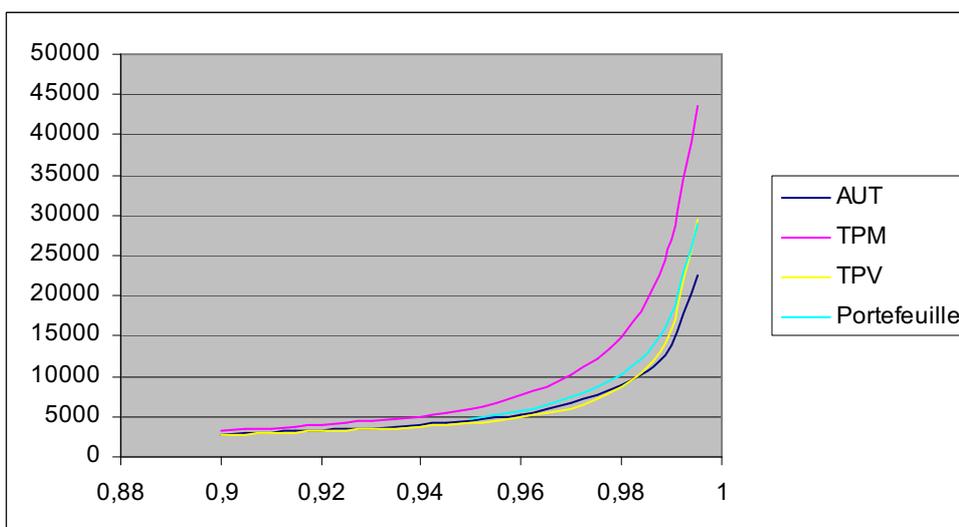
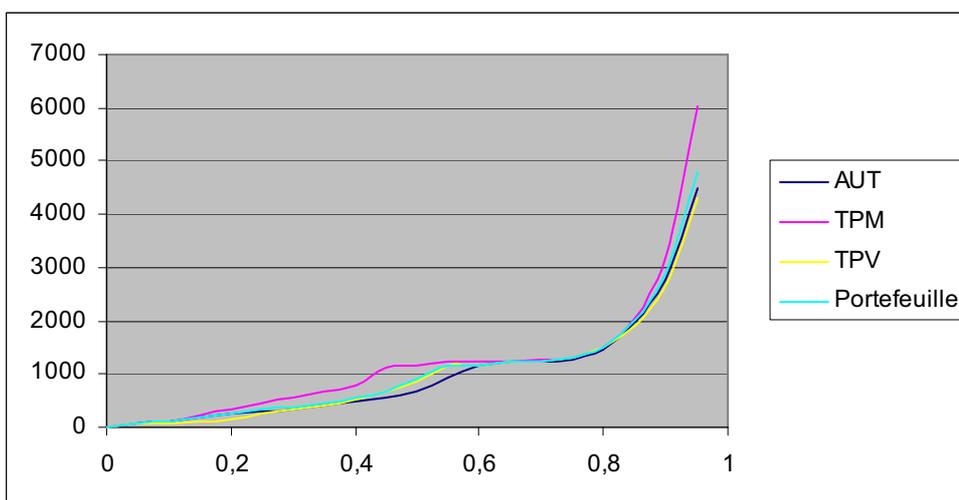
- les engins agricoles. Ces véhicules sont souvent P3T5, et peuvent avoir une forte valeur marchande (moissonneuses..). Ces véhicules ont des sinistres coûteux mais assez peu fréquents (faible vitesse, circulation peu dense en milieu rural).

- Les flottes TPM ont des sinistres largement plus élevés que la moyenne, aussi bien dans le cas ordinaire qu'exceptionnel. Ces véhicules sont très majoritairement des P3T5 (typiquement des poids lourds), et font partie de grosses flottes. Ces flottes sont à surveiller de près car elles sont beaucoup plus dangereuses que la moyenne. De plus, leur S/C est généralement médiocre.

- Les flottes TPV ont des sinistres en général raisonnables. Les quantiles des coûts sont inférieurs à ceux du portefeuille jusqu'à 99%. Ils sont cela dit légèrement supérieurs à ceux de la classe AUT. Par contre, la queue de sinistre est particulièrement lourde. A partir de 0.99 les coûts TPV dépassent ceux du portefeuille, et à partir de 0.998 ceux des TPM. En fait, ces sinistres particulièrement coûteux correspondent aux gros sinistres corporels (accidents de bus, d'autocars..). A titre d'illustration, les TPV ne représentent que 5% des contrats mais 12% des sinistres de plus de 100000 euros et 14% des plus de 500000. On remarque d'ailleurs que les coûts de sinistres sont beaucoup plus volatils que pour les autres usages (écart type de 30000 contre 20000). En somme, les flottes TPV sont globalement sûres mais avec un risque plus fort de sinistralité extrême.

| | Cout moyen | Cout médian | Quantile à 75% | Ecart type | Proportion du portefeuille |
|--------------|------------|-------------|----------------|------------|----------------------------|
| Portefeuille | 1885 | 890 | 1300 | 20810 | 100% |
| AUT | 1651 | 689 | 1283 | 19192 | 82% |
| TPM | 2362 | 1155 | 1328 | 20322 | 13% |
| TPV | 2016 | 874 | 1373 | 31829 | 5% |

Coûts des sinistres en fonction de l'usage



Loi du coût de sinistre pour les différents usages de véhicules

2.2.4.2 Loi des coûts par usage

On va maintenant modéliser les coûts de sinistres pour les différents usages. On utilise la même méthodologie que précédemment. On commence donc par chercher un seuil S au-delà duquel les sinistres seront jugés graves. Ensuite, on modélisera les sinistres attritionnels à l'aide des divers lois continues présentées en 2.2.3 et on choisira la meilleure à l'aide des statistiques du khi 2 et de Kolmogorov sur $[0; S]$. Enfin, les coûts en excès de S seront modélisés par une loi GPD.

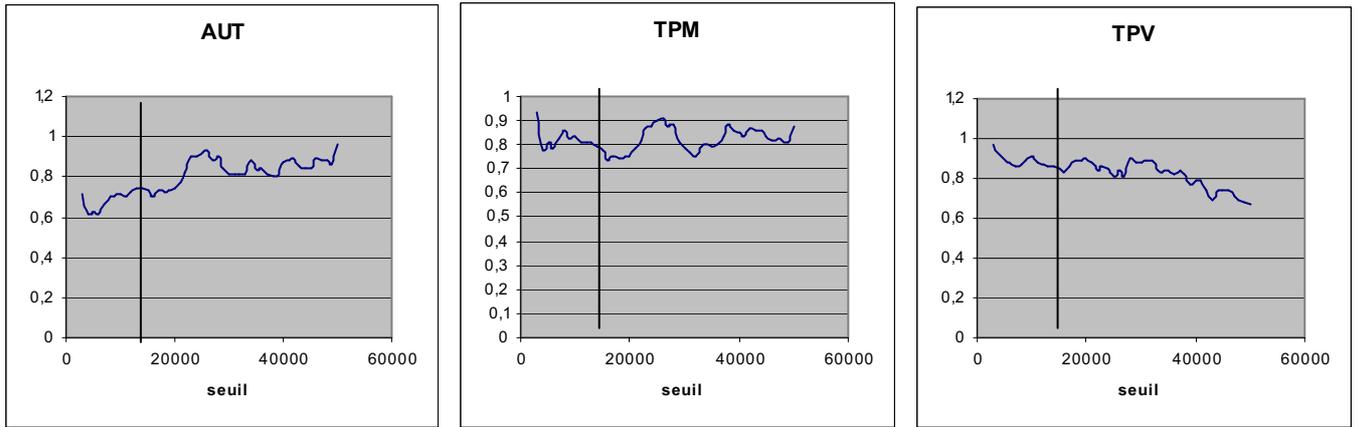
Les lois des coûts pour les différents usages, présentent les mêmes caractéristiques que précédemment : elles sont à queues très épaisses (les QQ plots contre la loi exponentielle sont nettement concaves). Les excès résiduels sont linéaires même pour des faibles valeurs de coûts. On ne donne pas ici ces graphes mais ils sont du même type que ceux obtenus pour les M3T5 et les P3T5. La méthodologie POT paraît donc pouvoir s'appliquer.

On va à présent déterminer le seuil S à retenir. A nouveau, comme les excès moyens sont linéaires même pour des coûts très faibles, on ne peut pas s'en servir pour choisir S . On va donc étudier trois estimateurs du paramètre α et choisir un seuil au voisinage duquel ils sont stables et cohérents entre eux. On utilise les trois mêmes que pour la modélisation par le type de véhicule : EMV de la loi de Pareto, estimateur de Hill et estimateur de la pente de l'excès moyen. Les graphes sont donnés ci-contre.

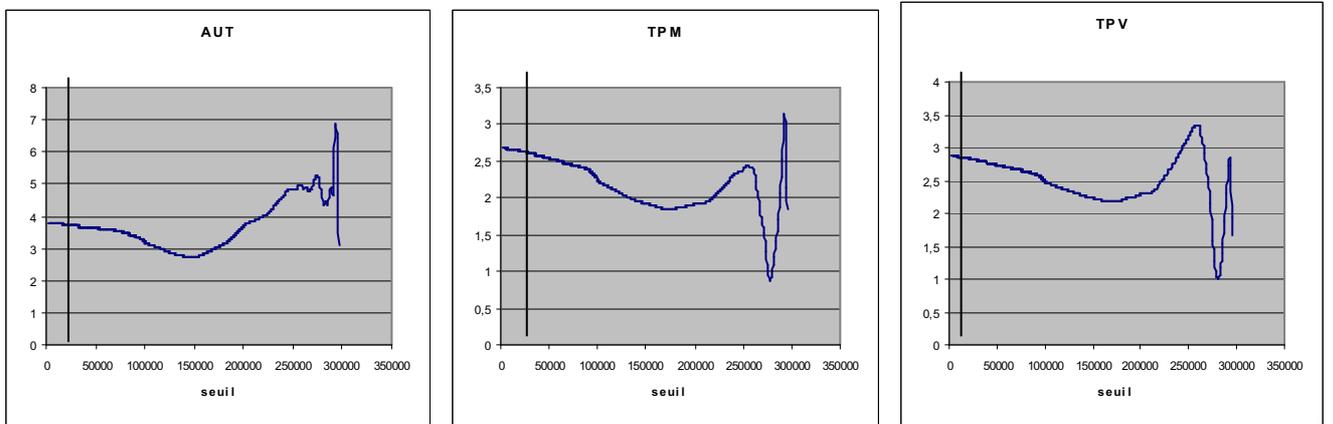
Au final, on a à nouveau choisi un seuil à 15000 euros. Celui-ci présente plusieurs avantages :

- ◇ les estimateurs sont pour la plupart assez stables au voisinage de 15000 euros.
- ◇ les valeurs obtenues ne diffèrent pas trop l'une de l'autre (sauf l'estimateur de la pente pour les TPV).
- ◇ le fait de choisir le même seuil que dans le modèle précédent permettra de les comparer plus efficacement.

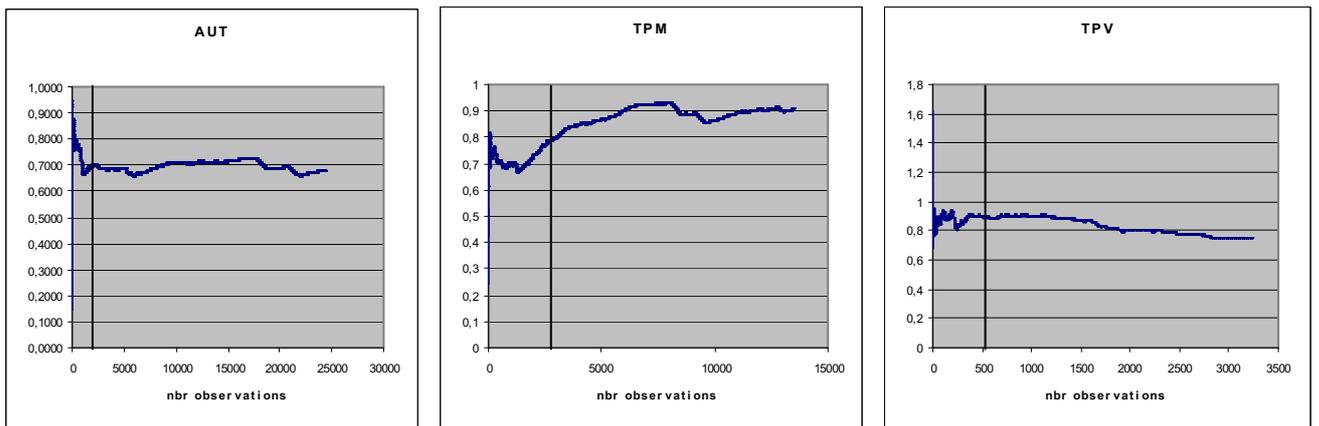
On donne aussi le nombre de sinistres dépassant 15000 euros pour chaque usage. On remarque que les TPM compte légèrement plus d'observations que les AUT alors qu'il y a cinq fois moins de contrats. On voit à nouveau que ces contrats sont plus dangereux que la moyenne. Les TPV comptent très peu d'observations, mais proportionnellement beaucoup plus que les AUT et ces sinistres sont plus graves, ce qui se traduit par des estimations de α sensiblement plus élevées.



Estimateurs du maximum de vraisemblance d'alpha en fonction du seuil



Pente de la droite de régression des excès moyens en fonction du seuil



Estimateur de Hill d'alpha en fonction du nombre d'observations

| | Hill | Pente | EMV | Nbr observations |
|-----|------|-------|------|------------------|
| AUT | 0,69 | 0,78 | 0,74 | 2322 |
| TPM | 0,77 | 0,72 | 0,76 | 2477 |
| TPV | 0,9 | 0,74 | 0,85 | 414 |

Estimateurs obtenus pour un seuil à 15000 euros

- On s'intéresse tout d'abord au cas des flottes AUT. En ce qui concerne les sinistres attritionnels, deux lois se détachent : les lois log normales et log gamma. La loi de Pareto a une adéquation moins bonne. On remarque toutefois que les valeurs de D_n et χ^2 restent élevées. De plus, la probabilité pour un sinistre d'être grave est assez nettement sous estimée. Au final, on privilégie la loi log normale car c'est celle pour laquelle l'erreur sur ce paramètre est la plus faible (même si l'adéquation de la loi log gamma est légèrement meilleure). On a tracé les lois empiriques et théoriques. Les deux semblent proches.

Pour ce qui concerne les sinistres graves, on va tout d'abord étudier la finitude des moments de la loi des coûts. On donne ci contre le test graphique. On rappelle qu'un moment fini se traduit par une courbe tendant vers 0. Il semble donc que l'espérance des coûts soit finie. Pour la variance et le moment d'ordre trois, c'est moins net. Les courbes semblent décroissantes, mais au bout de 2500 observations, elles valent respectivement 0.35 et 0.43. Il semble donc que ces moments soient infinis.

L'EMV de la loi GPD est de 0.735, ce qui confirmerait les observations ci-dessus. On a donné les graphiques des lois empiriques et GPD. L'ajustement paraît satisfaisant pour les sinistres de moins de 165000 euros (i.e. : 150000 euros de dépassement du seuil). Au-delà, la loi sous estime systématiquement les coûts. Par ailleurs, l'espérance de notre loi s'éloigne de façon importante de la valeur observée. Elle la surestime, ce qui est prudent, mais l'écart paraît très élevé.

- On veut à présent modéliser les coûts des sinistres des flottes TPM. L'adéquation des sinistres ordinaires est très médiocre : les valeurs de D_n et du khi deux sont élevées. Au final, c'est encore la loi log normale qui modélise le mieux les données. C'est aussi la loi qui donne la probabilité de sinistres graves la plus élevée. Cette probabilité reste néanmoins très sous-estimée (d'un facteur 3.5).

On a testé la finitude des moments. Le résultat graphique est très net : l'espérance est infinie et la variance finie, d'où un paramètre α compris entre 1 et 2, i.e. : ξ compris entre 0.5 et 1.

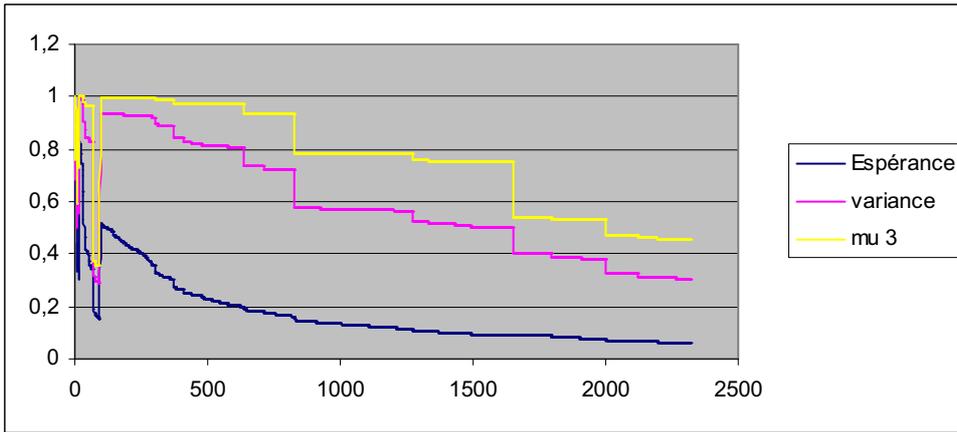
La valeur de α obtenue comme EMV est de 0.77. On remarque que les paramètres de la loi GPD sont assez proches de ceux obtenus pour les flottes AUT. C'est assez inattendu car on pensait les sinistres des TPM plus graves. C'est le cas mais moins qu'on ne l'aurait cru. En fait, le gros de la différence vient de la fréquence des sinistres graves (2% des sinistres pour les TPM contre 0.85% pour les AUT).

On a donné le graphe des lois GPD et empiriques pour les coûts en excès de 15000 euros. L'adéquation paraît bonne et contrairement au cas des flottes AUT, elle reste valable même pour les très gros coûts de sinistres. Un problème néanmoins : comme pour les AUT, il y a une forte surestimation du coût moyen des sinistres graves.

- On modélise enfin la loi des coûts de sinistres pour les flottes TPV. Ici, il convient d'être plus prudent car le nombre de sinistres observés est nettement plus faible (les TPV ne représentent que 5% des contrats). L'adéquation de la loi ordinaire est assez moyenne, du même ordre que pour les sinistres AUT. La meilleure loi semble être la loi de Pareto, devant les lois log normale et log gamma. Néanmoins, c'est la loi log normale qui sous estime le moins la probabilité pour un sinistre d'être grave. C'est donc elle que nous avons retenu.

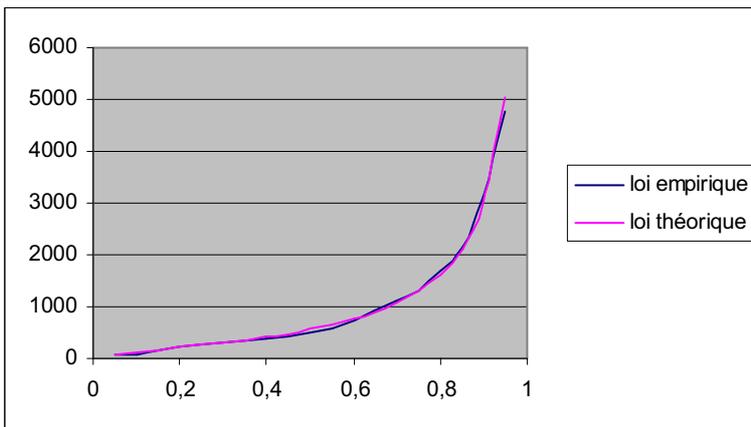
Le test graphiques des moments laisse supposer une queue extrêmement épaisse, la variance et le moment d'ordre 3 sont manifestement infinis et il y a doute pour l'espérance (ce qui signifie que le paramètre ξ pourrait être supérieur à 1). Les estimateurs ne vont pas dans ce sens et sont de l'ordre de 0.8. L'EMV est de 0.851. La loi est donc à queue plus épaisse que celles des autres catégories de flottes. L'adéquation de la loi GPD aux excès de coût constatés est bonne. Sur le deuxième graphique, c'est moins net mais cela vient du fait qu'il y a très peu de sinistres TPV de plus de 200000 euros. La courbe empirique est donc très irrégulière et ne peut pas être bien approchée avec seulement deux paramètres. Par contre, on constate à nouveau un écart important entre les moyennes théoriques et empiriques.

Au final, les TPV ont des sinistres attritionnels peu coûteux et une proportion moyenne de sinistres extrêmes moyenne (supérieure aux AUT mais inférieure aux TPM). Par contre, les coûts de ces sinistres graves sont particulièrement élevés (supérieurs de 40% à ceux des TPM).



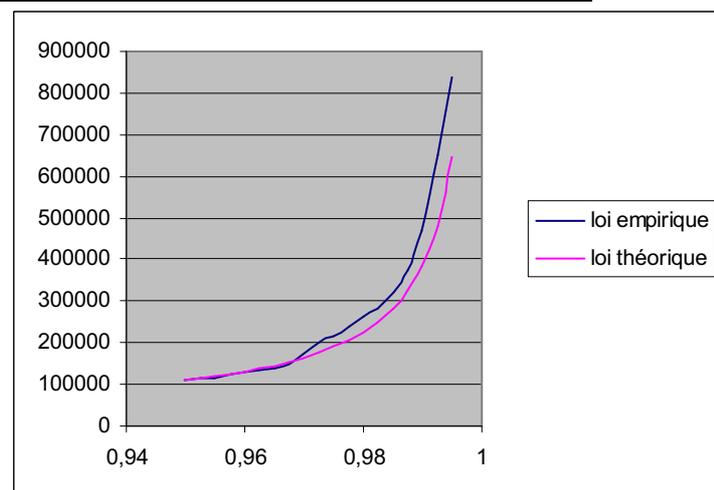
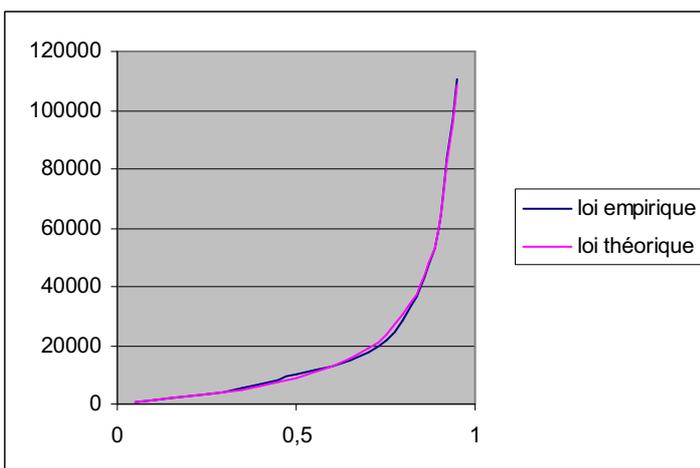
Finitude des moments des couts

| | Dn | Khi 2 | Probabilité de dépasser 15000 |
|-------------------|------------|-------|-------------------------------|
| Loi exponentielle | 0,23177778 | 607 | 3,E-05 |
| Loi gamma | 0,18444444 | 513 | 1,E-06 |
| Loi log normale | 0,06222222 | 312 | 0,51% |
| Loi log gamma | 0,064 | 259 | 0,30% |
| Loi de Pareto | 0,08333333 | 309 | 0,16% |
| Vraie valeur | | | 0,85% |

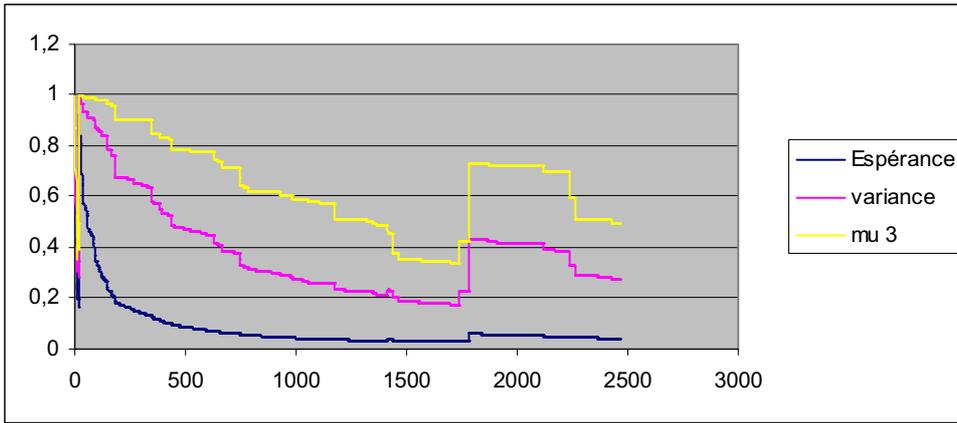


Adéquation des lois de sinistres attritionnels

| Type de flotte | Nombre d'observations | ksi | tau | Moyenne empirique | Moyenne théorique | Variance empirique | Variance théorique |
|----------------|-----------------------|-------|-----------|-------------------|-------------------|--------------------|--------------------|
| AUT | 2322 | 0,735 | 13491,945 | 37429 | 50913 | 3,5E+12 | infinie |

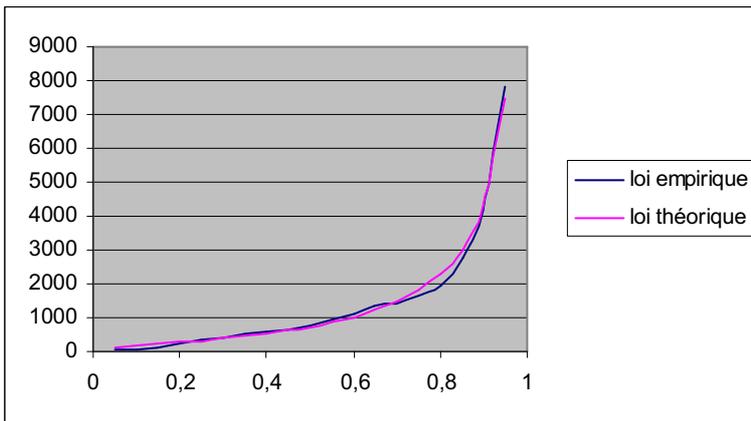


Adéquation de la loi des sinistres graves



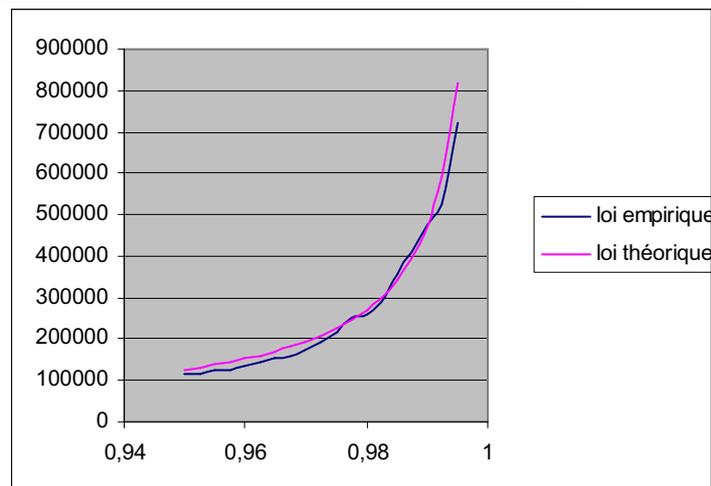
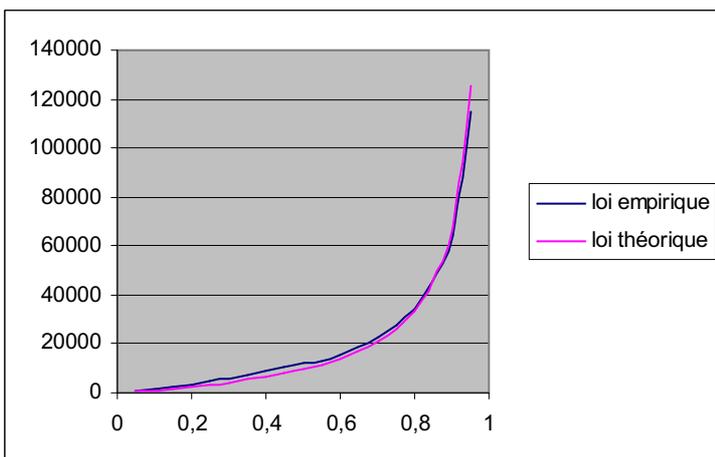
Finitude des moments des couts

| | Dn | Khi 2 | Probabilité de dépasser 15000 |
|-------------------|-------------|-------|-------------------------------|
| Loi exponentielle | 0,2805 | 1674 | 3,E-05 |
| Loi gamma | 0,263833333 | 1658 | 5,E-05 |
| Loi log normale | 0,126166667 | 1126 | 0,60% |
| Loi log gamma | 0,153333333 | 1176 | 0,40% |
| Loi de Pareto | 0,172833333 | 1434 | 0,20% |
| Vraie valeur | | | 2,00% |

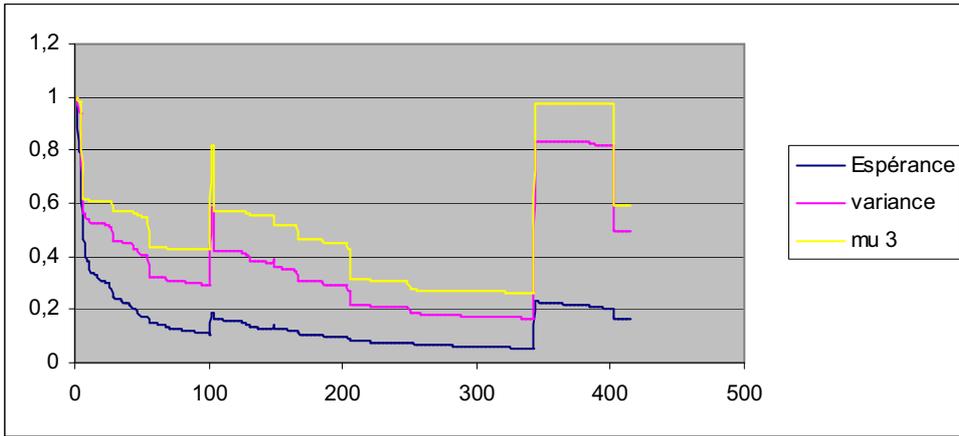


Adéquation des lois de sinistres attritionnels

| Type de flotte | Nombre d'observations | ksi | tau | Moyenne empirique | Moyenne théorique | Variance empirique | Variance théorique |
|----------------|-----------------------|------|----------|-------------------|-------------------|--------------------|--------------------|
| TPM | 2477 | 0,77 | 13987,68 | 45228 | 60816 | 4,1E+12 | infinie |

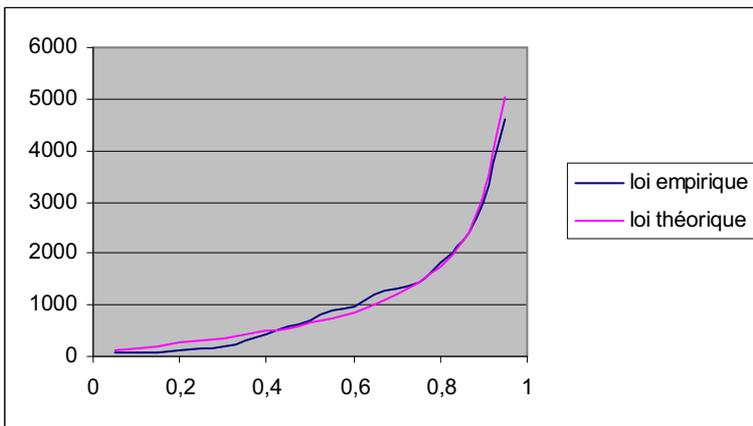


Adéquation de la loi des sinistres graves



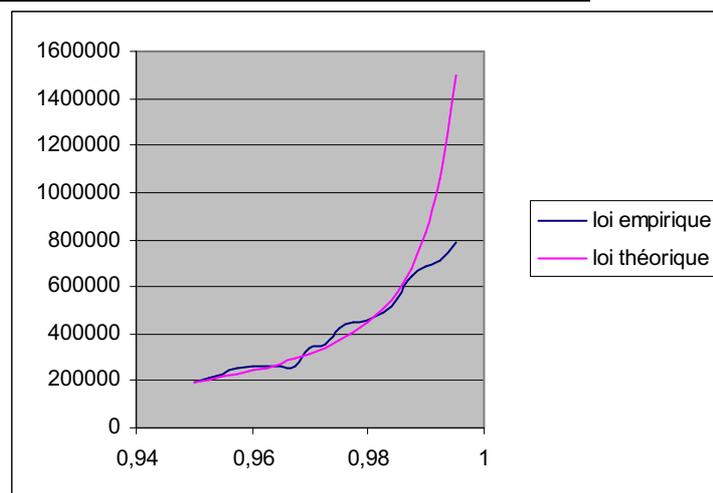
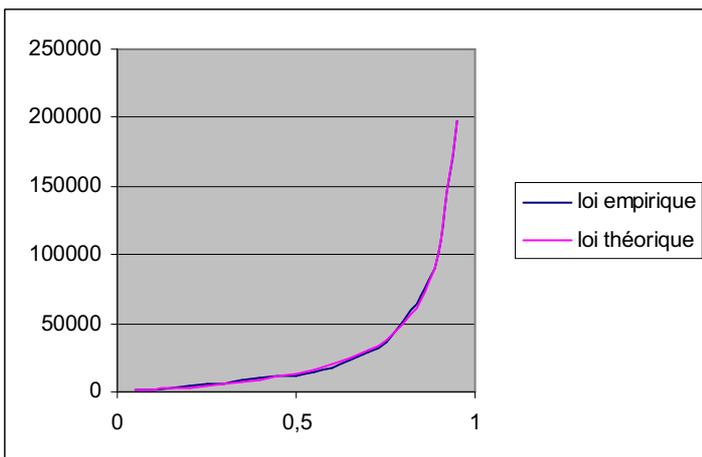
Finitude des moments des couts

| | Dn | Khi 2 | Probabilité de dépasser 15000 |
|-------------------|--------|-------|-------------------------------|
| Loi exponentielle | 0,1704 | 590 | 2,E-06 |
| Loi gamma | 0,1202 | 480 | 1,E-05 |
| Loi log normale | 0,073 | 719 | 0,42% |
| Loi log gamma | 0,0712 | 602 | 0,25% |
| Loi de Pareto | 0,0726 | 432 | 0,07% |
| Vraie valeur | | | 1,30% |



Adéquation des lois de sinistres attritionnels

| Type de flotte | Nombre d'observations | ksi | tau | Moyenne empirique | Moyenne théorique | Variance empirique | Variance théorique |
|----------------|-----------------------|-------|-----------|-------------------|-------------------|--------------------|--------------------|
| TPV | 414 | 0,851 | 15977,866 | 62967 | 107234 | 5,3E+13 | infinie |



Adéquation de la loi des sinistres graves

2.2.4.3 Comparaison des modèles

On a obtenu deux modèles différents pour les coûts de sinistres : en tenant compte du type de véhicule et en se basant sur l'usage. Reste à choisir entre ces eux. Plusieurs critères peuvent être envisagés. On va en utiliser trois : l'adéquation des lois aux données, la taille des classes et le caractère discriminant du modèle.

○ On veut tout d'abord que les lois obtenues modélisent le plus finement possible les données. L'adéquation des lois aux valeurs empiriques doit donc être la meilleure possible. On donne ci contre les statistiques d'adéquations des deux modèles. Les khis deux ont le même nombre de degré de liberté (17). La statistique de Kolmogorov a été normalisée en posant $D_n^* = \sqrt{n}D_n$. La loi de D_n^* ne dépend pas du nombre d'observations. Ainsi, on peut comparer les valeurs obtenues. On donne enfin le ratio $\frac{\text{Proportion théorique de sinistres graves}}{\text{Proportion empirique de sinistres graves}}$ et le quotient $\frac{\text{Coût moyen théorique des sinistres graves}}{\text{Coût moyen constaté des sinistres graves}}$. En ce qui concerne les lois de sinistres ordinaires, le premier modèle paraît nettement supérieur. Pour les lois de sinistres graves, le second paraît meilleur mais c'est moins net. Enfin, le second modèle sous estime fortement la probabilité de sinistres graves et surestime leurs coûts. Il paraît donc globalement moins bon.

○ Il est important que chaque classe ait une taille suffisante. En effet, si une classe est trop petite, un sinistre de très fort montant pourrait fausser largement le calcul du S/C. Par exemple, si une classe ne compte qu'un million d'euros de primes par an et qu'un sinistre extrême survient, le montant total des sinistres pourrait presque doubler. Via les mécanismes d'écrêtement et de crédibilité, cela entraînerait un doublement du S/C de tous les membres de la classe, ce qui est clairement inacceptable. Au final, il paraît préférable que chaque classe compte au moins 10 millions d'euros de primes, et idéalement plus de 20 millions.

On constate que pour le premier modèle, il n'y a qu'une catégorie totalisant moins de 10 millions et deux de moins de 20. Par contre, dans le second presque toutes les catégories de TPV et une partie de celles de TPM sont problématiques, voire très problématique (parfois moins d'un million).

○ Enfin, on souhaite segmenter le portefeuille de façon à ce que les coûts de sinistres soient les plus différents possibles entre les classes. On a procédé comme pour le choix des classes de cotisations : on calcul la variance interclasse dans le cas du premier modèle (2 classes : M3T5 et P3T5) et du second (3 classes : AUT, TPM et TPV) et on la rapporte à la variance totale. On constate que le R^2 du premier modèle est plus élevé, alors qu'il compte moins de classes. Néanmoins, vu la faible valeur des R^2 observés, il faut prendre ce résultat avec prudence.

Au final, le premier modèle semble donc être le meilleur et on distinguera les flottes selon le critère M3T5-P3T5.

| | Sinistres ordinaires | | | Sinistres graves | | |
|------|----------------------|-------|------------|------------------|-------|------------|
| | Dn normalisé | Khi 2 | Pth/Pemp | Dn normalisé | Khi 2 | Eth/Eemp |
| M3T5 | 18,23 | 106 | 0,938 | 0,95592468 | 792 | 1,02 |
| P3T5 | 15,39 | 79 | 1,059 | 3,1151963 | 1230 | 1,09 |
| AUT | 33,4841729 | 312 | 0,6375 | 0,91555557 | 928 | 1,36025542 |
| TPM | 45,2197833 | 1126 | 0,3 | 4,77786898 | 628 | 1,34465375 |
| TPV | 14,8533027 | 719 | 0,32307692 | 1,52602425 | 1489 | 1,70301904 |

Adéquation des lois de sinistres ordinaires et graves pour les deux modèles

| | M3T5 | P3T5 |
|---|-------|------|
| 1 | 12386 | 3069 |
| 2 | 6085 | 1988 |
| 3 | 4055 | 2542 |
| 4 | 1273 | 1830 |
| 5 | 259 | 495 |
| 6 | 27 | 71 |

| | M3T5 | P3T5 |
|---|------|------|
| 1 | 31 | 7,9 |
| 2 | 42 | 14 |
| 3 | 60 | 41 |
| 4 | 49 | 77 |
| 5 | 31 | 60 |
| 6 | 15 | 30 |

Nombre de contrats et primes (en million d'euros) pour les catégories du premier modèle
En rouge, les catégories à moins de 1 million, en rose celles à moins de 10 et en orange celles à moins de 20.

| | AUT | TPM | TPV |
|---|-------|------|-----|
| 1 | 14789 | 482 | 204 |
| 2 | 6806 | 771 | 496 |
| 3 | 4552 | 1396 | 649 |
| 4 | 1554 | 1208 | 341 |
| 5 | 310 | 362 | 82 |
| 6 | 35 | 56 | 7 |

| | AUT | TPM | TPV |
|---|-----|-----|-----|
| 1 | 37 | 1,5 | 0,7 |
| 2 | 48 | 5,7 | 3,6 |
| 3 | 68 | 23 | 10 |
| 4 | 61 | 51 | 14 |
| 5 | 37 | 45 | 10 |
| 6 | 18 | 24 | 2,9 |

Nombre de contrats et primes (en million d'euros) pour les catégories du second modèle
En rouge, les catégories à moins de 1 million, en rose celles à moins de 10 et en orange celles à moins de 20.

| | R2 |
|----------|----------|
| Modèle 1 | 3,41E-04 |
| Modèle 2 | 2,58E-04 |

Proportion de variance des coûts expliquée par les deux segmentations.

2.2.4.4 Modèle à double critère

l'étude précédente nous a amené à modéliser les lois en fonction du type de véhicules. Néanmoins, nous avons aussi vu que les différences liées à l'usage étaient nettes. On peut donc se demander s'il ne faut pas segmenter en tenant compte de ces deux critères. On doit déterminer les lois pour six types de flottes distinctes : M3T5 AUT, TPM et TPV et P3T5 AUT, TPM et TPV. On ne détaille pas le modèle ici. A nouveau, on modélise les coûts de sinistre par des lois log normales, et les coûts en excès de 15000 par une loi GPD. Le fait d'utiliser les memes lois que precedemment permet de mieux comparer les modèles.

On donne ci contre les résultats des tests d'adéquations. Pour les sinistres ordinaires, les résultats sont satisfaisants. Pour les M3T5, les valeurs obtenues sont beaucoup plus faibles que celles des modèles précédents. Pour les P3T5, les khis deux sont moins bons mais les D_n meilleurs. Par contre l'adéquation des sinistres graves est très mauvaise, surtout dans les classes ayant les effectifs les plus faibles. Cela vient sans doute du fait que la loi des coûts est plus irrégulière dans ces cas donc plus dure à approximer par des lois continues. Dans le cas des M3T5 TPM, l'adéquation est catastrophique. Cette classe comprenant très peu de contrats (530 au total), on pourrait la supprimer et affecter ses contrats soit aux P3T5 TPM, soit aux M3T5 AUT.

On donne les effectifs des différentes catégories quand on croise ce modèle avec les tailles de flottes. De très nombreuses classes ont des niveaux de primes beaucoup trop bas, ce qui rend le calcul du S/C instable. Au total, 23 classes sur 36 comptent moins de 10 millions de primes (et 8 moins de un million). En fait, l'une des classes ne comporte même aucun contrat. Les effectifs semblent donc insuffisants pour que ce modèle soit fonctionnel.

Le R^2 du modèle vaut 0.00041 donc est à peine meilleur que les précédents. Au final, on ne conserve donc pas le modèle à deux critères. Dans la suite, on distinguera les flottes uniquement suivant la taille et le type de véhicule.

| | | Sinistres ordinaires | | | Sinistres graves | | |
|------|-----|----------------------|--------------|-------|------------------|--------------|----------|
| | | Nombre | Dn normalisé | Khi 2 | Nombre | Dn normalisé | Khi deux |
| M3T5 | AUT | 232880 | 13,9947161 | 60 | 1660 | 1,30377912 | 1290 |
| | TPM | 6714 | 2,45817005 | 139 | 107 | 1,35507454 | 10887 |
| | TPV | 14515 | 5,30104141 | 191 | 121 | 0,517 | 2439 |
| P3T5 | AUT | 48570 | 5,50965062 | 154 | 663 | 1,26169053 | 1126 |
| | TPM | 131465 | 32,9948733 | 418 | 2369 | 5,4513059 | 575 |
| | TPV | 28806 | 9,84395165 | 514 | 289 | 0,5525 | 3042 |

Adéquation des sinistres ordinaires et graves pour le troisième modèle

| | M3T5 | | | P3T5 | | |
|---|-------|-----|-----|------|------|-----|
| | AUT | TPM | TPV | AUT | TPM | TPV |
| 1 | 12101 | 133 | 152 | 2668 | 349 | 52 |
| 2 | 5513 | 171 | 401 | 1293 | 600 | 95 |
| 3 | 3469 | 162 | 424 | 1083 | 1234 | 225 |
| 4 | 1135 | 60 | 78 | 419 | 1148 | 263 |
| 5 | 247 | 5 | 7 | 63 | 357 | 75 |
| 6 | 25 | 2 | 0 | 10 | 54 | 7 |

Nombre de contrats dans chacune des catégories du modèles (en rouge, celles comptant moins de 10 contrats)

| | M3T5 | | | P3T5 | | |
|---|------|-----|-----|------|-----|-----|
| | AUT | TPM | TPV | AUT | TPM | TPV |
| 1 | 30 | 0,4 | 0,5 | 6,3 | 1,1 | 0,1 |
| 2 | 38 | 1,3 | 2,9 | 9,3 | 4,4 | 0,7 |
| 3 | 52 | 2,5 | 6,4 | 16 | 20 | 3,9 |
| 4 | 44 | 2,2 | 2,8 | 17 | 49 | 11 |
| 5 | 29 | 0,7 | 0,6 | 7,5 | 43 | 9,4 |
| 6 | 14 | 0,8 | 0 | 3,9 | 23 | 2,9 |

Primes totales dans chacune des catégories du modèle (en orange, celles en dessous de 20 millions d'euros, en rose celle en dessous de 10 et en rouge celles en dessous de 1)

2.2.5 Résumé

On résume si dessus le modèle retenu pour les coûts de sinistres. Les flottes sont séparées suivant le critère M3T5/P3T5. La loi est obtenue à partir de deux composantes : une loi représentant les sinistres ordinaires (i.e. : inférieurs à 15000 euros) qui est log normale et une loi représentant les sinistres graves (i.e. : supérieurs à 15000 euros) qui est une loi de type Pareto. Au final, les lois de coûts ont donc 5 paramètres : espérance μ et variance σ^2 de la loi ordinaire, exposant α et paramètre de forme m de la loi de Pareto et la probabilité pour un sinistre d'être ordinaire p . p a été choisi de façon à ce que la fonction de répartition de la loi des coûts soit continue, c'est-à-dire de façon à ce que $P(\log N(\mu; \sigma^2) \leq 15000) = p$. On donne les propriétés mathématiques des lois retenues.

Remarque :

La loi que nous avons retenue peut être vue comme une loi mélange entre une loi log normale bornée à 15000 et une loi de Pareto, de paramètre de mélange p .

◇ Calcul de la fonction de répartition

Soit F_1 (resp F_2) la fdr de la loi des sinistres ordinaires (resp graves). La fonction de répartition de la loi des coûts F est donnée par :

$$* F(t) = F_1(t) \text{ si } t \leq 15000.$$

$$* F(t) = p \times F_2(t - 15000) \text{ si } t \geq 15000.$$

◇ Calcul des quantiles

Le quantile d'ordre α est donné par :

$$* q_\alpha = F_1^{-1}(\alpha) \text{ si } \alpha \leq p.$$

$$* q_\alpha = 15000 + F_2^{-1}\left(\frac{\alpha - p}{1 - p}\right) \text{ si } \alpha \geq p.$$

◇ Calcul des moments

L'espérance des coûts vaut : $(1 - p)\varphi(15000) + p\left(\frac{m}{1 - \alpha} + 15000\right)$ où $\varphi(u) = E(X|X < u)$ avec $X \sim \log N(\mu, \sigma^2)$.

Les moments d'ordre supérieur ou égaux à deux sont tous infinis car la loi des sinistres extrêmes a une variance infinie.

2.3 Simulation de la charge totale

A présent que nous disposons de loi pour la fréquence et la sévérité des sinistres, nous pouvons étudier le coût total $X = \sum_{i \leq N} C_i$. Le problème qui se pose est qu'il n'existe pas de formules exactes pour exprimer la loi de X . L'une des approches possibles est donc de recourir à une méthode de type Monte Carlo. On a donc réalisé un utilitaire EXCEL qui génère aléatoirement 100000 trajectoires de sinistres et permet d'estimer le comportement de la charge totale.

Une autre quantité utile est la charge totale écrêtée. Il est courant en assurance non vie de plafonner le montant des sinistres à un seuil M . Ce processus sera étudié au paragraphe 3.1. La charge totale écrêtée est donc $X^{ecr} = \sum_{i \leq N} \min(C_i, M)$. A nouveau, il est difficile d'étudier directement cette loi et on utilisera l'utilitaire Monte Carlo.

Enfin, nous avons besoin d'estimer les moments. Pour la charge totale, c'est simple : $E(X) = E(N)E(C)$ et on peut obtenir une valeur exacte. Les autres moments sont infinis car $V(C) = \infty$. Pour X^{ecr} , les moments sont tous finis car on a borné les coûts de sinistres. Par contre, leurs valeurs sont plus compliquées à obtenir. On les déterminera donc par simulation.

Les données obtenues dans cette partie nous serviront à plusieurs reprises dans la suite du mémoire :

- pour l'écrêtement, on a besoin de connaître les quantiles de la loi des coûts.
- pour étudier la stabilité du S/C, on détermine son coefficient de variation en fonction du niveau d'écrêtement.
- dans le cas de la crédibilité, pour déterminer si des contrats doivent avoir une crédibilité totale

2.3.1 Obtention de la loi de la charge

La charge totale (écrêtée ou non) est simulée par une macro EXCEL. On donne ici les étapes du programme.

1) Nombre de sinistre.

Excel génère un nombre aléatoire u uniforme sur $[0; 1]$. Le nombre de sinistres est $\min(n | P(N \leq n) \leq u)$. Concrètement, EXCEL inscrit dans une colonne les valeurs de $P(N \leq n)$ et détermine N par une fonction RECHERCHEV.

2) Coûts des sinistres.

EXCEL détermine la probabilité p_{ord} pour un sinistre de ne pas dépasser 15000 euros. Il génère ensuite un nombre aléatoire u .

★ Si $u \leq p_{ord}$ le sinistre est ordinaire et le coût est obtenu par $LOI.LOGNORMALE.INVERSE(u, \mu, \sigma)$.

★ Si $u > p_{ord}$, le sinistre est grave. La probabilité u est ramenée sur $[0; 1]$ en posant $v = \frac{u - p_{ord}}{1 - p_{ord}}$.

Le coût est $15000 + \tau((1 - v)^{-\xi} - 1)$.

Le programme réitère cette étape pour chaque sinistre.

3) Charge .

EXCEL détermine ensuite la charge (écrêtée ou non) $X_1 = \sum_{i \leq N} \min(C_i; M)$ avec $M = +\infty$ en l'absence d'écèlement.

4) Charge IDA .

Le nombre de sinistre forfaitaire IDA, N_{IDA} , est déterminé de la même façon que N . La charge IDA forfaitaire X_{IDA} est $1204 * N_{IDA}$.

5) Charge totale.

La charge totale est $X^{ecr} = X_1 + X_{IDA}$.

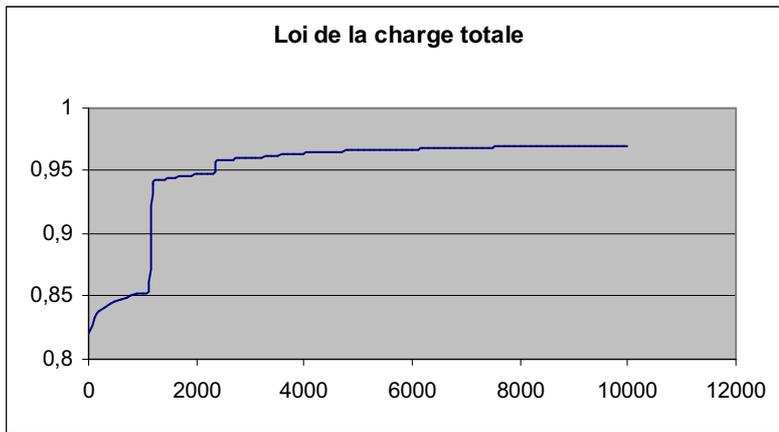
Le programme simule ainsi 100000 valeurs de X^{ecr} . Le temps de calcul reste assez court (de 6s pour la première classe de taille à 3 minutes pour ceux de la sixième).

On donne ci contre les graphiques des lois de charges pour un contrat M3T5 classe1 (charge non écrêtée et écrêtée à 3500 euros) et pour un M3T5 classe3 (charge non écrêtée et écrêtée à 3500 euros).

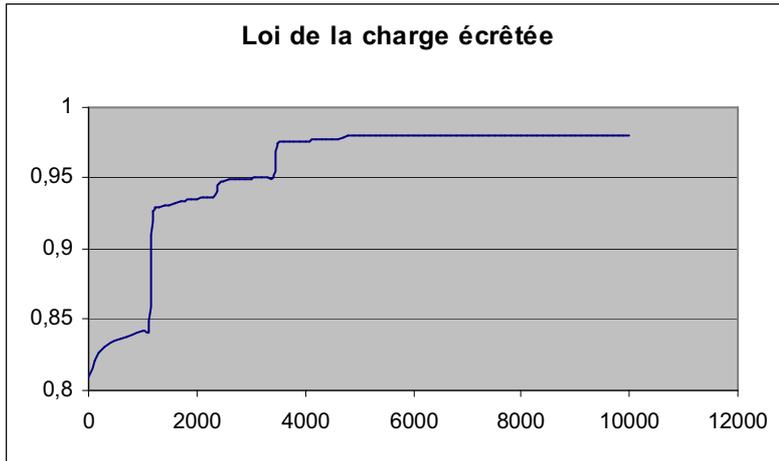
Dans le cas de la première classe, très peu de trajectoires ont plusieurs sinistres. Cela explique les seuils visibles sur les courbes :

- 0 qui correspond au cas où il n'y a aucun sinistre (82% des cas).
- 1200 euros : aucun sinistre classique mais un IDA (9% des cas).
- 2400 euros : deux sinistres IDA (1.5% des cas).
- 3500 euros : un sinistre au dessus du seuil d'écèlement (2.5% des cas).

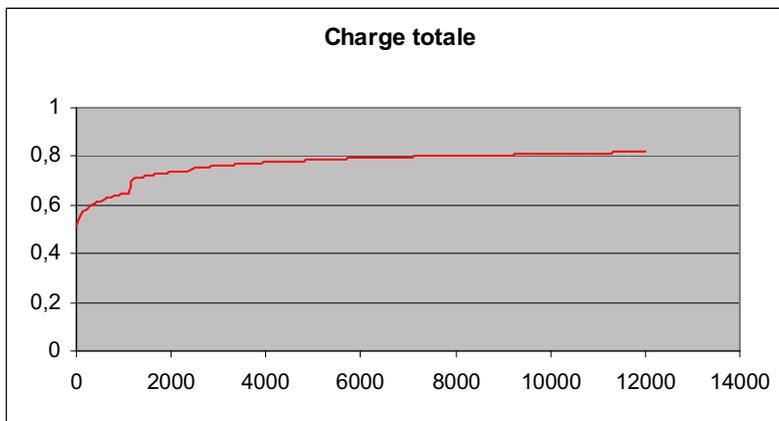
Dans le cas de la classe 3, le phénomène de seuil est moins visible car les trajectoires ont plus de sinistres mais ont distingué des sauts à 1200, 2400 et 10000 euros.



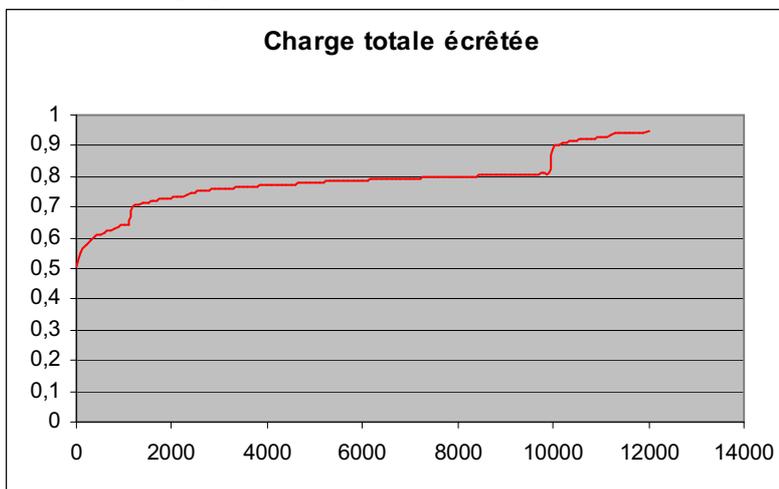
Loi de la charge pour un contrat taille 1 M3T5



Loi de la charge écrêtée à 3500 euros pour un contrat taille1 M3T5



Loi de la charge pour un contrat taille 3 M3T5



Loi de la charge écrêtée à 10000 euros pour un contrat taille3 M3T5

2.3.2 Estimation des moments

On veut à présent déterminer les moments de la charge. On rappelle que pour la charge non écrêtée, ces moments sont infinis dès l'ordre 2 car l'exposant de la loi des sinistres graves est inférieur à 2. Concrètement, les méthodes ci-dessous ne sont donc utilisées que pour la charge écrêtée et ne sont réalistes que pour des seuils M raisonnables (pour des très grands seuils, la volatilité est importante). On veut évaluer trois quantités : l'espérance $E(X)$, l'écart type σ_X et le coefficient de variation $T = \frac{\sigma_X}{E(X)}$.

On note $X_1 \dots X_{100000}$ les charges des trajectoires fournies par l'utilitaire.

- L'espérance est estimée $\bar{X} = \frac{1}{100000} \sum X_i$.

On sait que $\sqrt{n} \frac{\bar{X} - E(X)}{\sigma_X}$ tend vers une loi $N(0; 1)$.

Un intervalle de confiance à 5% est donc $[\bar{X} - 1.96 \frac{\sigma_X}{\sqrt{n}}; \bar{X} + 1.96 \frac{\sigma_X}{\sqrt{n}}]$.

- L'écart type est estimé par $\hat{\sigma}_X = \sqrt{\frac{1}{99999} \sum (X_i - \bar{X})^2}$.

On a de plus $\sqrt{n} 2\sigma_X \frac{\hat{\sigma}_X - \sigma_X}{\sqrt{\mu_4 - \sigma_X^4}} \rightarrow N(0, 1)$.

On a donc un intervalle de confiance à 5% : $[\hat{\sigma}_X - 1.96 \frac{\sqrt{\mu_4 - \sigma_X^4}}{\sqrt{n} * 2\sigma_X}; \hat{\sigma}_X + 1.96 \frac{\sqrt{\mu_4 - \sigma_X^4}}{\sqrt{n} * 2\sigma_X}]$.

Le moment d'ordre 4 μ_4 a été estimé par $\frac{1}{100000} \sum (X_i - \bar{X})^4$.

Remarque :

L'estimateur est biaisé. Néanmoins, il est asymptotiquement sans biais.

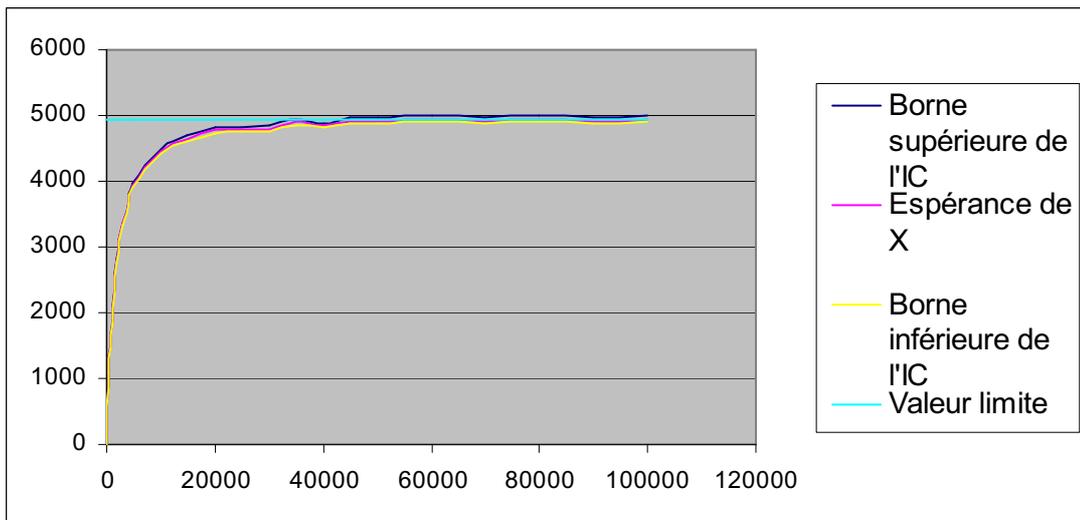
Vu la taille de l'échantillon, $\hat{\sigma}_X$ est donc acceptable.

- Le coefficient de variation $T = \frac{\sigma_X}{E(X)}$ est estimé par $\hat{T} = \frac{\hat{\sigma}_X}{\bar{X}}$. Cet estimateur est biaisé mais asymptotiquement sans biais. De plus, si note $[a; b]$ et $[c; d]$ les intervalles à 5% de $E(X)$ et σ_X , Un intervalle à 10% pour T est donné par $[\frac{c}{b}; \frac{d}{a}]$.

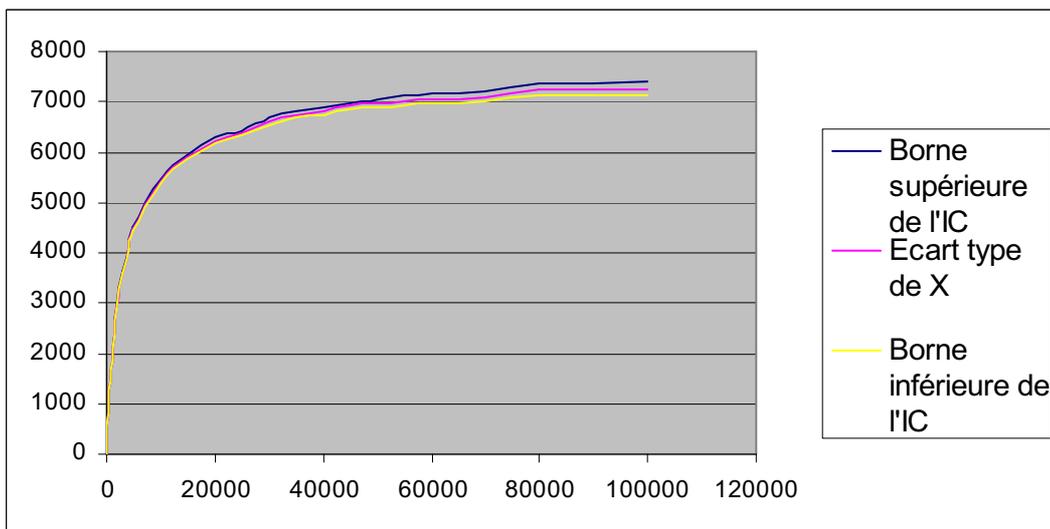
On donne ci-contre les différents moments de la charge d'un contrat taille 4 P3T5, en fonction du seuil d'écrêtement. On constate que ces courbes sont toutes croissantes concaves : plus on écrête à un niveau élevé, plus les sinistres coûtent chers et plus ils sont volatiles.

Dans le cas de l'espérance, au delà d'un seuil de 50000 euros, les résultats obtenus n'évoluent presque plus. En fait, très peu de sinistres dépassent le seuil donc l'augmenter ne change pas grand-chose. La valeur obtenue est approximativement l'espérance de la charge sans écrêtement.

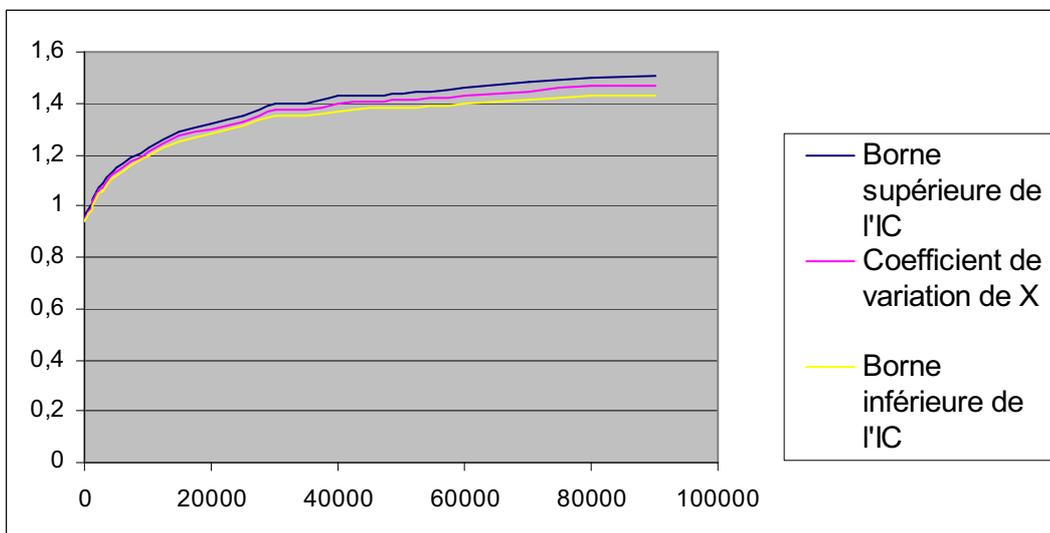
Dans le cas de l'écart type et du coefficient de variation, il n'y a pas convergence (ces quantités tendent vers l'infini avec le seuil). On remarque de plus que l'intervalle de confiance est plus large que celui de l'espérance.



Espérance de la charge en fonction du seuil d'écrêtement



Ecart type de la charge en fonction du seuil d'écrêtement



Coefficient de variation de la charge en fonction du seuil d'écrêtement

Synthèse de la deuxième partie

Au final, on a modélisé les différentes composantes de la sinistralité : fréquence, coût puis charge totale. On résume ici les résultats obtenus.

Les contrats ont été répartis en 6 classes en fonction du montant de leurs primes. La fréquence des sinistres d'un contrat suit une loi binomiale négative dont les paramètres dépendent de sa classe. Le nombre de sinistres IDA forfaitaires suit également une loi binomiale négative.

Pour l'étude de la sévérité, le portefeuille est divisé en deux : flottes M3T5 et P3T5. Les sinistres sont séparés en sinistres ordinaires (ou attritionnels) et graves. Dans le cas attritionnel, le montant suit une loi log normale. Dans le cas grave, une loi de Pareto. Le montant limite est de 15000 euros : au-delà, un sinistre est jugé grave.

Enfin, la charge totale et la charge totale écrêtée sont simulées par une méthode de type Monte Carlo en utilisant les lois ci-dessus.

Il convient de rester prudent avec le modèle obtenu. En effet, la sinistralité est un phénomène très complexe. De plus, le portefeuille étudié est important (plusieurs dizaines de milliers de contrats et plusieurs centaines de milliers de sinistres). Par contraste, notre loi ne fait intervenir que 8 paramètres (4 pour la fréquence et 4 pour le coût). L'adéquation ne peut donc qu'être partielle. Enfin, la segmentation reste assez simple. Nous n'avons en particulier pas retenu le critère de l'usage du véhicule. Les classes sont donc assez hétérogènes et les lois ne reflètent qu'un comportement moyen. Tout ceci doit nous conduire à ne pas surestimer les outils développés dans cette partie.

Chapitre 3

Processus de majoration QP

Quand un nouveau contrat intègre le portefeuille, une prime est calculée. Elle est déterminée à partir d'une grille tarifaire, donc en fonction de ses données propres : nombres et types de véhicules, zone géographique, usage. Néanmoins, l'information est partielle. Les chauffeurs de l'entreprise peuvent être plus ou moins bon, la fréquence d'utilisation des véhicules n'est pas connue, la valeur des marchandises transportées l'est rarement. Dans le cas des flottes ouvertes, même certaines données de bases peuvent manquer. L'hétérogénéité au sein d'une classe tarifaire est donc grande et certaines polices sont sous-tarifées. De plus, les contrats peuvent évoluer et devenir plus mauvais. Tout cela fait que pour une partie du portefeuille, la prime est insuffisante. Il convient donc de contrôler les contrats en cours et de voir ceux devant faire l'objet de majorations. Ce processus est appelé majoration QP pour « qualité du portefeuille ».

Un moyen simple d'évaluer la rentabilité d'un contrat est de calculer son ratio sinistres à cotisation ou S/C. On fixe un seuil correspondant au niveau d'équilibre et les contrats au dessus de ce seuil sont jugés non rentables et font l'objet d'une majoration. En pratique, ces ratios sont calculés chaque mois pour tout le portefeuille par la direction technique qui envoie ensuite la liste des contrats à majorer et dont l'échéance est proche aux directions régionales, ainsi que les majorations que nous recommandons. Celles-ci négocient ensuite avec les clients.

On pourrait être tenté d'évaluer un contrat par le ratio de l'année précédente : les sinistres survenus cette année là sur les primes acquises. Néanmoins ce n'est pas acceptable. En effet, le S/C est une donnée extrêmement volatile. Il est presque impossible de déduire le ratio de l'année à venir à partir de celui de l'année passée. On utilise donc la moyenne des ratios des dernières années. Cependant les données anciennes sont moins pertinentes. En pratique, on utilise les ratios de l'année en cours et des trois années précédentes. Cela limite la volatilité, même si elle reste importante.

Deux phénomènes principaux génèrent cette instabilité. Tout d'abord les très gros sinistres. Pour un contrat moyen, un sinistre corporel peut facilement augmenter le S/C d'une année de 1 (voire de beaucoup plus). La solution à ce problème est l'écrêtement : on fixe une valeur maximale aux sinistres, le reste du coût étant mutualisé. Ensuite, les nombre et les coûts des sinistres (même écrêtés) sont très instables. On limite ce facteur en ayant recours à la crédibilité : le S/C d'un contrat sera une moyenne de son ratio et de celui du portefeuille (ou de sa catégorie).

Dans un premier temps, on expose les problématiques d'écrêtement, et les méthodes que nous avons retenues. Ensuite, nous présenterons brièvement plusieurs modèles de crédibilité avant de les appliquer au portefeuille. Enfin, nous détaillerons le processus de calcul du S/C et de majoration QP.

3.1 Ecrêtement des sinistres

Un problème classique en assurance non vie est le poids très important des sinistres extrêmes. Le ratio S/C est très sensible à la présence de grands sinistres. L'illustration la plus connue de ce phénomène est l'automobiliste qui en 1976 fit dérailler un train directement dans le canal voisin à Bar le Duc. Au final, 16 millions d'euros de dégâts pour moins de 100 euros de primes soit un S/C supérieur à 100000, pour un conducteur pas spécialement mauvais. Sans chercher aussi loin, pour un contrat de moins de 5000 euros de primes (la moitié du portefeuille), n'importe quel gros sinistre corporel garantit un S/C supérieur à 1 et même potentiellement très supérieur. Pourtant, ces montants très élevés n'ont pas une grande valeur explicative, ils peuvent largement être mis sur le compte du manque de chance. Sur mille conducteurs de même niveau (même mauvais), seuls un ou deux renverseront un piéton et auront une charge totale bien supérieure aux autres. Pour résoudre ce problème, les sinistres sont écrêtés : ils sont plafonnés à un niveau M . La charge en excès (ou surcrête) $S - M$ n'est pas imputée à l'auteur du sinistre mais répartie entre tous les assurés. Reste à déterminer comment fixer le seuil M et comment répartir la surcrête entre les assurés.

On peut en particulier se demander si le seuil M doit être fixe ou varier selon les assurés. En flotte automobile d'entreprise, l'usage est de choisir un seuil croissant avec la cotisation. Deux raisons possibles :

- Un gros contrat peut plus facilement "absorber" une perte importante qu'un petit.
- Le seuil d'écrêtement correspond à un sinistre exceptionnel. Or la notion de gros sinistres est très relative, et liée au nombre de véhicules de la flotte. Un véhicule a environ un sinistre de 15000 euros ou plus tout les 200 ans. Pour une flotte de 10 véhicules, un tel sinistre peut donc être vu comme un événement exceptionnel, sans doute dû à la malchance. Par contre, pour un contrat de 300 véhicules, un tel sinistre a plus d'une chance sur deux de se produire, il n'est donc pas exceptionnel et ne doit pas être écrêté.

Dans un premier temps, nous allons déterminer le seuil d'écrêtement par une méthode de type quantile, puis le lisser en faisant intervenir les coefficients de variation des coûts. Ensuite, nous étudierons la mutualisation de la surcrête. Enfin, nous affinerons le processus d'écrêtement en le décomposant en deux phases.

3.1.1 Détermination du seuil d'écrêtement

Il y a plusieurs méthodes pour évaluer le seuil d'écrêtement, par exemple :

- choisir un certain quantile du coût des sinistres (typiquement 5,2 ou 1%).
- choisir un seuil tel que la charge surcrête représente un certain pourcentage de la charge totale. On sait qu'en assurance automobile, une part importante de la charge provient de quelques sinistres exceptionnels.

Dans les deux cas, la valeur de M ne varie pas avec la taille du contrat, contrairement à ce qui est souhaité. L'idée retenue est de définir le seuil comme un quantile de la loi des sinistres **pour une flotte de ce type**. On jugera un sinistre exceptionnel (et donc à écrêter) si la probabilité qu'une flotte de ce type ait au moins un tel sinistre est inférieure à un certain niveau α .

Mathématiquement, si N désigne le nombre des sinistres d'un contrat et $C_1..C_N$ leurs coûts, le seuil M est donné par :

$$P(\exists i \leq N, C_i > M) = \alpha.$$

En d'autres termes, le seuil est le quantile à $1 - \alpha$ de la loi du maximum des sinistres.

On donne à titre d'exemple les seuils obtenus pour un contrat taille 3 M3T5 en fonction de α . On constate que le niveau décroît fortement avec α , surtout quand α est proche de 0. Dans la pratique, le risque α a été fixé à 2%.

On donne maintenant une formule pour déterminer M .

Posons $F(t) = P(C \leq t)$ la fonction de répartition de C et $g(u) = \sum_n P(N = n)u^n$ la série génératrice de N .

$$\begin{aligned} P(\exists i \leq N, C_i > M) &= \alpha \\ P(\forall i \leq N, C_i \leq M) &= 1 - \alpha \\ \sum_n P(N = n)P(\forall i \leq n, C_i \leq M) &= 1 - \alpha \\ \sum_n P(N = n)P(C \leq M)^n &= 1 - \alpha \\ g(F(M)) &= 1 - \alpha \\ M &= F^{-1}(g^{-1}(1 - \alpha)) \end{aligned}$$

Ainsi, M est le quantile d'ordre $g^{-1}(1 - \alpha)$ de F . En pratique, on utilise les lois de N et C déterminées précédemment. Dans ce cas, $g(u) = \left(\frac{p}{1 - (1 - p)u}\right)^r$. On donne les quantiles de la loi du coût utilisés comme seuil et on en déduit une valeur pour les seuils d'écrêtement.

| alpha | seuil |
|-------|-------|
| 0,05 | 5603 |
| 0,04 | 6406 |
| 0,03 | 7544 |
| 0,02 | 9375 |
| 0,015 | 10858 |
| 0,01 | 13250 |
| 0,005 | 18323 |

Seuil d'écêtement pour un contrat taille3 M3T5 en fonction du risque alpha.

| Classe | Quantile |
|--------|----------|
| 1 | 82,488% |
| 2 | 93,879% |
| 3 | 97,611% |
| 4 | 99,254% |
| 5 | 99,828% |
| 6 | 99,961% |

Quantiles de la loi des couts utilisés pour le calcul des seuils.

| Classe | seuil M3T5 | seuil P3T5 |
|--------|------------|------------|
| 1 | 2755 | 3740 |
| 2 | 5605 | 7084 |
| 3 | 9375 | 11399 |
| 4 | 16462 | 19282 |
| 5 | 31543 | 35473 |
| 6 | 59108 | 63929 |

Seuils donnés par la formule (alpha=2%)

3.1.2 Utilisation des coefficients de variation :

Le recours à l'écrêtement sert entre autre à garantir une certaine stabilité dans le temps de la charge totale d'un contrat en éliminant les sinistres exceptionnels. Concrètement, on veut donc minimiser l'écart type de $X^{ecr} = \sum_{i \leq N} C_i^{ecr}$. Cela étant, un plus gros contrat aura par nature un écart

type plus important. L'indicateur utilisé sera donc le coefficient de variation : $T = \frac{\sigma(X^{ecr})}{E(X^{ecr})}$.

Il est assez délicat de calculer T directement. On va donc utiliser l'utilitaire Monte Carlo présenté au paragraphe 3.

On constate que le coefficient de variation croît assez rapidement avec le seuil d'écrêtement. En effet, plus le seuil est bas et moins il y a d'incertitude sur le coût des sinistres. A l'extrême, ($M \sim 0$), tout les coûts de sinistres sont quasi identiques et il ne reste plus que l'incertitude sur le nombre des sinistres. Dans ce cas, $T = \frac{\sigma(N)}{E(N)}$.

A l'inverse, si M tends vers l'infini, T tends vers $\frac{\sigma_X}{E(X)} = +\infty$.

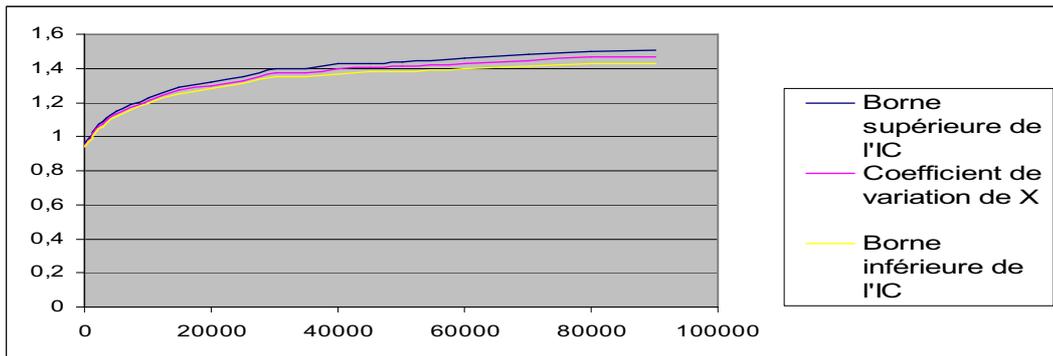
On redonne ci contre le graphe de T en fonction de M .

Au final, on a obtenu pour toutes les tailles de flottes les valeurs de T des flottes plus et moins de 3.5 tonnes et la valeur minimale (simulée en écrétant les sinistres à 1 euro). Elles sont indiquées dans le tableau ci contre.

On constate que le coefficient de variation décroît assez rapidement avec la taille de la flotte. Les grosses flottes, ayant plus de sinistres, ont une sinistralité qui s'écarte moins de la moyenne (en particulier, un sinistre important n'impactent pas trop le S/C). C'est d'autant plus notable que leurs seuils d'écrêtement sont plus élevés ce qui rends le S/C plus volatil.

On cherche à avoir des comportements cohérents entre les tailles de flottes et les types (M3T5 et P3T5). On veut donc homogénéiser les rapports $\frac{P3T5}{M3T5}$ et $\frac{M3T5}{minimal}$. On constate que les seuils classe 6 (M3T5 et surtout P3T5) doivent être réduit et les coûts classe 1 légèrement augmentés. Sinon, les valeurs obtenues semblent équitables.

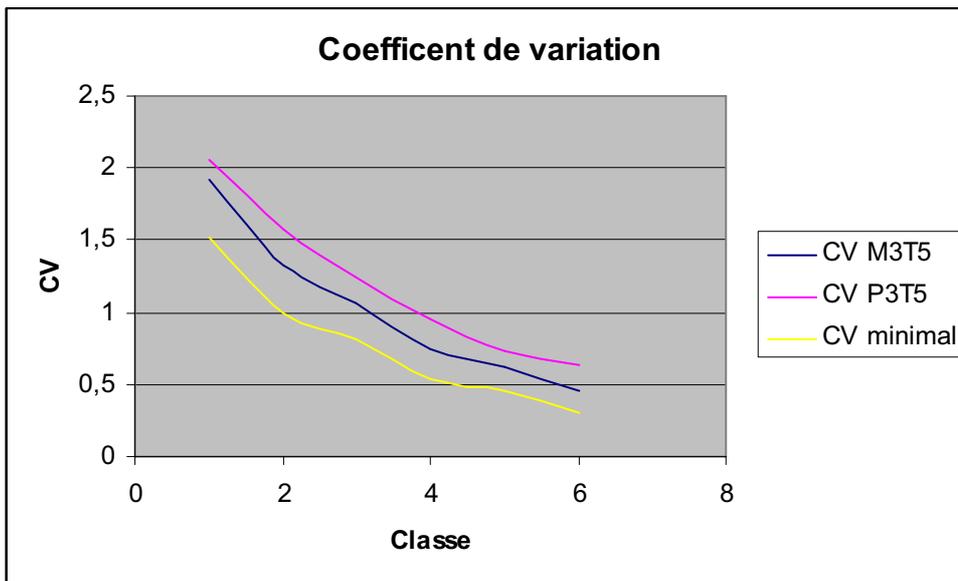
On en a déduit les seuils finalement retenus. Ils assurent que les deux ratios que nous cherchons à contrôler diffèrent peu d'une classe à l'autre.



CV en fonction du seuil d'écèlement.

| Classe | CV M3T5 | CV P3T5 | CV minimal | Rapport P3T5/M3T5 | Rapport M3T5/minimal |
|--------|---------|---------|------------|-------------------|----------------------|
| 1 | 1,92 | 2,06 | 1,52 | 1,07 | 1,26 |
| 2 | 1,33 | 1,58 | 0,99 | 1,19 | 1,34 |
| 3 | 1,06 | 1,24 | 0,82 | 1,17 | 1,29 |
| 4 | 0,75 | 0,95 | 0,54 | 1,27 | 1,39 |
| 5 | 0,62 | 0,73 | 0,46 | 1,18 | 1,35 |
| 6 | 0,45 | 0,64 | 0,3 | 1,42 | 1,5 |

Coefficient de variation de la charge des différents types de flottes



| Classe | Seuil écèlement M3T5 | Seuil écèlement P3T5 |
|--------|----------------------|----------------------|
| 1 | 3500 | 4000 |
| 2 | 6000 | 9000 |
| 3 | 9000 | 12000 |
| 4 | 16000 | 20000 |
| 5 | 30000 | 45000 |
| 6 | 45000 | 60000 |

Seuils finalement retenus.

3.1.3 Mutualisation de la surcrête

L'idée de la méthode d'écrêtement est que le niveau de certains sinistres n'est pas représentatif et que la totalité du coût ne doit donc pas être attribuée à l'assuré. Reste la question de savoir quoi faire de la charge écrêtée (ou surcrête).

○ Une solution est de l'ignorer entièrement. Néanmoins la charge surcrête correspond à un coût réel que l'assureur devra payer. Cette méthode n'est donc pas satisfaisante.

○ Une autre approche consiste à répartir les coûts entre tous les assurés. L'idée est que ces très gros sinistres sont largement liés à la malchance et qu'ils auraient tout aussi bien pu arriver à quelqu'un d'autre. Il est donc logique que tous les assurés supportent une part de la charge. Cela étant, il y a des différences de comportement assez nettes entre les assurés des différentes classes.

○ On peut enfin mutualiser la charge surcrête des sinistres d'une classe entre les contrats de cette classe. C'est l'approche retenue

Reste à la répartir entre les contrats. On note ω_i la proportion de la surcrête attribuée au contrat i . Il faut donc choisir les poids ω_i . Il y a deux approches classiques : proportionnels à la sinistralité ou à la taille.

◇ On peut choisir des poids proportionnels à la charge souscrête (restant à la charge de l'assuré après l'écrêtement) ou au nombre de sinistres. L'idée est que les grands sinistres sont souvent des petits sinistres qui ont "mal tournés". Ainsi, plus un contrat a de petits sinistres, plus son risque d'avoir un sinistre grave augmente. Ces méthodes présentent deux défauts. Tout d'abord, elle attribue plus de charges aux contrats déjà touchés. Cela revient à les pénaliser deux fois. De plus, un contrat qui n'a aucun sinistre se retrouverait avec une charge totale nulle. La logique de l'approche étant de dire qu'un sinistre pourrait arriver à n'importe qui, cela n'est pas souhaitable.

◇ On peut aussi choisir des poids proportionnel à la taille de la flotte.

L'idée est que le risque de gros sinistres augmente avec l'exposition. Celle ci pourrait être représentée par le nombre de véhicules, mais cette donnée n'est pas disponible pour toutes les flottes et certains véhicules (les P3T5 par exemple) ont beaucoup plus de risque de gros sinistres. Le choix fait à été de prendre ω proportionnel aux primes. Ainsi, chaque assuré se voit attribuer une charge surcrête*primes de l'assuré/Prime totale de la classe. Concrètement, cela revient à majorer le S/C de tous les assurés de la classe i par $\frac{\text{surcrête totale de la classe } i}{\text{total des primes de la classe } i}$.

Remarque :

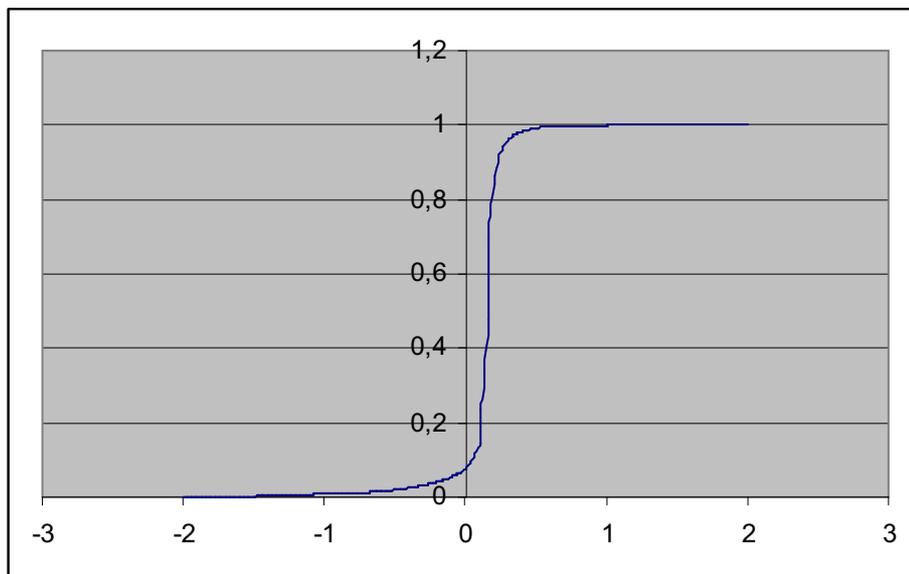
Le modèle proposé est additif. On aurait aussi pu utiliser un modèle multiplicatif du type : coût attribué au contrat = coût sous crête * λ

où le paramètre λ est calculé pour que la somme des coûts attribués soit égale à la somme des sinistres. Néanmoins avec cette méthode aussi un assuré qui n'aurait eu aucun sinistre aurait un S/C nul.

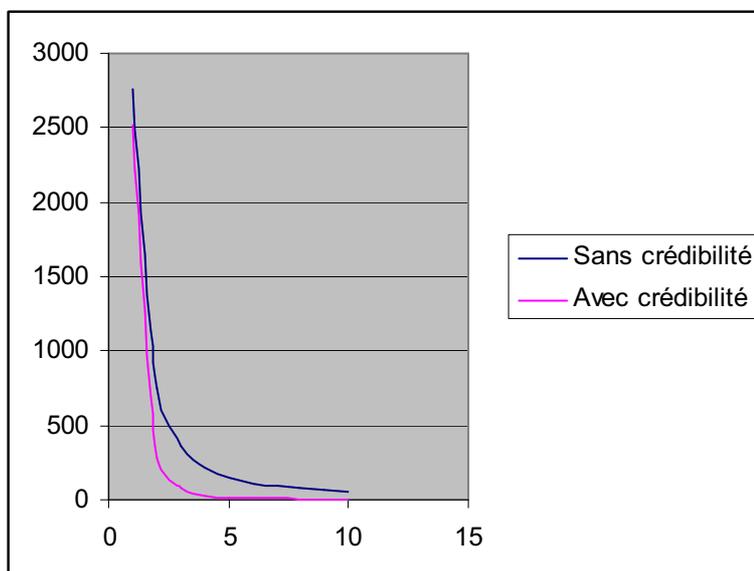
On donne ci contre la loi de $S/C_{\text{écrêté}} - S/C_{\text{non écrêté}}$. On constate que dans la majorité des cas, le processus d'écrêtement mutualisation augmente le S/C. Ce n'es pas surprenant car les seuils ont été choisis de façon à ce n'avoir que 2% de contrats écrêtés par an. Par contre, pour les contrats qui en bénéficient, le gain peut être énorme. Les contrats pour qui l'écrêtement est le plus défavorable voient leur S/C augmenter de 0.3. Par contre dans les cas les plus favorables, il baisse de 120. Ainsi, l'écrêtement permet un "nivellement par le milieu" en augmentant légèrement le S/C de la plus part des contrats pour baisser fortement celui des plus mauvais. On donne aussi le nombre de contrats dépassant un certain niveau de S/C avant et après l'écrêtement. On constate que le nombre de contrat dépassant 1 est à peu près similaire, par contre pour les seuils plus élevés, l'écrêtement réduit nettement le nombre de cas. Ainsi, l'essentiel des contrats à priori très sous tarifés étaient en fait des contrats ayant eu un sinistre exceptionnel, et donc sans doute pas si mauvais.

| Type | Taille | Charge écrêtée en % |
|------|--------|---------------------|
| M3T5 | 1 | 32,8 |
| | 2 | 22,49 |
| | 3 | 17 |
| | 4 | 21,19 |
| | 5 | 21,46 |
| | 6 | 17,34 |
| P3T5 | 1 | 34,56 |
| | 2 | 36,23 |
| | 3 | 28,24 |
| | 4 | 23,13 |
| | 5 | 26,16 |
| | 6 | 32,74 |

Pourcentage de la charge écrêtée dans chaque catégorie.



Loi de S/C avant écrêtement moins S/C écrêté mutualisé



Nombre de contrat dont le S/C dépasse un certain seuil, avant et après écrêtement

3.1.4 Deuxième niveau d'écrêtement

La méthode présentée ci-dessus permet de bien gérer les gros sinistres. Par contre, en ce qui concerne les très gros sinistres, elle est moins satisfaisante.

- La charge surcrête d'un sinistre est répartie entre les assurés de la classe. De ce fait, la présence de quelques sinistres extrêmes au sein d'une classe affecte tous les membres. Pourtant, au delà d'un certain niveau de coût, on peut considérer que le montant ne relève plus que de la malchance. Pour illustrer cela, on donne le pourcentage de charge écrêtée sein de chaque classe en 2005. Plus une flotte est grosse, plus son seuil d'écrêtement est élevé et plus ce pourcentage devrait être faible. Pour les flottes M3T5 qui ont peu de sinistres très élevés, c'est le cas (sauf pour la classe 5 qui a eu beaucoup de très gros sinistres). Par contre, pour les flottes P3T5, on constate que la situation est plus complexe. Les irrégularités sont en fait dues aux plus gros sinistres.

- De plus, le calcul des seuils a été effectué à l'aide des lois de sévérité obtenue au paragraphe 3. Or, ces lois sous-estiment les risques de sinistres de plus de 300000 euros. Ces sinistres, bien que très rares, (un sur douze mille environ) représentent tous les ans de l'ordre de 5 à 8 pour cent de la charge totale. Il paraît plus prudent de les amortir en les répartissant sur tout le portefeuille.

Au final, on effectue donc un écrêtement à deux seuils :

- * les sinistres sont tronqués à 100000 euros, la surcrête étant mutualisée sur l'ensemble du portefeuille.
- * la souscrête est ensuite tronquée à un seuil dépendant de la classe du risque, le montant écrêté est réparti entre les seuls assurés de cette classe.
- * la charge restante est attribuée à l'assuré auteur du sinistre.

Cette méthode permet de limiter les irrégularités entre classes.

Sur le portefeuille 2005, l'écrêtement à 100000 mutualise 7% de la charge. Le pourcentage mutualisé par le second seuil est donné ci-dessous. Les irrégularités sont beaucoup moins notables qu'avec la méthode à un seul seuil d'écrêtement.

| Classe | M3T5 | P3T5 |
|--------|-------|-------|
| 1 | 0,114 | 0,109 |
| 2 | 0,118 | 0,273 |
| 3 | 0,11 | 0,156 |
| 4 | 0,091 | 0,123 |
| 5 | 0,18 | 0,196 |
| 6 | 0,094 | 0,329 |

S/C additionnel venant du processus d'écèlement (méthode à un seuil)

| Classe | M3T5 | P3T5 |
|--------|-------|-------|
| 1 | 0,088 | 0,099 |
| 2 | 0,079 | 0,143 |
| 3 | 0,09 | 0,121 |
| 4 | 0,064 | 0,077 |
| 5 | 0,06 | 0,05 |
| 6 | 0,023 | 0,053 |

S/C additionnel venant du deuxième écèlement (méthode à deux seuils)

| Classe | M3T5 | P3T5 |
|--------|------|------|
| 1 | 2 | 1 |
| 2 | 2 | 6 |
| 3 | 8 | 10 |
| 4 | 7 | 15 |
| 5 | 10 | 17 |
| 6 | 3 | 18 |

Nombres de sinistres au dessus de 100000 euros

La surcrête à 100000 euros est mutualisée sur l'ensemble du portefeuille. Théoriquement, cela suppose que ces sinistres soient indépendants des différentes segmentations : taille (1 à 6), poids (M3T5 ou P3T5) et usage (TPM, TPV, AUT).

Dans le cas du poids et de l'usage, c'est improbable. En effet, on a vu que les P3T5 ont des sinistres plus élevés que les M3T5. Les TPM ont aussi des coûts plus importants et les TPV ont plus de sinistres extrêmes. On donne ci contre le pourcentage de charge écrêté au dessus de 100000 euros, pour ces différents types de flottes. On a également effectué un test d'indépendance du khi 2 sur la base des sinistres entre la variable $1_{\text{supérieur à 100000 euros}}$ et les indicatrices des variables poids et usage. On donne les p-valeurs ci contre. Les deux analyses confirment qu'il n'y a pas indépendance entre la sinistralité extrême et le poids ou l'usage.

On peut se demander s'il y a indépendance par rapport à la taille. On donc effectué un test du χ^2 comme ci dessus. A priori, il rejette l'indépendance (p-valeur de 2.5%). Les plus grosses flottes ont proportionnellement plus de très gros sinistres que les petites. Cependant, cela vient sans doute du fait que ces flottes sont la plupart du temps des P3T5 et/ou des TPM/TPV. On a donc refait le test en se limitant aux sinistres des flottes M3T5 puis à ceux des P3T5. Cette fois ci, le test ne rejette pas l'indépendance. Il en va de même si on se limite aux flottes ayant un usage donné, mais c'est moins net.

Cette analyse nous indique qu'il est préférable, du point de vue de la rigueur de mutualiser la surcrête à 100000 euros séparément pour les M3T5 et les P3T5. Néanmoins, pour des raisons de simplicité et de lisibilité, on a conservé le système précédent (mutualisation sur l'ensemble du portefeuille). Cela permet de plus une mutualisation maximale et donc d'amortir le plus possible le choc des plus gros sinistres (certains pouvant dépasser le million d'euros).

| Type de flotte | Pourcentage de charge au dessus de 100000 euros |
|--------------------------|---|
| Ensemble du portefeuille | 9,81% |
| M3T5 | 7,65% |
| P3T5 | 11,34% |
| AUT | 8,78% |
| TPM | 9,54% |
| TPV | 16,22% |

Pourcentage de charge écrêtée au seuil de 100000 euros par type de flotte

| Type de flotte | P-valeur du chi deux d'indépendance |
|--------------------------|-------------------------------------|
| Ensemble du portefeuille | 2,47% |
| M3T5 | 14,77% |
| P3T5 | 12,35% |
| AUT | 4,50% |
| TPM | 7,80% |
| TPV | 6,50% |

Test du khi deux d'indépendance entre le fait de dépasser 100000 euros et la taille de la flotte (1 à 6). Test effectué sur l'ensemble de la base sinistre puis sur les sinistres des différents types de flottes.

| Variable | P-valeur du chi deux d'indépendance |
|----------------------|-------------------------------------|
| Poids (M3T5 P3T5) | 0,06% |
| Usage (AUT TPM TPV) | 0,11% |

Test du khi deux d'indépendance entre le fait de dépasser 1000000 euros et le type de flotte. Test effectué pour le poids et pour l'usage.

3.1.5 Résumé et exemples

Décomposition des coûts de sinistres :

Si un sinistre survient, son coût C est réparti en $C = C^{ecr1} + C^{souscrête 1}$ avec

◇ $C^{ecr1} = \max(0, C - 100000)$ la part mutualisée au sein du portefeuille.

◇ $C^{souscrête 1} = \min(C, 100000)$ qui est attribuée à la classe de l'assuré.

Cette souscrête est ensuite décomposée en $C^{souscrête 1} = C^{souscrête} + C^{ecr2}$

◇ $C^{ecr2} = \max(0, C^{souscrête 1} - \text{seuil})$ la part mutualisée au sein de la classe

◇ $C^{souscrête} = \min(C, \text{seuil})$ restant attribuée à l'assuré.

Donnons un exemple. Un assuré classe 1 M3T5 (seuil=3500 euros) a un sinistre de 120000 euro.

Somme mutualisée sur tout le portefeuille : $120000 - 100000 = 20000$. Reste donc 100000 euros de coût.

Somme mutualisée entre les contrats classe1 M3T5 : $100000 - 3500 = 96500$

Somme restant attribuée à l'assuré $120000 - 20000 - 96500 = 3500$.

Obtention du S/C d'un contrat :

Au final, le S/C d'un assuré sera donc la somme de trois termes :

$$\diamond S/C^{ecr1} = \frac{\sum_{C^{ecr1}} \text{sinistres du portefeuille}}{\text{prime totale du portefeuille}}$$

$$\diamond S/C^{ecr2} = \frac{\sum_{C^{ecr2}} \text{sinistres de la classe}}{\text{prime totale de la classe}}$$

$$\diamond S/C^{propre} = \frac{\sum_{C^{souscrête}} \text{sinistres de l'assuré}}{\text{prime totale du contrat}}$$

Par exemple, supposons qu'un assuré classe 1 M3T5 (seuil=3500 euros) ait 2000 euros de primes, et un seul sinistres de 500 euros.

Au total, le montant des primes est de 2 000 000 euros et le portefeuille compte 3 gros sinistres de montants 150 000, 200 000 et 320 000.

La classe 1 M3T5 totalise 500000 euros de primes et a 5 sinistres de plus de 3500 euros : 5000 5000 10 000 50 000 et 150 000.

Au final,

$$* S/C^{ecr1} = (50\ 000 + 100\ 000 + 220\ 000)/2\ 000\ 000 = 0.185$$

$$* S/C^{ecr2} = (1500 + 1500 + 6500 + 46\ 500 + 96\ 500)/500\ 000 = 0.30$$

$$* S/C^{propre} = 500/2\ 000 = 0.25$$

Au total, son ratio S/C vaut 0.645 (contre 0.31 sans écrêtement/mutualisation) . Les bons assurés sont donc perdants puisqu'ils doivent assumer une partie des sinistres importants.

Supposons à présent qu'il n'ait eu qu'un sinistre mais que ce soit celui à 150000 euros. Son S/C mutualisé serait de 2.145 alors que son S/C aurait été de 75 sans écrêtement !

3.2 Crédibilité

On cherche à évaluer la rentabilité des contrats en portefeuille. L'indicateur le plus classique, et que nous utilisons est le S/C. Néanmoins, un problème se pose. Celui-ci varie de façon très importante d'une année sur l'autre : la sinistralité est très volatile. Plusieurs facteurs peuvent générer ces variations :

- ★ la présence d'un gros sinistre peut affecter de façon importante le S/C. Ce problème a été résolu via l'écrêtement en mutualisant la surcrête.
- ★ la présence d'un ou plusieurs sinistres intermédiaires (élevés mais en dessous du seuil d'écrêtement).
- ★ une fréquence plus (ou moins) élevée des sinistres. Typiquement, les petits contrats (moins de 2000 euros de primes) ont en moyenne un sinistre tous les 3 ans. Cela entraîne tous les 3ans un doublement du S/C.

Pour illustrer ce phénomène, on donne les corrélations entre les S/C des années 2004 à 2007 pour les différents contrats du portefeuille. Les valeurs s'étalent entre 7 et 14%. Ainsi, si on avait dû prédire fin 2006 la rentabilité en 2007 d'un contrat, utiliser le seul S/C 2006 n'aurait pas donné un bon résultat. Une solution courante pour limiter cela est de considérer la moyenne des années précédentes. On donne aussi la corrélation entre le S/C d'une année et la moyenne des trois autres. Elle est meilleure mais reste faible : de 13 à 21%.

On peut aussi évaluer la variabilité du S/C en déterminant pour chaque contrat son coefficient de variation : l'écart type de son S/C divisé par sa moyenne. L'essentiel des contrats ont un CT élevé, compris entre 1.5 et 4.

Les S/C individuels sont dangereux car très variables. Par contre, on dispose d'un indicateur plus stable dans le temps : le S/C global. Comme on a vu que la sinistralité variait de façon importante avec les différentes catégories de flottes, on a calculé le S/C global de chacune de ces catégories pour les années 2004 à 2007 : la somme de leurs sinistres sur la somme de leurs cotisations. On constate que les corrélations entre les différentes années sont excellentes, le plus souvent supérieures à 0.9. Cela indique que les fluctuations aléatoires jouent un rôle mineur quand on se place au niveau des catégories. Les CT sont également beaucoup plus faibles, en général inférieurs à 0.1. En somme, le passé est un indicateur assez fiable de la rentabilité, mais seulement au niveau global.

On pourrait donc être tenté de juger les contrats à partir du S/C du groupe auquel ils appartiennent mais cela n'est pas judicieux. En effet, les classes sont très hétérogènes. Certaines polices peuvent s'écarter de façon nette du comportement moyen. Leur attribuer le S/C collectif n'est donc pas approprié. De plus, vu la taille du portefeuille AXA, chaque contrat ne contribue que pour une proportion infime au S/C de sa classe (sauf peut être pour les flottes de taille 5 et 6). Le juger à partir du ratio global revient donc à ignorer sa sinistralité propre. Or, c'est elle que nous cherchons à mesurer.

On a ainsi deux approches. On peut juger un contrat par son expérience propre, mais c'est dangereux du fait de la volatilité du S/C (instabilité dans le temps). On peut aussi le juger par l'expérience du groupe auquel il appartient mais c'est dangereux du fait de l'hétérogénéité de ce groupe (instabilité dans l'espace). L'idée qui sous-tend la théorie de la crédibilité est de chercher une solution intermédiaire entre les deux approches. A chaque contrat on associe un facteur Z appelé crédibilité qui mesure la fiabilité de son expérience propre ($0 \leq Z \leq 1$). La valeur retenue pour le S/C est alors :

$$S/C_{\text{retenu}} = Z * S/C_{\text{propre}} + (1 - Z) * S/C_{\text{groupe}}$$

Si $Z = 0$, l'information du contrat est jugée comme absolument non fiable et on se base uniquement sur les résultats du groupe. Si $Z = 1$, l'information est jugée parfaitement fiable et on se base

uniquement sur l'expérience propre. La question se pose de savoir comment choisir la valeur de Z . Plusieurs critères paraissent naturels : plus l'expérience d'un contrat est grande (donc plus il est ancien) et plus elle paraît crédible. De plus, certains contrats ont des résultats plus stables que d'autres et doivent donc être davantage crédibilisés. Dans le cas des flottes d'entreprises, le critère qui influe le plus sur la stabilité du S/C est la taille de la flotte. Les plus grosses ont naturellement un nombre de sinistres moins variable et un coût moyen plus stable du fait du plus grand nombre de sinistres. A titre d'illustration, on a calculé les corrélations des S/C annuels pour les classes 5 et 6. On constate qu'elles sont deux fois plus élevées que celles du portefeuille pris dans son ensemble.

La théorie de la crédibilité a été développée au début du siècle par les actuaires nord américains pour déterminer si un contrat devait ou non être jugé uniquement à partir de ses statistiques propres (i.e. : si $Z = 1$). Cette méthode, baptisé « méthode des fluctuations limitées » ou « crédibilité américaine », permet aussi de gérer les cas de crédibilité partielle ($0 \leq Z < 1$). Elle manque cependant de fondements mathématiques et est moins utilisée de nos jours. L'approche actuelle, ou « crédibilité européenne », repose sur un article d'H Buhlmann publié en 1967. Ce modèle reste néanmoins frustré, car il suppose tous les contrats identiques à priori. Il a été amélioré pour prendre en compte la différence de poids entre les contrats : c'est le modèle de Buhlmann Straub, qui est sans doute le plus utilisé actuellement. Néanmoins, dans ce modèle les contrats restent identiques à un facteur de taille près. Enfin, les modèles hiérarchiques (introduits par Jewell) permettent de tenir compte de la subdivision du portefeuille en catégories et sous catégories.

Dans cette partie, nous rappellerons les modèles utilisés et nous les appliquerons au portefeuille, avant de choisir celui qui semble le plus approprié. Nous en considérerons cinq : fluctuations limitées, Buhlmann Straub, Jewell à trois et quatre niveaux et un dernier modèle proposé par Buhlmann qui fait intervenir trois estimateurs au lieu de 2.

| | 2004 | 2005 | 2006 | 2007 |
|------|-------|-------|-------|-------|
| 2004 | | 0,089 | 0,092 | 0,074 |
| 2005 | 0,089 | | 0,145 | 0,14 |
| 2006 | 0,092 | 0,145 | | 0,115 |
| 2007 | 0,074 | 0,14 | 0,115 | |

| | 2004 | 2005 | 2006 | 2007 |
|------|-------|-------|-------|-------|
| 2004 | | 0,217 | 0,212 | 0,17 |
| 2005 | 0,217 | | 0,2 | 0,167 |
| 2006 | 0,212 | 0,2 | | 0,2 |
| 2007 | 0,17 | 0,167 | 0,2 | |

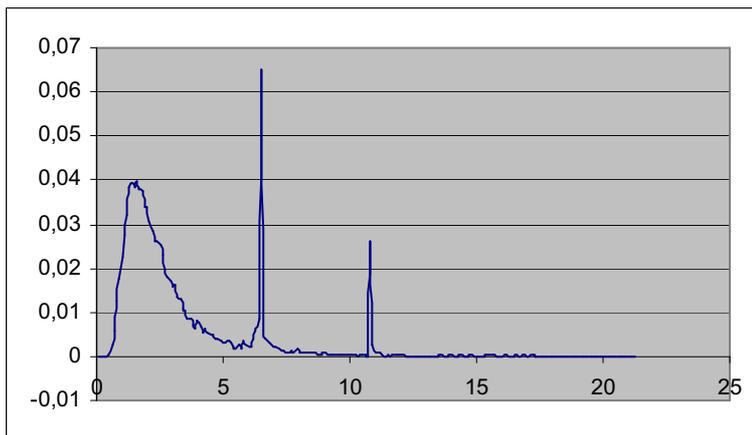
Corrélations entre les S/C des différentes années sur tout le portefeuille (tableau 1) et sur les contrats de taille 5 et 6 (tableau 2)

| | |
|------|-------|
| 2004 | 0,131 |
| 2005 | 0,212 |
| 2006 | 0,19 |
| 2007 | 0,176 |

Corrélation entre le S/C d'une année et la moyenne des S/C des autres années

| | 2004 | 2005 | 2006 | 2007 |
|------|-------|-------|-------|-------|
| 2004 | | 0,914 | 0,922 | 0,855 |
| 2005 | 0,914 | | 0,949 | 0,908 |
| 2006 | 0,922 | 0,949 | | 0,906 |
| 2007 | 0,855 | 0,908 | 0,906 | |

Corrélations entre les S/C annuels des catégories de contrats



Densité des coefficients de variations des S/C des contrats

| M3T5 | | | | | | P3T5 | | | | | |
|-------|-------|-------|-------|-------|-------|-------|-------|------|-------|-------|-------|
| 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 |
| 0,127 | 0,064 | 0,044 | 0,056 | 0,085 | 0,104 | 0,068 | 0,066 | 0,05 | 0,081 | 0,034 | 0,078 |

Coefficients de variations des S/C des catégories.

3.2.1 Description des différents modèles

3.2.1.1 Modèle des fluctuations limitées

Le modèle des fluctuations limitées est sans doute le premier modèle de crédibilité. On veut évaluer la sinistralité d'un contrat et on dispose d'un historique de ses coûts : $C_1 \dots C_n$ (dans notre cas, on appliquera la méthode au ratio S/C mais cela ne change rien au raisonnement). On peut donc estimer $E(C)$ par sa moyenne \bar{C} . La question est : cet estimateur est-il approprié, i.e. : doit-on crédibiliser pleinement le contrat. L'information apportée par \bar{X} est d'autant plus fiable que l'estimateur est stable. Le critère proposé par Whitney est :

$$P\left(\frac{|(\bar{C}) - E(\bar{C})|}{E(\bar{C})} > \alpha\right) < \varepsilon \quad (1) \text{ pour des constantes } \alpha \text{ et } \varepsilon \text{ fixée (typiquement } \alpha = \varepsilon = 0.1).$$

Ainsi, l'estimateur propre du contrat est jugé acceptable s'il a peu de chances de s'écarter nettement de la vraie valeur.

En pratique, la condition (1) est très restrictive et rarement vérifiée. Dans ce cas, on fait intervenir un estimateur stable, typiquement la moyenne C_0 des sinistralités du portefeuille. La valeur de $E(C)$ est estimée par : $\hat{C} = Z\bar{C} + (1-Z)C_0$ (2). Reste à choisir Z . Comme l'information propre du contrat est a priori plus pertinente, on souhaite prendre Z le plus grand possible, tout en conservant un estimateur raisonnablement stable. En pratique, on néglige les fluctuations de C_0 (si le portefeuille est assez grand, le hasard joue peu de rôle). La volatilité vient donc essentiellement du deuxième terme de (2). La condition (1) est donc remplacée par :

$$P\left(Z \frac{|(\bar{C}) - E(\bar{C})|}{E(\bar{C})} > \alpha\right) < \varepsilon \quad (3)$$

Finalement, on choisira donc la plus grande valeur de Z qui vérifie (3).

Les critères (1) et (3) sont simples mais ont le défaut d'être peu commode à manipuler. Il n'est en particulier pas évident de déterminer Z à partir de (3). Les actuaires nord américains ont développés plusieurs formules basées sur divers approximations. On donne ici l'une des plus classiques. On fait l'hypothèse que la statistique \bar{C} suit une loi normale (cette hypothèse découle du fait que $\sqrt{n} \frac{\bar{C} - E(C)}{\sigma(C)}$ tends en loi vers une loi $N(0;1)$). Dans ce cas, on obtient des formules simples pour Z .

Théorème : Fluctuations limitées

Soit $C_1 \dots C_n$ des variables réelles iid de variances finies. On suppose que la moyenne \bar{C} suit une loi normale.

$$\rightarrow \text{On a } P\left(Z \frac{|(\bar{C}) - E(\bar{C})|}{E(\bar{C})} > \alpha\right) < \varepsilon \text{ ssi } n > n_0 = \left(\frac{\chi_{1-\varepsilon/2}^2}{\alpha}\right)^2 \frac{V(C)}{E^2(C)}.$$

Où χ est le quantile de la loi normale.

$$\rightarrow \text{Si } n < n_0, \text{ la solution de } P\left(Z \frac{|(\bar{C}) - E(\bar{C})|}{E(\bar{C})} > \alpha\right) = \varepsilon \text{ est donnée par } Z = \sqrt{\frac{n}{n_0}}.$$

En pratique, on prendra $Z = 1$ si $n > n_0$ et $Z = \sqrt{\frac{n}{n_0}}$ sinon.

3.2.1.2 Modèle de Buhlmann Straub

Jusque dans les années 60, de nombreux modèles de crédibilité avaient été développés mais ils présentaient deux défauts :

- certains, comme le modèle des fluctuations limitées manquaient de fondations mathématiques solides
- d'autres modèles, de type paramétrique, comme le modèle de poisson gamma de Delaporte reposaient sur des hypothèses très fortes concernant les lois des données.

En 1967, Buhlmann proposa un premier modèle semi paramétrique dont les bases mathématiques étaient rigoureuses. Ce modèle comportait un défaut : il supposait tous les contrats identiques à priori, sans tenir compte de leur taille. En 1970, Buhlmann et Straub ont amélioré le modèle pour inclure l'information liée à la taille du contrat. C'est ce modèle que nous présentons ici.

On dispose d'un portefeuille de N contrats observés pendant T années (il n'est pas nécessaire que chaque contrat soit observé chaque année). On dispose pour chaque contrat et chaque année ou il a été observé d'une observation X_{jt} . Le modèle s'applique à de nombreuses grandeurs : nombre de sinistres, coût des sinistres, ratio de sinistralité, taux de destruction, coût moyen... Dans notre étude, X sera le ratio S/C. A chaque observation, on associe un poids ω_{jt} . Il y a plusieurs façon de choisir le poids : lié à la taille du contrat, à l'exposition, au nombre de sinistres ... On suppose de plus que chaque contrat possède un paramètre de risque Θ_j et que Θ et ω déterminent la loi des X_{jt} . Dans la pratique, le paramètre Θ est bien sur inobservable. On fait donc une hypothèse assez forte : la loi de X_j ne varie pas au cours du temps, au facteur de poids près. Dans la pratique, cela impose de ne prendre en compte que des données relativement récentes.

Le modèle de Buhlmann Straub repose sur les hypothèses suivantes :

Modèle de Buhlmann Straub :

(BS1) Les variables X_{jt} sont de carré intégrables.

(BS2) Les contrats sont indépendants :

Les vecteurs $(\Theta_j, X_{jt} \ t \geq 0)$ sont indépendants.

(BS3) Les contrats ont a priori le même risque :

Les vecteurs $(\Theta_j, X_{jt} \ t \geq 0)$ ont même loi.

(BS4) La valeur moyenne ne dépend que du paramètre de risque :

$E(X_{jt}|\Theta_j = \theta)$ ne dépend ni de j ni de t .

(BS5) La variance du risque à θ fixé est inversement proportionnelle au poids :

$V(X_{jt}|\Theta_j = \theta) = \frac{\sigma^2(\theta)}{\omega_{jt}}$ où $\sigma(\theta)$ ne dépend ni de j ni de t .

(BS6) Les différentes observations d'un même contrat n'influent entre elles que par le paramètre de risque :

Conditionnellement à Θ_j , les variables $X_{jt}, t \geq 0$ sont non corrélées.

On va à présent introduire plusieurs notations qui seront utiles par la suite :

◇ $\mu_0 = E(X_{jt})$. μ_0 est le S/C à priori, c'est le S/C que l'on doit attendre d'un contrat si on ne dispose pas de ses données.

◇ $\mu(\theta) = E(X_{jt}|\Theta_j = \theta)$. $\mu(\theta)$ est le S/C probable d'un contrat de paramètre θ . En pratique, comme on ne connaît pas θ , on ne connaît pas non plus $\mu(\theta)$. C'est la quantité que l'on veut estimer.

◇ $a = V(E(X_{jt}|\Theta))$. a est la variance inter contrat (ou dans l'espace). Il mesure l'hétérogénéité du portefeuille. Plus a est grand et plus les contrats peuvent s'éloigner de la moyenne du portefeuille, donc plus il faut se baser sur les données propres (i.e. : crédibiliser plus).

◇ $s^2 = E(V(X_{jt}|\Theta))$. s^2 est la variance intra contrat (ou dans le temps). Elle mesure la tendance du S/C d'un contrat à fluctuer dans le temps. Plus s^2 est grand et plus les S/C individuels sont volatils. Dans ce cas, l'expérience propre devient peu fiable et on doit se reposer plus sur les données collectives (i.e. : crédibiliser moins).

On rappelle que $V(X_{jt}) = a + s^2$.

Le but de la méthode de crédibilité est de trouver un estimateur le meilleur possible \widehat{X}_j du ratio S/C du contrat j , donc de $\mu(\theta_j)$. On prendra le meilleur au sens quadratique. On cherche donc \widehat{X}_j mesurable par rapport aux X_{jt} , $t \leq T$ qui minimise $E((X_{jT+1} - \widehat{X}_j)^2 | X_{j1} \dots X_{jt})$. La solution de ce problème est l'estimateur de Bayes $E(X_{jT+1} | X_{j1} \dots X_{jt})$ mais en pratique on ne sait pas la calculer. On va donc restreindre davantage la forme de l'estimateur. Deux approches :

- La crédibilité inhomogène : on cherche \widehat{X}_j comme fonction affine des observations, donc du type $l_0 + \sum_{t \leq T} \sum_{i \leq N} l_{it} X_{it}$.
- La crédibilité homogène : on cherche \widehat{X}_j^h comme fonction linéaire des observations, donc du type $\sum_{t \leq T} \sum_{i \leq N} l_{it} X_{it}$.

Ces estimateurs ont l'avantage d'avoir une expression simple sous les hypothèses (BS1) à (BS6).

Théorème : estimateurs de Buhlmann Straub

Sous les hypothèses (BS1) à (BS6), les estimateurs sont donnés par :

Cas inhomogène : $\widehat{X}_j = Z_j \bar{X}_j + (1 - Z_j) \mu_0$.

Cas homogène : $\widehat{X}_j^h = Z_j \bar{X}_j + (1 - Z_j) \bar{X}$.

Avec $\omega_j = \sum_t \omega_{jt}$, $\omega = \sum_{it} \omega_{it}$, $Z_j = \frac{\omega_j}{\omega_j + \frac{a}{s^2}}$, $Z = \sum_i Z_i$, $\bar{X}_j = \frac{1}{\omega_j} \sum_t \omega_{jt} X_{jt}$ et $\bar{X} = \frac{1}{Z} \sum_i Z_i \bar{X}_i$.

On remarque qu'en pratique les deux résultats reviennent au même. En effet, on ne connaît pas μ_0 donc il faut l'estimer et l'estimateur naturel est la moyenne du portefeuille. Reste à choisir la pondération. Celle utilisée pour \bar{X} , la pondération par les crédibilités, garantie une variance minimale de l'estimateur.

Reste un problème : les estimateurs dépendent des paramètres structuraux a et s^2 , or ces paramètres sont inconnus. Il faut donc eux même les estimer. Il existe plusieurs choix possibles. On donne ci-dessous un choix classique (estimateurs de De Vylder) et que l'on a utilisé dans la suite de l'étude.

→ $\widehat{s}^2 = \frac{1}{N(T-1)} \sum_{i,t} (X_{it} - \bar{X}_i)^2$ est un estimateur sans biais et consistant de s^2 .

→ $\widehat{a} = \frac{1}{N-1} \sum_i Z_i (\bar{X}_i - \bar{X})^2$ est un pseudo estimateur sans biais et consistant de a

L'estimation de s^2 ne pose pas de problème, par contre, celles de μ_0 et de a est plus complexe. En effet, les estimateurs font intervenir les crédibilités Z_i qui elles mêmes dépendent de a . On n'obtient donc pas de formules fermées à partir des observations. Néanmoins, on peut résoudre le problème par une procédure itérative :

- 1) déterminer \widehat{s}^2 .
- 2) fixer arbitrairement des valeurs a^0 et μ^0 pour a et μ_0 .
- 3) en déduire les crédibilités Z_i .
- 4) en déduire les valeurs de $\widehat{\mu}_0$ et \widehat{a} .

Réitérer le calcul à partir de l'étape 3.

Sur nos données, l'algorithme stationne en moins de 20 étapes (10 décimales exactes), et la valeur obtenue ne change pas si on modifie la condition initiale (a^0, μ^0) .

3.2.1.3 Modèle de Jewell

Le modèle de Buhlmann Straub présente un défaut notable : tous les contrats sont supposés identiques à un facteur de poids près (qui n'influe que sur la variance du risque). En pratique, on sait que certains contrats sont plus risqués, ou plus rentable. Dans notre cas, le portefeuille a été segmenté selon la taille (les six classes de cotisations définies au 2.1.2) et selon le type de véhicule (distinction entre les M3T5 et les P3T5). On aimerait tenir compte de cette subdivision dans les calculs de crédibilité. L'outil permettant de le faire est la notion de modèle hiérarchique. Ces modèles ont été introduits par Jewell en 1975. L'idée est que le portefeuille (objet de niveau 0) est divisé en classes (niveau 1) elles même divisée en sous classes (niveau 2) qui peuvent elles mêmes être divisées en sous objet et ainsi de suite. A chaque niveau, on augmente donc l'homogénéité (ce qui est une bonne chose) mais on diminue la taille des groupes (ce qui peut devenir gênant). Aux derniers niveaux, on trouve les contrats individuels puis les observations. Le S/C des contrats sera crédibilisé successivement à chaque étage. On remarque que le modèle de Buhlmann Straub est un cas particulier du modèle de Jewell (niveaux : portefeuille, contrats et observations).

On a considéré deux modèles de ce type. Le premier avec un seul niveau intermédiaire : le portefeuille est subdivisé en 12 classes (M3T5 taille 1 à 6 et P3T5 taille 1 à 6) qui elles même contiennent les contrats. Le second avec deux niveaux intermédiaires : le portefeuille est subdivisé en six classes (critère de taille) elles même divisées en deux catégories (M3T5 et P3T5). Dans la suite de ce paragraphe, on va présenter le modèle à 2 niveaux intermédiaires. La généralisation à un nombre quelconques se fait naturellement.

On dispose d'un portefeuille de N contrats observés pendant T années. Le portefeuille est divisés en M classes C_i , elles même divisées en M_i catégories $C_{ij} \subset C_i$. On dispose pour chaque contrat c_{ijk} de C_{ij} et chaque année t où il a existé d'une observation X_{ijkt} , dans notre cas le ratio S/C. A chaque observation, on associe un poids ω_{ijkt} . On suppose de plus que chaque niveau possède un paramètre de risque. Celui de la classe C_i est noté Θ_i , celui de la catégorie C_{ij} Θ_{ij} et celui du contrat c_{ijk} Θ_{ijk} . Les Θ et ω déterminent la loi des X_{ijkt} . On fait à nouveau l'hypothèse que la loi du S/C ne varie pas au cours du temps, au facteur de poids près.

On donne les hypothèses J1 à J8 du modèle de Jewell.

Modèle de Jewell à deux niveaux intermédiaires :

(J1) *Les variables X_{ijkt} sont de carré intégrables.*

(J2) *Les classes sont indépendantes :*

Les vecteurs $(\Theta_i, \Theta_{ij}$ avec $c_{ij} \subset C_i, \Theta_{ijk}$ avec $c_{ijk} \in C_i, X_{ijkt}, t \geq 0)$ sont indépendants.

(J3) *Les différentes catégories d'une même classe n'influent l'une sur l'autre que par le paramètre de risque de la classe :*

Les vecteurs $(\Theta_{ij}$ avec $c_{ij} \subset C_i, \Theta_{ijk}$ avec $c_{ijk} \in C_{ij}, X_{ijkt}, t \geq 0)$ sont indépendants conditionnellement à Θ_i .

(J4) *Les différents contrats d'une même catégorie n'influent l'un sur l'autre que par le paramètre de risque de la catégorie :*

Les vecteurs $(\Theta_{ijk}$ avec $c_{ijk} \in C_{ij}, X_{ijkt}, t \geq 0)$ sont indépendants conditionnellement à Θ_{ij} .

(J5) *Les différentes observations d'un même contrat n'influent l'une sur l'autre que par le paramètre de risque du contrat :*

Les variables $(X_{ijkt}, t \geq 0)$ sont indépendantes conditionnellement à Θ_{ijk} .

(J6) *Les contrats ont a priori le même risque :*

Les vecteurs $(\Theta_i, \Theta_{ij}, \Theta_{ijk}, X_{ijkt} \geq 0)$ ont même loi.

(J7) *La valeur moyenne ne dépend que du paramètre de risque :*

$E(X_{ijkt}|\Theta_{ijk} = \theta)$ ne dépend que de θ On le note $\mu_3(\theta)$.

(J8) La variance du risque à θ fixé est inversement proportionnelle au poids :

$$V(X_{ijkt}|\Theta_{ijk} = \theta) = \frac{\sigma^2(\theta)}{\omega_{ijkt}} \text{ où } \sigma(\theta) \text{ ne dépend que de } \theta.$$

On donne quelques notations utiles pour la suite :

$\diamond \mu_0 = E(X_{ijkt})$ C'est la valeur à priori, celle que nous devrions attribuer à un contrat si nous ne savions rien sur lui.

$\diamond \mu_1(\theta) = E(X_{ijkt}|\Theta_i = \theta)$ C'est la valeur que l'on doit attribuer à un contrat si l'on sait seulement qu'il est dans une classe de paramètre θ . On peut le voir comme la valeur attribuer si l'on veut donner le même S/C aux contrats de chaque classe.

$\diamond \mu_2(\theta) = E(X_{ijkt}|\Theta_{ij} = \theta)$ C'est la valeur que l'on doit attribuer à un contrat si l'on sait seulement qu'il est dans une catégorie de paramètre θ . On peut le voir comme la valeur attribuer si l'on veut donner le même S/C aux contrats de chaque catégorie.

$\diamond \mu_3(\theta) = E(X_{ijkt}|\Theta_{ijk} = \theta)$ C'est la valeur que l'on doit attribuer à un contrat de paramètre θ . Le but du modèle est d'estimer cette quantité car c'est celle qui correspond au S/C probable du contrat.

On définit à présent les paramètres structuraux du modèle.

- On note $s^2 = E(\sigma^2(\Theta_{ijk}))$ qui représente la variance due aux fluctuations dans le temps.
 - On note $a = E(V(\mu(\Theta_{ijk}|\Theta_{ij})))$ qui mesure l'hétérogénéité entre les contrats des catégories (variance intra catégorie).
 - On note $b = E(V(\mu(\Theta_{ij}|\Theta_i)))$ qui mesure l'hétérogénéité entre les différentes catégories des classes (variance intraclasse).
 - On note $c = V(E(\mu(\Theta_i)))$ qui mesure l'hétérogénéité des classes (variance interclasse).
- On rappelle que $V(X_{ijkt}) = a + b + c + s^2$.

On va attribuer aux éléments de chaque niveau un poids.

- Chaque observation a déjà été munie d'un poids ω_{ijkt} .
- Chaque contrat est muni du poids $\omega_{ijk} = \sum_t \omega_{ijkt}$, i.e. : de la somme des poids de ses observations.
- Chaque catégorie est munie d'un poids $\omega_{ij} = \sum_{c_{ijk} \subset C_{ij}} Z_{ijk}$ i.e. : la somme des crédibilité de ses contrats.
- Chaque classe est munie d'un poids $\omega_i = \sum_{C_{ij} \subset C_i} Z_{ij}$, i.e. : la somme des crédibilité de ses catégories.
- Le portefeuille est munie d'un poids $\omega = \sum_i Z_i$, i.e. : la somme des crédibilités de les classes.

On va attribuer aux éléments de chaque niveau une crédibilité.

★ La crédibilité du contrat c_{ijk} est $\frac{\omega_{ijk}}{\omega_{ijk} + \frac{s^2}{a}}$. Un contrat est d'autant plus crédible qu'il est stable

dans le temps (s^2 faible) et que les catégories sont hétérogènes (a grand).

★ La crédibilité de la catégorie C_{ij} est $\frac{\omega_{ij}}{\omega_{ij} + \frac{b}{a}}$. Une catégorie est d'autant plus crédible qu'elle est

homogène (a faible) et que les classe sont hétérogènes (b grand).

★ La crédibilité de la classe C_i est $\frac{\omega_i}{\omega_i + \frac{c}{b}}$. Une classe est d'autant plus crédible qu'elle est homogène

(b faible) et que le portefeuille est hétérogènes (c grand).

On définit enfin les moyennes à chaque niveau, relativement au poids :

$$\rightarrow \text{La moyenne du contrat } c_{ijk} \text{ est } \bar{X}_{ijk} = \frac{1}{\omega_{ijk}} \sum_t \omega_{ijkt} X_{ijkt}$$

$$\rightarrow \text{La moyenne de la catégorie } C_{ij} \text{ est } \bar{X}_{ij} = \frac{1}{\omega_{ij}} \sum_k \omega_{ijk} \bar{X}_{ijk}$$

$$\rightarrow \text{La moyenne de la classe } C_i \text{ est } \bar{X}_i = \frac{1}{\omega_i} \sum_j \omega_{ij} \bar{X}_{ij}$$

$$\rightarrow \text{La moyenne du portefeuille est } \bar{X} = \frac{1}{\omega} \sum_i \omega_i \bar{X}_i$$

On peut à présent résoudre le problème : arriver à trouver la meilleure estimation possible au sens L^2 du ratio des contrats : $\mu_3(\theta)$. On obtiendra aussi les meilleurs estimations pour les ratios des différents niveaux : $\mu_1(\theta)$ et $\mu_2(\theta)$. On considérera à nouveau les estimateurs inhomogènes (i.e. : affine par rapport aux observations) et homogènes (i.e. : linéaire par rapport aux observations).

Théorème : modèle de Jewell homogène

(1) L'estimateur de crédibilité homogène de la classe i est :

$$\hat{X}_i = Z_i \bar{X}_i + (1 - Z_i) \bar{X}$$

(2) L'estimateur de crédibilité homogène de la catégorie j est :

$$\hat{X}_{ij} = Z_{ij} \bar{X}_{ij} + (1 - Z_{ij}) \hat{X}_i$$

(3) L'estimateur de crédibilité homogène du contrat k est :

$$\hat{X}_{ijk} = Z_{ijk} \bar{X}_{ijk} + (1 - Z_{ijk}) \hat{X}_{ij}$$

Théorème : modèle de Jewell inhomogène

(1) L'estimateur de crédibilité inhomogène de la classe i est :

$$\hat{X}_i = Z_i \bar{X}_i + (1 - Z_i) \mu_0$$

(2) L'estimateur de crédibilité inhomogène de la catégorie j est :

$$\hat{X}_{ij} = Z_{ij} \bar{X}_{ij} + (1 - Z_{ij}) \hat{X}_i$$

(3) L'estimateur de crédibilité inhomogène du contrat k est :

$$\hat{X}_{ijk} = Z_{ijk} \bar{X}_{ijk} + (1 - Z_{ijk}) \hat{X}_{ij}$$

Reste à estimer les paramètres structuraux μ_0 , s^2 , a , b et c . En ce qui concerne μ_0 on prendra \bar{X} . Pour s^2 , on utilise le même estimateur que dans le modèle de Buhlmann Straub :

$$\hat{s}^2 = \frac{1}{N(T-1)} \sum_{ijkl} (X_{ijkl} - \bar{X}_{ijk})^2.$$

Ces deux estimateurs sont non biaisés et consistants.

Reste à estimer a , b et c .

$$\text{On estime } a \text{ par } \hat{a} = \frac{1}{\sum_{ij} (\text{card}(C_{ij}) - 1)} \sum_{ijk} (\bar{X}_{ijk} - \bar{X}_{ij})^2.$$

$$\text{On estime } b \text{ par } \hat{b} = \frac{1}{\sum_i (M_i - 1)} \sum_{ij} (\bar{X}_{ij} - \bar{X}_i)^2.$$

$$\text{On estime } c \text{ par } \hat{c} = \frac{1}{M-1} \sum_i (\bar{X}_i - \bar{X})^2.$$

Les pseudos estimateurs \hat{a} , \hat{b} et \hat{c} sont non biaisés et consistant.

Comme dans le cas du modèle de Buhlmann Straub, les crédibilités sont définies à partir des poids qui sont eux même définis à partir des crédibilités. On a donc de nouveau recours à une méthode itérative

1) On fixe arbitrairement a , b et c

- 2) On en déduit les crédibilités à chaque niveau
- 3) On en déduit les poids à chaque niveau
- 4) On en déduit les moyennes à chaque niveau
- 5) On en déduit les estimateurs de a , b et c

On recommence à l'étape 2.

A nouveau, la méthode converge en dix à quinze étapes et le résultat reste stable si on modifie les valeurs initiales.

Le tableau ci-dessous résume les estimations pour le modèle à deux niveaux intermédiaires.

| Niveau | Portefeuille | Classe | Catégorie | Contrat | Obs |
|----------------------|--|--|--|--|-----------------|
| Indice | | i | j | k | t |
| Nombre d'observation | 1 | 6 | 12 | 33000 | 89000 |
| Paramètre de risque | | Θ_i | Θ_{ij} | Θ_{ijk} | |
| S/C probable | $\mu_0 = E(X_{ijkt})$ | $\mu_1(\theta) = E(X_{ijkt} \Theta_i = \theta)$ | $\mu_2(\theta) = E(X_{ijkt} \Theta_{ij} = \theta)$ | $\mu_3(\theta) = E(X_{ijkt} \Theta_{ijk} = \theta)$ | |
| Paramètre structurel | $c = V(\mu_1(\Theta_i))$ | $b = E(V(\mu_2(\Theta_{ij}) \Theta_i))$ | $a = E(V(\mu_3(\Theta_{ijk}) \Theta_{ij}))$ | $s^2 = E(\sigma^2(\Theta_{ijk}))$ | |
| Poids | $\omega = \sum_i Z_i$ | $\omega_i = \sum_j Z_{ij}$ | $\omega_{ij} = \sum_k Z_{ijk}$ | $\omega_{ijk} = \sum_t \omega_{ijkt}$ | ω_{ijkt} |
| Crédibilité | | $Z_i = \frac{\omega_i}{\omega_i + \frac{b}{c}}$ | $Z_{ij} = \frac{\omega_{ij}}{\omega_{ij} + \frac{a}{b}}$ | $Z_{ijk} = \frac{\omega_{ijk}}{\omega_{ijk} + \frac{s^2}{a}}$ | |
| Moyenne | $\bar{X} = \frac{1}{\omega} \sum_i \omega_i \bar{X}_i$ | $\bar{X}_i = \frac{1}{\omega_i} \sum_j \omega_{ij} \bar{X}_{ij}$ | $\bar{X}_{ij} = \frac{1}{\omega_{ij}} \sum_k \omega_{ijk} \bar{X}_{ijk}$ | $\bar{X}_{ijk} = \frac{1}{\omega_{ijk}} \sum_t \omega_{ijkt} X_{ijkt}$ | |
| Estimateur | $\hat{x} = \mu_0$ ou \bar{x} | $\hat{X}_i = Z_i \bar{X}_i + (1 - Z_i) \hat{X}$ | $\hat{X}_{ij} = Z_{ij} \bar{X}_{ij} + (1 - Z_{ij}) \hat{X}_i$ | $\hat{X}_{ijk} = Z_{ijk} \bar{X}_{ijk} + (1 - Z_{ijk}) \hat{X}_{ij}$ | |

On donne aussi les notations pour le modèle à un seul niveau intermédiaire.

| Niveau | Portefeuille | Classe | Contrat | Obs |
|----------------------|--|--|--|----------------|
| Indice | | i | j | t |
| nombre d'observation | 1 | 12 | 33000 | 89000 |
| Paramètre de risque | | Θ_i | Θ_{ij} | |
| S/C probable | $\mu_0 = E(X_{ijt})$ | $\mu_1(\theta) = E(X_{ijt} \Theta_i = \theta)$ | $\mu_2(\theta) = E(X_{ijt} \Theta_{ij} = \theta)$ | |
| Paramètre structurel | $b = V(\mu_1(\Theta_i))$ | $a = E(V(\mu_2(\Theta_{ij}) \Theta_i))$ | $s^2 = E(\sigma^2(\Theta_{ij}))$ | |
| Poids | $\omega = \sum_i Z_i$ | $\omega_i = \sum_j Z_{ij}$ | $\omega_{ij} = \sum_t \omega_{ijt}$ | ω_{ijt} |
| Crédibilité | | $Z_i = \frac{\omega_i}{\omega_i + \frac{a}{b}}$ | $Z_{ij} = \frac{\omega_{ij}}{\omega_{ij} + \frac{s^2}{a}}$ | |
| Moyenne | $\bar{X} = \frac{1}{\omega} \sum_i \omega_i \bar{X}_i$ | $\bar{X}_i = \frac{1}{\omega_i} \sum_j \omega_{ij} \bar{X}_{ij}$ | $\bar{X}_{ij} = \frac{1}{\omega_{ij}} \sum_t \omega_{ijt} X_{ijt}$ | |
| Estimateur | $\hat{X} = \mu_0$ ou \bar{x} | $\hat{X}_i = Z_i \bar{X}_i + (1 - Z_i) \hat{X}$ | $\hat{X}_{ij} = Z_{ij} \bar{X}_{ij} + (1 - Z_{ij}) \hat{X}_i$ | |

3.2.1.4 Modèle de Buhlmann à trois estimateurs

Le modèle suivant a été proposé par Buhlmann en 1973. L'idée est d'utiliser la décomposition de la charge totale en fréquence et coût. En pratique, la fréquence est plus fiable que le coût et on peut être tenté de ne pas tenir compte des coûts observés. On a donc trois estimateurs du S/C d'un contrat :

$$(1) \text{ estimateur propre : } \frac{\text{somme des sinistres du contrat}}{\text{prime du contrat}}$$

$$(2) \text{ estimateur global : } \frac{\text{somme des sinistres du portefeuille}}{\text{somme des primes du portefeuille}}$$

$$(3) \text{ estimateur mixte : } \frac{\text{nombre de sinistre du contrat}}{\text{prime du contrat}} * \text{Coût moyen des sinistres du portefeuille}$$

L'idée de ce modèle est d'estimer les S/C des contrats comme barycentre de ces trois quantités.

On se place sous des hypothèses du type Buhlmann simple : le portefeuille est composé de M contrats observés pendant T années. On suppose de plus que chaque contrat possède deux paramètres de risques : Θ qui influe sur ses coûts de sinistres et Λ qui influe sur la fréquence de ces sinistres. On note X_{it} le S/C observé du contrat i l'année t , N_{it} son nombre de sinistres et C_{itk} le sinistre numéro k du contrat i l'année t . On détaille les hypothèses du modèle.

(B3 1) Les observations X_{it} sont de carré intégrables.

(B3 2) Les contrats sont indépendants :

Les vecteurs $(\Theta_i, \Lambda_i, X_{it}, N_{it}, C_{itk}, t \geq 0, k \leq N_t)$ sont indépendants.

(B3 3) Les contrats ont à priori même loi :

Les vecteurs $(\Theta_i, \Lambda_i, X_{it}, N_{it}, C_{itk}, t \geq 0, k \leq N_t)$ ont même loi.

(B3 4) L'espérance et la variance de la fréquence et coût des sinistres dépend des paramètres et pas du contrat ni de la date.

Il existe des fonctions μ_C, μ_N, σ_C et σ_N indépendantes de i, t et k telles que :

$$E(N_{it} | \Lambda_i = \lambda) = \mu_N(\lambda)$$

$$V(N_{it} | \Lambda_i = \lambda) = \sigma_N^2(\lambda)$$

$$E(C_{itk} | \Theta_i = \theta) = \mu_C(\theta)$$

$$V(C_{itk} | \Theta_i = \theta) = \sigma_C^2(\theta)$$

(B3 5) Indépendance fréquence coût :

Pour tout i , les vecteurs $(\Lambda_i, N_{it}, t \geq 0)$ et $(\Theta_i, C_{itk}, t \geq 0, k \geq 0)$ sont indépendants.

On introduit les paramètres structurels :

◇ $\mu_N = E(N)$ le nombre probable de sinistres à priori.

◇ $s_N^2 = E(V(N|\Lambda))$ la volatilité moyenne dans le temps du nombre de sinistres.

◇ $a_N = V(E(N|\Lambda))$ l'hétérogénéité du portefeuille du point de vue de la fréquence.

◇ $\mu_C = E(C)$ le coût probable à priori.

◇ $s_C^2 = E(V(C|\Theta))$ la volatilité moyenne dans le temps du coût.

◇ $a_C = V(E(C|\Theta))$ l'hétérogénéité du portefeuille du point de vue des coûts.

On note $\mu = E(X) = \frac{E(C)E(N)}{\text{prime totale}}$ l'estimateur global du S/C, $X_i = \frac{1}{T} \sum_t X_{it}$ l'estimateur propre

et $M_i = \frac{1}{T} \sum_t \frac{N_{it} * E(C)}{\text{Primes de } t}$ l'estimateur mixte. On va estimer $\mu(\Theta, \Lambda)$ par une expression du type

$\hat{X}_i = Z_1 * \bar{X}_i + Z_2 * M_i + (1 - Z_1 - Z_2) * \mu$. Les coefficients Z_1 et Z_2 doivent minimiser $E(\mu(\Theta_i, \Lambda_i) - \hat{X}_i)^2$. Z_1 est la crédibilité propre du contrat, Z_2 sa crédibilité fréquentielle et $Z = Z_1 + Z_2$ sa crédibilité.

Les valeurs des crédibilités sont données ci-dessous.

Théorème : crédibilité à trois estimateurs

L'estimateur de crédibilité est donné par $\hat{X}_i = Z_1 * \bar{X}_i + Z_2 * M_i + (1 - Z_1 - Z_2) * \mu$ avec :

$$* Z = Z_1 + Z_2 = \frac{T}{T + \frac{s_N^2}{a_N}}$$

$$* Z_1 = Z_1 + Z_2 = \frac{T}{T + \frac{\mu_N s_C^2 + s_N^2 a_C}{\mu_N^2 a_C + a_N a_C}}$$

Reste à estimer les paramètres structuraux.

On estime μ_N par $\frac{1}{MT} \sum_{it} N_{it}$.

On estime s_N^2 par $\frac{1}{M(T-1)} \sum_{it} (N_{it} - \bar{N}_i)^2$ On estime a_N par $\frac{1}{M-1} \sum_i (\bar{N}_i - \bar{N})^2$.

Les estimateurs des paramètres correspondant aux coûts (μ_C , a_C et s_C^2) sont obtenus de même en utilisant les montants de sinistres au lieu de leurs nombres.

3.2.2 Choix des poids

Dans deux des modèles que nous avons présenté, il faut munir les observations d'un poids. De nombreuses approches existent :

- on peut munir toutes les observations d'un poids identique. Le poids total d'un contrat est alors égal à son nombre d'observation. Cela revient baser la crédibilité uniquement sur l'ancienneté du contrat.
- on peut munir les observations d'un poids lié à leur ancienneté. L'idée est qu'en pratique, les contrats évoluent au cours du temps (même si nous supposons l'inverse dans le modèle). Les données plus anciennes paraissent donc moins fiables.
- on peut prendre comme poids un indicateur de la taille du contrat. L'idée est qu'en général, le S/C des gros contrats est beaucoup plus stable que celui des petits. Ils devraient donc être plus crédibles. Divers indicateurs de tailles peuvent être envisagés : nombre de véhicule, nombre d'années véhicules, primes acquises, cotisations potentielles...
- on peut prendre un poids lié à la sinistralité : nombres de sinistres ou coûts totaux.
- on peut prendre un poids proportionnel à l'exposition aux risques.

Dans le cas des flottes d'entreprise, le facteur qui influe le plus sur la stabilité des résultats est la taille du contrat. C'est donc le système de poids que nous avons choisi. Comme nous ne disposons pas du nombre exact de véhicules de certaines flottes, nous avons décidé de pondérer par les cotisations potentielles, c'est-à-dire par le montant de primes que nous nous attendons à percevoir sur l'exercice de l'observation.

Ce critère n'est pas totalement pertinent. En effet, certains contrats sont résiliés en cours d'années. Ils sont alors moins gros que ce que leur cotisation potentielle laisserait à penser. Par exemple, si un gros contrat n'existe que pendant deux mois d'une année, son S/C sur cette année risque d'être moins stable. On prendra donc la cotisation potentielle au prorata du temps pendant lequel le contrat a effectivement existé. Dans l'exemple précédent, la cotisation serait divisée par 6 pour l'observation de cette année.

Il faut aussi tenir compte du facteur ancienneté. On a donc décidé de multiplier le poids de chaque observation par un facteur dépendant de l'ancienneté. On veut un terme du type $C\lambda^n$ où n est le nombre d'années d'ancienneté. Pour choisir λ , on a déterminé les corrélations entre les S/C de divers années. On en a donné un exemple au début de ce paragraphe. Au final, l'effet ancienneté semble assez faible et nous avons choisi de prendre $\lambda = 0.9$. La constante C a été choisie de façon à ce que la somme des facteurs d'ancienneté (pour les exercices N-3 à N) vaille 1. Finalement, les coefficients ont été arrondis pour plus de lisibilité. Ils sont donnés ci-après.

Enfin, les modèles utilisés présentent un défaut : ils ne tiennent pas compte de la stabilité propre des contrats. Certains ont en effet un S/C nettement plus stable que d'autres, même à taille égale. Il paraît donc raisonnable de leur accorder une crédibilité plus grande (donc un poids plus important). On a donc décidé de multiplier le poids total d'un contrat par une constante multiplicative liée à la stabilité de son S/C. On l'a évaluée par sa variance empirique s_{ijk}^2 .

On fixe un seuil de modification maximal K : c'est-à-dire que les contrats les plus (resp moins) stables verront leur S/C majoré (resp minorés) de K pourcent. On va multiplier le contrat par une constante multiplicative $f = \phi\left(\frac{s_{ijk}^2}{s_{\text{médian}}^2}\right)$. La fonction ϕ doit vérifier quatre conditions :

→ ϕ doit être décroissante : plus un contrat est instable et moins on accorde de poids à son expérience.

→ $\phi(0) = 1 + K$: si un contrat est parfaitement stable, son poids est majoré au maximum.

→ $\lim_{+\infty} \phi(t) = 1 - K$: si un contrat est infiniment instable, son poids doit être le plus diminué possible.

→ $\phi(1) = 1$: si la stabilité d'un contrat vaut la médiane du portefeuille, son poids n'est pas modifié.

La fonction retenue est :
$$\phi(t) = \frac{(1 - K)t + (1 + K)}{t + 1}$$

On donne ci contre le graphe de la fonction et la répartition des $\phi(t)$ sur le portefeuille.

Remarque :

Pour les contrats n'ayant qu'une ou deux observations, cette méthode n'est guère pertinente et nous posons d'office $f = 1$, i.e. : on ne tient pas compte de la stabilité dans le temps.

On résume ici les poids retenus.

Poids de l'observation du contrat k l'année t :

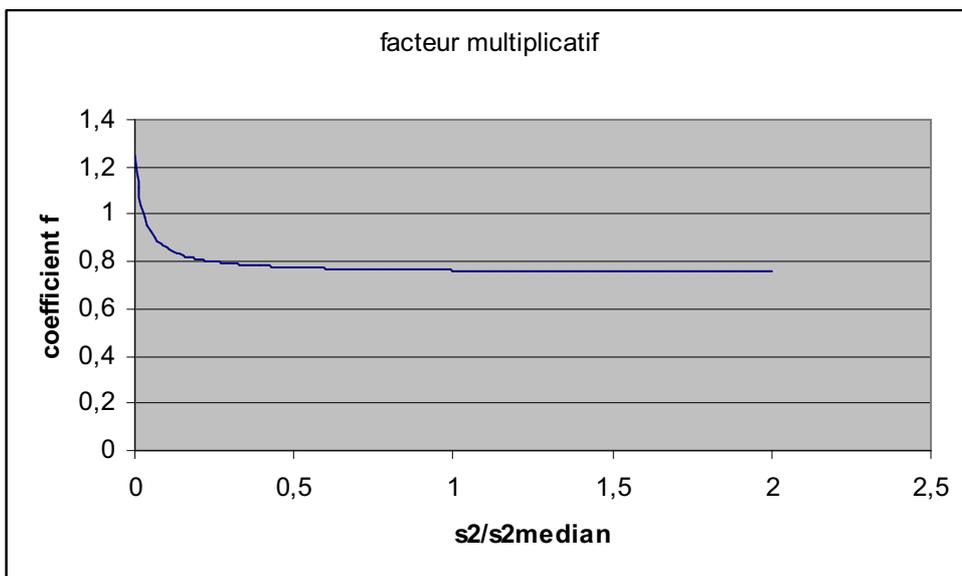
$\omega_{ijkt} = \text{Cotisation potentielle} * \text{Prorata de l'année où le contrat a existé} * \text{Facteur d'ancienneté}$

Poids total du contrat k :

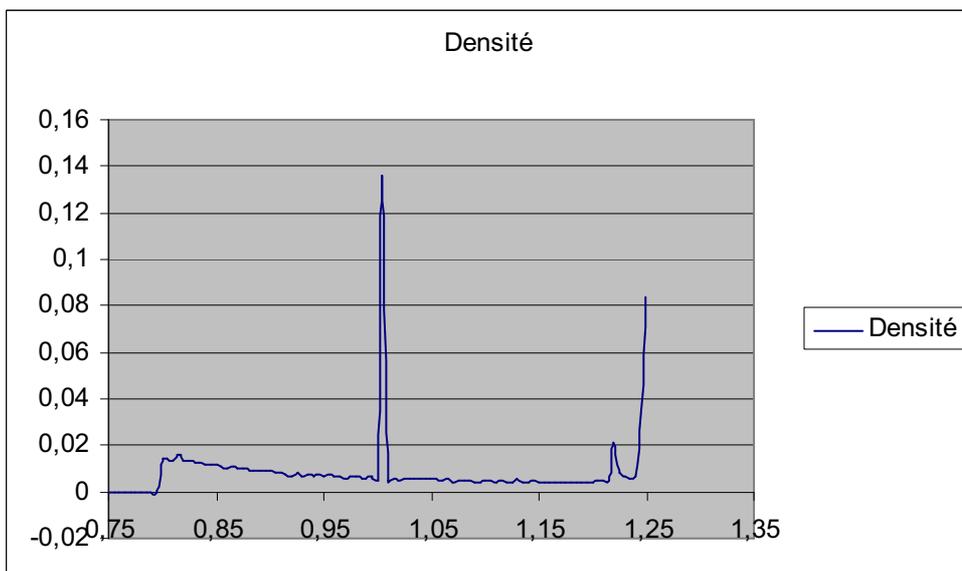
$\omega_{ijk} = (Poids_{\text{année N}} + Poids_{\text{année N-1}} + Poids_{\text{année N-2}} + Poids_{\text{année N-3}}) * \text{Facteur de stabilité}$

| Exercice | N | N-1 | N-2 | N-3 |
|----------------------|-----|------|------|-----|
| Facteur d'ancienneté | 0,3 | 0,25 | 0,25 | 0,2 |

Facteurs d'ancienneté



Graphe de la fonction phi



Répartition du coefficient f au sein du portefeuille

3.2.3 Application des modèles de crédibilité au portefeuille

On détaille à présent les résultats obtenus par les différents modèles de crédibilité.

3.2.3.1 Application de la crédibilité américaine

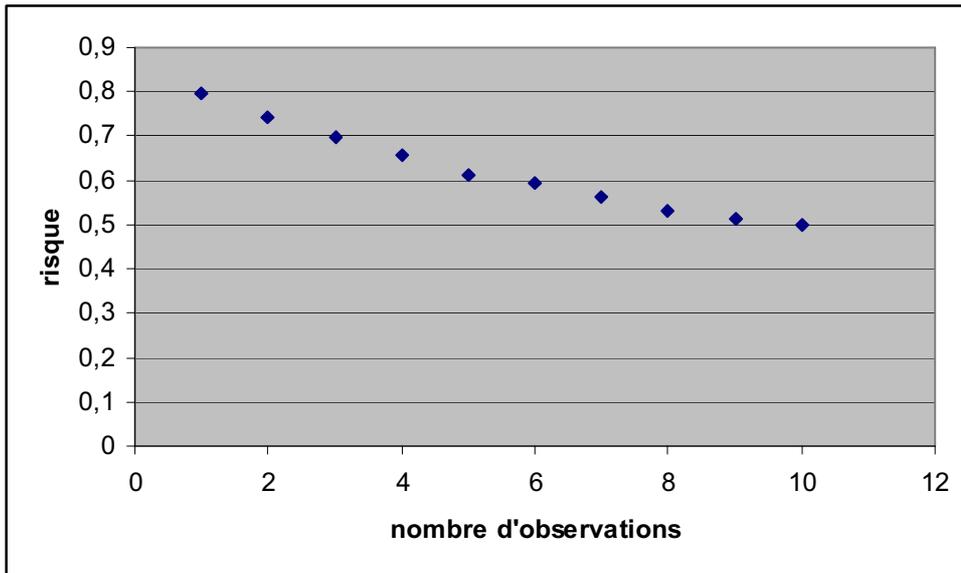
On a d'abord appliqué le modèle des fluctuations limitées au portefeuille. Dans ce modèle, on s'intéresse à la loi de $\frac{\bar{X} - E(X)}{E(X)}$ où X désigne le S/C écrêté d'un contrat et \bar{X} la moyenne des S/C écrêtés observés sur les 4 dernières années. Le problème est que cette loi est complexe à manipuler. On a utilisé deux approches :

- En utilisant l'utilitaire Monte Carlo décrit au 2.3. Cette méthode est la plus rigoureuse mais présente des défauts. Le temps de calcul reste assez long (près de 30 minutes pour l'ensemble de l'étude). De plus, la crédibilité des contrats d'une même classe ayant la même ancienneté est nécessairement identique.
- En utilisant les formules découlant de l'approximation normale. Dans ce cas, on a quand même besoin du coefficient de variation des S/C écrêtés, qui est obtenu par l'utilitaire Monte Carlo.

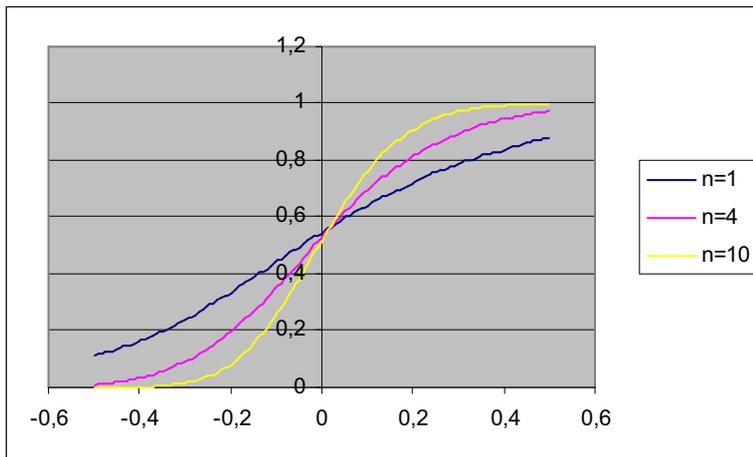
On fixe les paramètres : $\alpha = \varepsilon = 0.1$. Ainsi, la crédibilité est choisie de façon à ce qu'un contrat ne s'écarte pas de plus de 10% de sa moyenne dans 90% des cas.

Dans un premier temps, on cherche à savoir si certaines classes peuvent être pleinement crédibilisées. On a donné pour les contrats classe 6 P3T5 (les plus gros) la loi de $Y = \frac{\bar{X} - E(X)}{E(X)}$ si on dispose d'une, quatre et dix observations. Comme on pouvait s'y attendre, plus le nombre d'années est grand, plus la loi de Y est concentrée en 0. Néanmoins, les queues restent épaisses. On donne $P(|Y| > 0.1)$ pour un nombre d'observation compris entre 1 et 10. On constate que même avec des données remontant à 10 ans, près de la moitié des trajectoires s'éloignent de plus de 10% de la vraie valeur. On arrive à la même constatation pour toutes les classes de contrats. On ne peut donc pas crédibiliser pleinement les contrats.

On a ensuite déterminé le nombre d'années n_0 pour une crédibilité. Le nombre d'années d'observation nécessaires varie de 60 à plus de 2500 ans. Ces valeurs sont considérables, et n'ont bien sûr guère de sens en pratique. Les crédibilités sont par conséquent très faibles, et ne dépassent jamais 0.25. Ces valeurs sont manifestement trop basses, et ne permettraient pas de distinguer les bons contrats des mauvais. Dans le cas des classes 1 (près de la moitié du portefeuille), la crédibilité ne dépasse jamais 4% (6 avec les approximations normales). Ainsi, le S/C d'un contrat ne dépendrait pratiquement que du S/C moyen, et on ne pourrait pas identifier les contrats posant problème. On remarque par ailleurs que les deux méthodes donnent des résultats assez différents. L'approximation normale paraît cependant discutable car on ne fait la moyenne que de 4 observations. On constate qu'elle conduit à sous-estimer les fluctuations du S/C.



Risque de s'écarter de plus de 10% de la vraie valeur en fonction du nombre d'années d'observation



Loi de l'écart relatif entre la moyenne des coûts totaux et son espérance

| | | M3T5 | | | | | | P3T5 | | | | | |
|---|----------------|------|------|------|------|------|------|------|------|------|------|------|------|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 |
| | n0 | 2615 | 2124 | 1234 | 623 | 138 | 61 | 2321 | 1830 | 1256 | 402 | 142 | 63 |
| Z | 1 observation | 0,02 | 0,02 | 0,03 | 0,04 | 0,09 | 0,13 | 0,02 | 0,02 | 0,03 | 0,05 | 0,08 | 0,13 |
| | 2 observations | 0,03 | 0,03 | 0,04 | 0,06 | 0,12 | 0,18 | 0,03 | 0,03 | 0,04 | 0,07 | 0,12 | 0,18 |
| | 3 observations | 0,03 | 0,04 | 0,05 | 0,07 | 0,15 | 0,22 | 0,04 | 0,04 | 0,05 | 0,09 | 0,15 | 0,22 |
| | 4 observations | 0,04 | 0,04 | 0,06 | 0,08 | 0,17 | 0,26 | 0,04 | 0,05 | 0,06 | 0,10 | 0,17 | 0,25 |

| | | M3T5 | | | | | | P3T5 | | | | | |
|---|----------------|------|-------|-------|-------|-------|-------|------|------|-------|-------|-------|------|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 |
| | n0 | 1053 | 505,2 | 320,9 | 160,7 | 109,8 | 57,84 | 1212 | 713 | 439,2 | 257,8 | 152,2 | 117 |
| Z | 1 observation | 0,03 | 0,04 | 0,06 | 0,08 | 0,10 | 0,13 | 0,03 | 0,04 | 0,05 | 0,06 | 0,08 | 0,09 |
| | 2 observations | 0,04 | 0,06 | 0,08 | 0,11 | 0,13 | 0,19 | 0,04 | 0,05 | 0,07 | 0,09 | 0,11 | 0,13 |
| | 3 observations | 0,05 | 0,08 | 0,10 | 0,14 | 0,17 | 0,23 | 0,05 | 0,06 | 0,08 | 0,11 | 0,14 | 0,16 |
| | 4 observations | 0,06 | 0,09 | 0,11 | 0,16 | 0,19 | 0,26 | 0,06 | 0,07 | 0,10 | 0,12 | 0,16 | 0,18 |

Valeur des crédibilités suivant la méthode Monte Carlo (tableau 1) et loi normale (tableau 2)

3.2.3.2 Utilisation du modèle de Buhlmann Straub

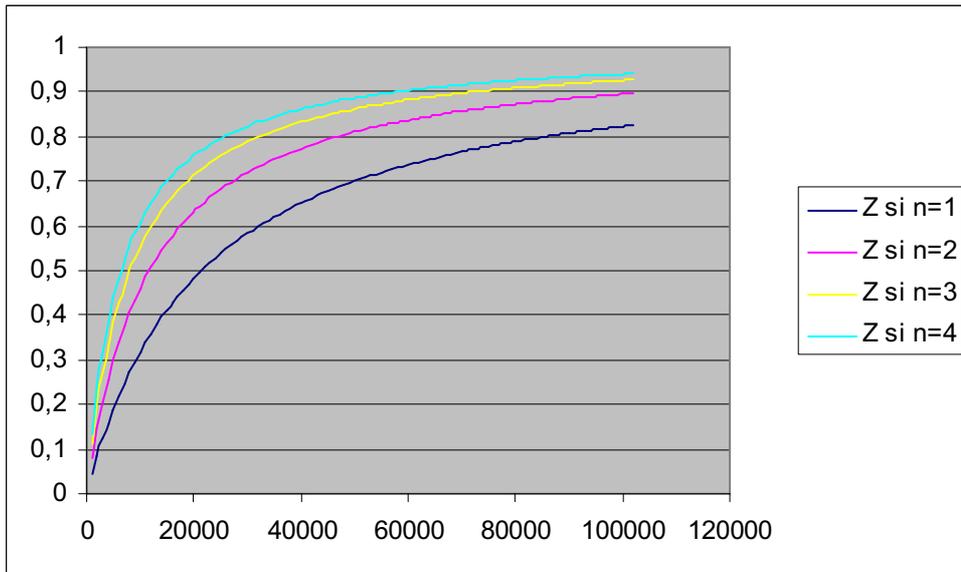
On a appliqué le modèle de Buhlmann Straub au portefeuille. La crédibilité Z d'un contrat est $\frac{\omega}{\omega + 6424}$ où ω est le poids du contrat. Le S/C final est $ZS/C_{\text{contrat}} + (1 - Z)S/C_{\text{portefeuille}}$. A titre d'illustration, pour un contrat ayant au moins 4 ans d'existence, l'expérience propre est prépondérante (i.e. : $Z > 0.5$) si la cotisation annuelle est supérieure à 6424 euros soit 45% du portefeuille. (en pratique, compte tenu de la baisse des poids pour les contrats instables, ce pourcentage est de l'ordre de 30%). On constate que dans la majorité des cas, c'est le ratio du portefeuille qui est donc prépondérant. Cela vient du fait que les S/C individuels sont extrêmement instables, et qu'il est donc dangereux de se baser sur eux.

On a donné le graphe de la crédibilité des contrats en fonction de leur cotisation annuelle et du nombre d'années d'existence. Le résultat paraît satisfaisant. La crédibilité croît aussi bien avec la taille qu'avec l'ancienneté. Elle tend vers 0 pour les très petits contrats, ce qui est souhaitable car ils ont des S/C particulièrement volatiles. Pour les très gros contrats, la crédibilité tend vers 1. Par exemple, elle dépasse 95% au bout de 4 ans pour les contrats de plus de plus de 122000 euros et dès la première année pour ceux de plus de 406000 euros (en pratique, compte tenu de la plus grande stabilité des gros contrats, ces seuils sont sans doute plus bas). Il est courant dans les modèles de crédibilité que les plus gros contrats soient presque entièrement crédibilisés. En l'occurrence, on peut néanmoins s'interroger. On a vu en effet que le S/C d'un gros contrat pouvait fluctuer de façon notable (cf. le paragraphe sur la crédibilité américaine et les calculs de corrélation en introduction du 3.2).

On donne pour chaque classe la moyenne des S/C et des crédibilités sur le portefeuille. En ce qui concerne les crédibilités, cela confirme les observations ci-dessus : les plus petits contrats sont très peu crédibilisés, alors que les plus gros le sont presque entièrement (respectivement 90 et 96 % de moyenne pour les deux dernières classes). L'étendue de la variable Z est grande, surtout pour les premières classes. Cela vient de plusieurs facteurs. Au sein d'une même classe les contrats peuvent avoir des cotisations assez différentes (la cotisation maximale de la classe 1 est de 5000 euros mais près du tiers de ces contrats comptent moins de 1000 euros). On a également modifié les poids de certains contrats pour tenir compte de leur plus ou moins grande stabilité. Enfin, le facteur ancienneté joue. Pour un contrat taille 1, la crédibilité moyenne est de 7% au bout d'un an mais de 25% au bout de 4 ans. Un problème apparaît néanmoins. Comme on s'y attendait, le S/C des contrats évoluent avec la taille. Or, le modèle de Buhlmann Straub ne prend pas en compte ce facteur. Cela nous incite à recourir plutôt à un modèle hiérarchique.

| K | s2 | a |
|------|-----|------------|
| 6424 | 619 | 0,09635741 |

Paramètres du modèle de Buhlmann Straub



Crédibilité en fonction de la cotisation potentielle

| | CATEGORIE | SCMOYEN | SCMIN | SCMAX | ZMOYEN | ZMIN | ZMAX |
|------|-----------|---------|---------|---------|---------|---------|---------|
| M3T5 | 1 | 0.56146 | 0.26685 | 1.31550 | 0.19255 | 0.00060 | 0.78083 |
| | 2 | 0.56743 | 0.30157 | 2.04328 | 0.40563 | 0.00203 | 0.82878 |
| | 3 | 0.58782 | 0.17357 | 4.15650 | 0.58847 | 0.01198 | 0.95006 |
| | 4 | 0.61972 | 0.18005 | 1.96186 | 0.77902 | 0.06812 | 0.93816 |
| | 5 | 0.69544 | 0.15774 | 1.91703 | 0.91221 | 0.65891 | 0.98134 |
| | 6 | 0.74563 | 0.36956 | 1.39625 | 0.97254 | 0.87863 | 0.99565 |
| P3T5 | 1 | 0.55658 | 0.36484 | 1.24403 | 0.19935 | 0.00061 | 0.73708 |
| | 2 | 0.54893 | 0.27068 | 1.71122 | 0.41034 | 0.01974 | 0.79553 |
| | 3 | 0.57238 | 0.25305 | 1.65599 | 0.59119 | 0.01152 | 0.86680 |
| | 4 | 0.63968 | 0.20352 | 1.82133 | 0.78579 | 0.09187 | 0.95063 |
| | 5 | 0.70486 | 0.18867 | 2.23922 | 0.90462 | 0.31073 | 0.97487 |
| | 6 | 0.77798 | 0.18734 | 1.54305 | 0.97155 | 0.87532 | 0.99635 |

Résultats obtenus sur le portefeuille

3.2.3.3 Modèle de Jewell à un niveau intermédiaire

On applique à présent au portefeuille le modèle de Jewell à un niveau intermédiaire. On décompose le portefeuille en 12 catégories, en fonction de la taille des flottes et du type de véhicule. On dispose pour chaque contrat d'un S/C noté S/C_{contrat} . Le S/C moyen d'une catégorie est $S/C_{\text{catégorie}}$.

On estime le S/C d'une catégorie par $\widehat{S/C}_{\text{catégorie}} = Z_{\text{cat}}S/C_{\text{catégorie}} + (1 - Z_{\text{cat}})S/C_{\text{portefeuille}}$ avec $Z_{\text{cat}} = \frac{\omega}{\omega + 8.11}$ et ω le poids de la catégorie.

On estime le S/C d'un contrat par $\widehat{S/C}_{\text{contrat}} = Z_{\text{contrat}}S/C_{\text{contrat}} + (1 - Z_{\text{contrat}})\widehat{S/C}_{\text{catégorie}}$ avec $Z_{\text{contrat}} = \frac{\omega}{\omega + 6907}$ et ω le poids du contrat.

On constate que les crédibilités sont plus faibles que celles du modèle BS. Cela vient du fait que le complément du S/C du contrat est plus précis : on a utilisé la moyenne de la catégorie plutôt que celle du portefeuille. Comme ce complément est plus fiable (les catégories sont plus homogènes), le poids qui lui est associé (1-crédibilité) est plus grand.

Une autre conséquence de ce changement de complément est que l'écart entre les S/C moyens est beaucoup plus net. Dans le modèle BS, le fait d'effectuer un barycentre avec la moyenne du portefeuille reportait une partie du coût des gros contrats (souvent déficitaire ou à peine rentable) sur les petits (en général plus rentables). Concrètement, les petits contrats paient pour les gros. Dans le modèle de Jewell, ce n'est pas le cas (en fait, il y a une petite mutualisation via le calcul des $\widehat{S/C}_{\text{catégories}}$ mais elle pèse très peu). Du coup, les petits contrats sont en moyenne bien meilleurs que dans le modèle précédent.

Enfin, on remarque que les crédibilités Z_{cat} des catégories sont toutes très proches de 1 (sauf pour les tailles 6). En première approximation, on peut donc négliger la mutualisation à ce niveau. On a alors :

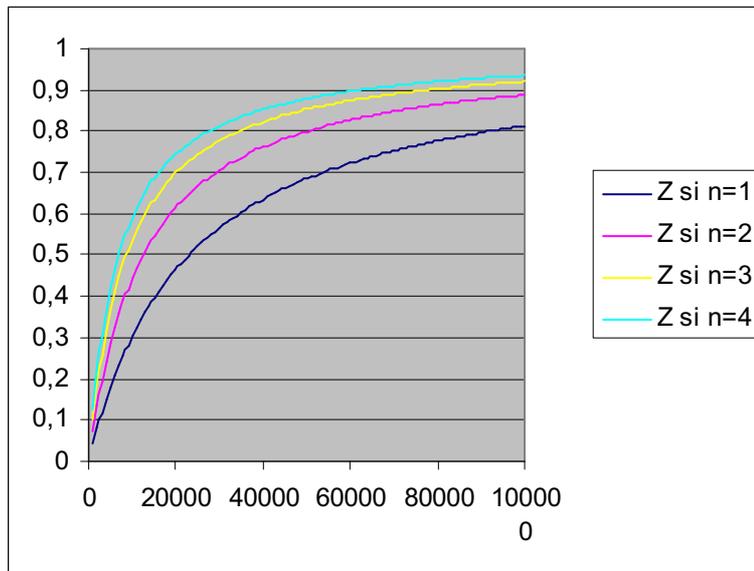
$$S/C_{\text{contrat}} \sim Z_{\text{contrat}}S/C_{\text{contrat}} + (1 - Z_{\text{contrat}})S/C_{\text{catégorie}} \text{ avec } Z_{\text{contrat}} = \frac{\omega}{\omega + 6907}$$

En somme, le modèle de Jewell est obtenu à partir du modèle BS en changeant deux points :

- une légère baisse des crédibilités (passage de K de 6424 à 6907)
- l'utilisation comme complément de la moyenne des catégories plutôt que celle du portefeuille

| s2 | a | b | K | Kcatégorie |
|-----|--------|---------|------|------------|
| 619 | 0,0896 | 0,01104 | 6907 | 8,11 |

Paramètres du modèle de Jewell à un niveau intermédiaire



Crédibilité en fonction de la cotisation potentielle et du nombre n d'années d'ancienneté

| | CATEGORIE | SCMOYEN | SCMIN | SCMAX | ZMOYEN | ZMIN | ZMAX |
|------|-----------|---------|---------|---------|---------|---------|---------|
| M3T5 | 1 | 0,4714 | 0,24614 | 1,17015 | 0,18259 | 0,00056 | 0,76817 |
| | 2 | 0,53515 | 0,2899 | 1,98505 | 0,38967 | 0,00189 | 0,81825 |
| | 3 | 0,57517 | 0,17358 | 3,98487 | 0,57246 | 0,01115 | 0,94651 |
| | 4 | 0,61638 | 0,18236 | 1,90479 | 0,76744 | 0,06366 | 0,93382 |
| | 5 | 0,69923 | 0,16917 | 1,90986 | 0,90655 | 0,64244 | 0,97996 |
| | 6 | 0,74602 | 0,37016 | 1,39551 | 0,97059 | 0,87068 | 0,99533 |
| P3T5 | 1 | 0,44553 | 0,3185 | 1,11004 | 0,18921 | 0,00057 | 0,72279 |
| | 2 | 0,48848 | 0,25176 | 1,6031 | 0,39453 | 0,01838 | 0,78348 |
| | 3 | 0,55373 | 0,249 | 1,61236 | 0,5756 | 0,01073 | 0,85821 |
| | 4 | 0,64554 | 0,20817 | 1,81195 | 0,77464 | 0,086 | 0,94712 |
| | 5 | 0,7117 | 0,19234 | 2,22626 | 0,89873 | 0,29542 | 0,97303 |
| | 6 | 0,78016 | 0,18832 | 1,54383 | 0,96951 | 0,86719 | 0,99607 |

Résultats obtenus sur le portefeuille

| M3T5 | | | | | | P3T5 | | | | | |
|-------|-------|-------|-------|-------|-------|-------|------|-------|-------|-------|-------|
| 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 |
| 0,997 | 0,997 | 0,997 | 0,992 | 0,967 | 0,764 | 0,987 | 0,99 | 0,994 | 0,994 | 0,982 | 0,893 |

Crédibilité des catégories

3.2.3.4 Modèle de Jewell à deux niveaux intermédiaires

On applique le modèle de Jewell à deux niveaux intermédiaires au portefeuille. Celui-ci est décomposé en 6 classes en fonction de la cotisation potentielle, et chacune des classes est divisée en deux catégories : M3T5 et P3T5.

Les crédibilités à chaque étage (contrat, classe, catégorie) sont données par :

$Z = \frac{poids}{poids + K}$. Les valeurs de K à chaque étage sont données ci contre.

On estime le S/C d'une classe par $\widehat{S/C}_{classe} = ZS/C_{classe} + (1 - Z)S/C_{portefeuille}$.

On estime le S/C d'une catégorie par $\widehat{S/C}_{categorie} = ZS/C_{categorie} + (1 - Z)\widehat{S/C}_{classe}$.

On estime le S/C d'un contrat par $\widehat{S/C}_{contrat} = ZS/C_{contrat} + (1 - Z)\widehat{S/C}_{categorie}$.

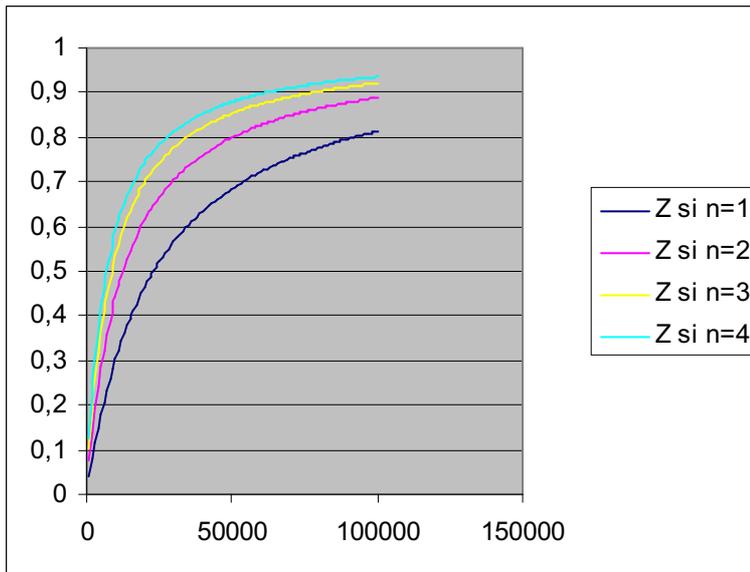
Les résultats obtenus par ce modèle sont très proches de ceux du précédent. En particulier, la valeur de K est pratiquement la même. Les crédibilités des contrats sont donc identiques. Les valeurs de S/C sont globalement similaires à celles du modèle précédent. On constate à nouveau que le S/C se dégrade très rapidement avec la taille des contrats.

Les classes sont globalement très crédibles (de l'ordre de 90%). Les catégories le sont encore plus (presque toujours au dessus de 95%). Les seules exceptions sont les catégories de la classe 6. Leur moindre crédibilité vient du faible nombre de contrats. En somme, pour ces contrats, le complément est un mélange des S/C classe 6 M3T5 et Classe 6 P3T5 (plus un petit rôle joué par ceux des autres classes, via le S/C du portefeuille).

Au final, ce modèle donne des résultats qui ne diffèrent guère de ceux du précédent. Les crédibilités sont les mêmes et les compléments diffèrent peu, sauf pour les contrats classe 6. De plus, ces très gros contrats ayant une crédibilité proche de 1, le complément joue très peu pour eux.

| s2 | a | b | c | K | Kcatégorie | Kclasse |
|-----|--------|--------|-------|------|------------|---------|
| 619 | 0,0896 | 0,0025 | 0,012 | 6907 | 35,6 | 0,2 |

Paramètres du modèle de Jewell à deux niveaux intermédiaires



Crédibilité en fonction de la cotisation potentielle et du nombre n d'années d'ancienneté

| | CATEGORIE | SCMOYEN | SCMIN | SCMAX | ZMOYEN | ZMIN | ZMAX |
|------|-----------|---------|---------|---------|---------|---------|---------|
| M3T5 | 1 | 0,47133 | 0,24612 | 1,17007 | 0,18259 | 0,00056 | 0,76817 |
| | 2 | 0,53505 | 0,28983 | 1,98498 | 0,38967 | 0,00189 | 0,81825 |
| | 3 | 0,57513 | 0,17357 | 3,98481 | 0,57246 | 0,01115 | 0,94651 |
| | 4 | 0,61658 | 0,18248 | 1,90526 | 0,76744 | 0,06366 | 0,93382 |
| | 5 | 0,69945 | 0,16952 | 1,91024 | 0,90655 | 0,64244 | 0,97996 |
| | 6 | 0,7467 | 0,37048 | 1,3961 | 0,97059 | 0,87068 | 0,99533 |
| P3T5 | 11 | 0,44651 | 0,31904 | 1,11092 | 0,18921 | 0,00057 | 0,72279 |
| | 12 | 0,48943 | 0,25213 | 1,60404 | 0,39453 | 0,01838 | 0,78348 |
| | 13 | 0,55383 | 0,24905 | 1,61242 | 0,5756 | 0,01073 | 0,85821 |
| | 14 | 0,64545 | 0,20814 | 1,81189 | 0,77464 | 0,086 | 0,94712 |
| | 15 | 0,71162 | 0,19231 | 2,22602 | 0,89873 | 0,29542 | 0,97303 |
| | 16 | 0,77996 | 0,18826 | 1,54357 | 0,96951 | 0,86719 | 0,99607 |

Résultats obtenus sur le portefeuille

| Classe | 1 | | 2 | | 3 | | 4 | | 5 | | 6 | |
|-------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Catégorie | M3T5 | P3T5 |
| Z catégorie | 0,985 | 0,943 | 0,985 | 0,957 | 0,985 | 0,976 | 0,965 | 0,975 | 0,87 | 0,926 | 0,424 | 0,656 |
| Z classe | 0,906 | | 0,906 | | 0,907 | | 0,906 | | 0,899 | | 0,844 | |

Crédibilité des classes et des catégories

3.2.3.5 Modèle à trois estimateurs

On applique à présent le modèle de Buhlmann à trois estimateurs. On rappelle que l'on estime le S/C d'un contrat par :

$$S/C = Z_1 * S/C_{\text{contrat}} + Z_2 * S/C_{\text{fréquentiel}} + (1 - Z_1 - Z_2) * S/C_{\text{portefeuille}}$$

$$\text{où } S/C_{\text{fréquentiel}} = \frac{\text{Nombre de sinistres du contrat}}{\text{Prime du contrat}} * \text{Coût moyen des sinistres du portefeuille.}$$

Les deux crédibilités sont données par $Z_1 = \frac{T}{T + K'}$ et $Z_1 + Z_2 = \frac{T}{T + K}$ Où T est le nombre d'observations du contrat. On donne ci contre les paramètres obtenus pour le portefeuille.

Tout d'abord, ce modèle ne tiens pas compte du système de poids. Par conséquent, les gros contrats ne sont pas plus crédibilisés que les petits. Le niveau de crédibilité est moyen, allant de 23% pour un nouveau contrat à 55 pour ceux existant depuis plus de trois ans. Ces niveaux paraissent trop élevés pour les petits contrats, et sans doute trop faibles pour les gros.

On remarque que l'estimateur $S/C_{\text{fréquentiel}}$ n'est pratiquement pas utilisé. La crédibilité correspondante reste inférieure à 1% même pour les contrats les plus anciens. Cet estimateur présente un défaut : les coûts de sinistres sont très volatils et à queue épaisse. La moyenne correspond au quantile d'ordre 0.8. Ainsi, pour le calcul du $S/C_{\text{fréquentiel}}$, les sinistres contribuent chacun à un coût important. Pour un contrat de 1000 euros de prime, chaque sinistre augmente le $S/C_{\text{fréquentiel}}$ de près de 1.5. Cet estimateur s'avère donc au final dangereux et son poids est presque nul.

On peut comparer les résultats de ce modèle à ceux du modèle BS. Les deux utilisent le même complément : le S/C global du portefeuille (on néglige le $S/C_{\text{fréquentiel}}$ car son apport est très faible). Ils diffèrent par le choix des poids, donc par les crédibilités. Le modèle BS pondère par la taille sur l'ensemble de la période et le modèle à trois estimateurs par l'ancienneté. BS crédibilise donc moins les petits contrats et plus les gros. Au final, les S/C issus du modèle à trois estimateurs sont légèrement plus faibles. On notera que dans le modèle à trois estimateurs certains contrats ont des S/C beaucoup plus élevés (parfois supérieur à 5). Cela vient de la forte crédibilité des petits contrats.

| s2N | aN | s2C | aC | K | K' |
|-------|--------|--------|-------|------|------|
| 0,575 | 0,1725 | 684172 | 91223 | 3,35 | 3,43 |

Paramètres du modèle de Buhlmann à trois estimateurs

| Nombre d'observations | Z1 | Z2 |
|-----------------------|--------|--------|
| 1 | 0,2299 | 0,0042 |
| 2 | 0,3738 | 0,0055 |
| 3 | 0,4724 | 0,0059 |
| 4 | 0,5442 | 0,0059 |

Crédibilité en fonction de l'ancienneté

| | | SC MOYEN | SC MIN | SC MAX | Z1 MOYEN | Z1 MIN | Z1 MAX | Z2 MOYEN | Z2 MIN | Z2 MAX |
|------|---|-------------|-----------|-----------|-------------|-----------|-----------|-------------|-----------|-----------|
| M3T5 | 1 | 0,4784 | 0,3147 | 5,376 | 0,46 | 0,23 | 0,54 | 0,0055 | 0,0042 | 0,0059 |
| | 2 | 0,5208 | 0,2996 | 2,2536 | 0,47 | 0,23 | 0,54 | 0,0056 | 0,0042 | 0,0059 |
| | 3 | 0,5367 | 0,3063 | 1,4195 | 0,48 | 0,23 | 0,54 | 0,0057 | 0,0042 | 0,0059 |
| | 4 | 0,55 | 0,2745 | 1,1502 | 0,48 | 0,23 | 0,54 | 0,0057 | 0,0042 | 0,0059 |
| | 5 | 0,5786 | 0,3589 | 1,0005 | 0,48 | 0,23 | 0,54 | 0,0057 | 0,0042 | 0,0059 |
| | 6 | 0,5996 | 0,439 | 0,8699 | 0,49 | 0,23 | 0,54 | 0,0057 | 0,0042 | 0,0059 |
| P3T5 | 1 | 0,4766 | 0,3297 | 3,8798 | 0,46 | 0,23 | 0,54 | 0,0056 | 0,0042 | 0,0059 |
| | 2 | 0,4991 | 0,3095 | 1,9362 | 0,46 | 0,23 | 0,54 | 0,0056 | 0,0042 | 0,0059 |
| | 3 | 0,5275 | 0,3032 | 1,4889 | 0,47 | 0,23 | 0,54 | 0,0056 | 0,0042 | 0,0059 |
| | 4 | 0,5658 | 0,3222 | 1,2608 | 0,48 | 0,23 | 0,54 | 0,0057 | 0,0042 | 0,0059 |
| | 5 | 0,5956 | 0,3302 | 1,9637 | 0,48 | 0,23 | 0,54 | 0,0057 | 0,0042 | 0,0059 |
| | 6 | 0,618 | 0,3477 | 0,9657 | 0,48 | 0,23 | 0,54 | 0,0057 | 0,0042 | 0,0059 |

Résultat obtenu sur le portefeuille

3.2.4 Choix du modèle

Après avoir appliqué les différents modèles au portefeuille, il reste à choisir celui que nous allons utiliser.

◇ Le modèle des fluctuations limitées conduit à des crédibilités très faibles. C'est à priori assez pertinent vu la forte volatilité des S/C (en particulier il est sans doute discutable d'avoir des crédibilités proche de 1, même pour les plus gros contrats). Néanmoins, dans notre cas, cela pose un problème. En effet, nous cherchons à distinguer les contrats sous tarifés des autres. Avec des niveaux de crédibilité aussi faibles, l'expérience propre ne joue pas un grand rôle et il est donc difficile de repérer les contrats devant être majoré.

◇ Le modèle de Buhlmann Straub possède un gros défaut. Le complément utilisé, le S/C moyen du portefeuille, est le même pour tous les contrats. Pourtant les différentes segmentations (surtout celle par la taille), influent sur la rentabilité. Au final, le ratio des petits contrats est donc surestimé et celui des gros contrats sous-estimé. Cela joue en particulier sur la proportion de contrats repérés. Comme la plupart des contrats sont des taille 1 ou 2, qui sont défavorisés par le complément, le taux de repérés est supérieur à ceux obtenus par les méthodes de Jewell (17.5% contre 11). Or pour des raisons pratiques, nous souhaitons un pourcentage proche de 10%. Au final, le modèle BS ne paraît donc pas optimal.

◇ Le modèle à trois estimateurs s'avère décevant. Son principal intérêt à priori était d'introduire un estimateur ne dépendant que de la fréquence, qui est plus stable que la charge totale. Cet estimateur s'avère médiocre, et son poids est infime. Pour le reste, le modèle est très grossier, il ne prend pas en compte la taille des flottes, et crédibilise uniquement en fonction de l'ancienneté. Nous ne l'avons donc pas retenu.

◇ Reste donc les deux modèles de type Jewell. Le modèle à deux niveaux intermédiaires est à priori plus précis, mais dans la pratique l'ajout du niveau supplémentaire ne modifie guère les résultats. Les S/C obtenus par les deux méthodes sont très proches. On a donc décidé de conserver le modèle le plus simple à savoir celui à un seul niveau intermédiaire.

3.3 Description du processus de majoration QP

On va à présent décrire les différentes étapes du processus de majoration QP. Dans un premier temps, on calcule le ratio S/C de chaque contrat pour l'année en cours et les années N-1, N-2 et N-3 (si le contrat existait ces années là). Ce calcul nécessite la mise sous forme as if des primes et des charges de sinistres ainsi que l'écrêtement et le développement à l'ultime des coûts. Ensuite, on en déduit le S/C historique comme une moyenne pondérée des S/C des différentes années. Pour les contrats présentant une tendance nette (contrat en amélioration ou en dégradation nette), on modifie le S/C en anticipant son évolution. Enfin, on obtient le S/C final en crédibilisant le S/C historique.

Pour qu'un contrat soit rentable, il faut qu'il coûte moins qu'il ne rapporte. On matérialise cela en demandant que le S/C soit inférieur à un certain seuil. On peut se demander si ce seuil doit être uniforme sur le portefeuille ou pas. Typiquement, certaines classes de contrats sont moins rentables que d'autres. Cela signifie que les règles de tarification initiale diffèrent. Il paraît donc logique de différencier aussi celles du processus QP. On va donc définir un seuil d'équilibre variant avec la taille de la flotte. Les contrats dont le S/C dépasse ce niveau seront jugés non rentables et majorés. La majoration sera celle qui appliquée il y a un an aurait permis de les amener à l'équilibre.

On va tout d'abord détailler le calcul du S/C, en particulier du traitement des primes et des coûts de sinistres (l'écrêtement et la crédibilité ayant fait l'objet de chapitres spécifiques). Ensuite, on étudiera le choix du seuil d'équilibre, enfin on affinera le modèle en réglant les divers paramètres de façon à obtenir des résultats les plus exploitables possible.

3.3.1 Calcul du S/C par années

Dans un premier temps, il faut déterminer le ratio S/C du contrat pour chaque année depuis N-3. Il faut donc trouver la charge de sinistres et le montant des primes du contrat pour ces années.

★ Montant des primes :

On utilisera les primes acquises à l'exercice. Ce sont les primes servant à couvrir les sinistres survenus au cours de l'année.

Un problème se pose pour les contrats ayant déjà fait l'objet de majorations. Par exemple, considérons un contrat qui a eu de mauvais résultats les années N-3 et N-2, puis à fait l'objet d'une majoration qui l'a ramené à l'équilibre. Dans le calcul du S/C historique, on continuerait à prendre en compte les mauvais ratios N-3 et N-2 alors que le problème a déjà été corrigé.

Pour éviter ce biais, on applique à la prime les majorations du contrat. Par exemple, si un contrat a été majoré de 10 % depuis N - 3, sa prime N - 3 est augmentée de 10%. Concrètement, on pose :

Prime acquise as if (année i) = prime acquise (année i) * $\frac{CT_{\text{année N}}}{CT_{\text{année i}}}$

où CT_i est le coefficient technique du contrat pour l'année i , c'est à dire l'écart entre sa prime et le tarif facial.

★ Montant par sinistre :

Chaque sinistre est affecté au calcul du S/C de son année de survenance, même si les remboursements s'étalent sur plusieurs années (c'est typiquement le cas pour les sinistres corporels).

Les coûts des sinistres sont séparés en trois grands pôles : RC corporel, RC matérielle et dommage.

La valeur prise en compte est le coût total probable :

Coût total probable = Coût payé + Prévision de versements - Recours - Prévision de recours

Le problème est que les prévisions peuvent être sur ou sous estimées. On va donc développer à l'ultime ces coûts. Les facteurs de développement sont donnés ci contre. Ils ont été obtenus par une méthode de type « chain ladder » appliquée au triangle des coûts + provisions.

L'inflation est ensuite prise en compte. Les taux retenus sont donnés ci contre.

Le coût total du sinistre est la somme des coûts par type de dommage :

Coût = coût RC corp + coût RC mat + coût DOMM.

Enfin, le sinistre est écrêté deux fois suivant la méthode décrite au paragraphe 3.1.

★ Calcul du S/C de l'année

Au final, le S/C est obtenu par la formule donnée en 3.1 :

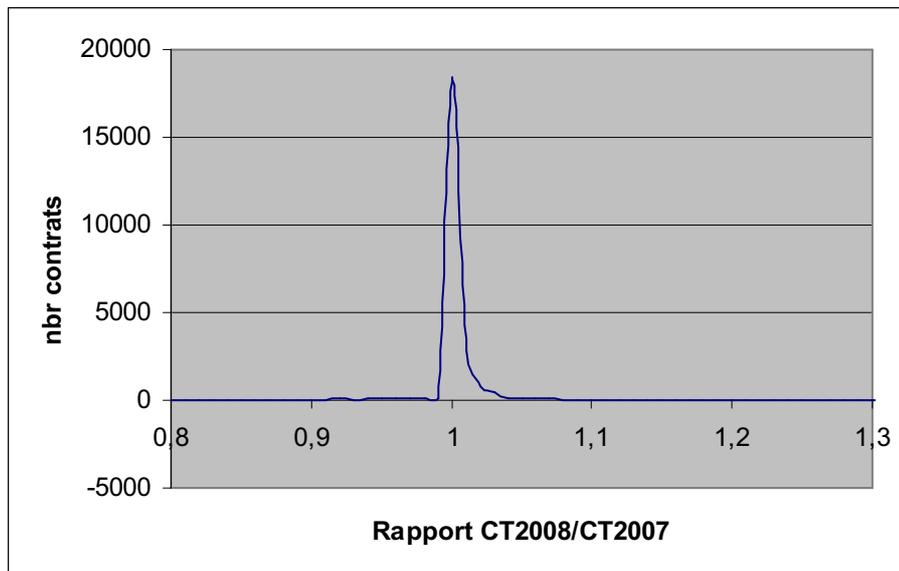
$$S/C = \frac{\sum_{\text{contrat}} \text{coûts non écrétés}}{\text{Prime du contrat}} + \frac{\sum_{\text{catégorie}} \text{coût écrêté deuxième seuil}}{\text{Primes de la catégorie}} + \frac{\sum_{\text{portefeuille}} \text{coût écrêté premier seuil}}{\text{Primes du portefeuille}}$$

| Exercice de survenance | RC corporel | RC matérielle | Domage |
|------------------------|-------------|---------------|--------|
| N | 1,3 | 1,02 | 0,98 |
| N-1 | 1,2 | 0,95 | 1 |
| N-2 | 1,01 | 0,99 | 1 |
| N-3 | 1,005 | 1 | 1 |

Facteurs de développement

| Type de garantie | Taux d'inflation |
|------------------|------------------|
| RC corporel | 5% |
| RC matérielle | 1% |
| Domage | 1% |

Taux d'inflation



Loi du rapport CT2008/CT2007

3.3.2 Calcul du S/C historique et du S/C crédibilisé

Une fois que l'on dispose du S/C pour chaque année, le S/C historique se calcule facilement :

$$S/C_{historique} = \frac{\sum_i \text{poids}_{\text{année } i} * S/C_{\text{année } i}}{\sum_i \text{poids}_{\text{année } i}}$$

On rappelle que les poids utilisés sont ceux définis au 3.2, c'est-à-dire les cotisations potentielles au prorata du temps d'existence du contrat, multiplié par un facteur lié à sa stabilité.

Remarque :

Conventionnellement, si un contrat n'existait pas pendant une année, son S/C pour cette année vaut 0. Cette valeur n'a pas d'importance car le poids qui lui est affecté est nul.

Le défaut de cette méthode de calcul est qu'il ne tient pas compte des évolutions au cours du temps, par exemple du fait qu'un contrat puisse être en amélioration ou en dégradation. Le S/C obtenu va donc être modifié pour les contrats en hausse (resp en baisse) constante sur les trois dernières observations. Dans ce cas, on calcule la hausse h (resp la baisse b) annuelle moyenne du S/C. On augmente le ratio obtenu de $C * h$ (resp diminue de $C * b$), où C est une constante à déterminer.

Par exemple, prenons $C = 0.75$. Si le S/C annuel d'un contrat augmente de 20% par an en moyenne, le S/C historique sera augmenté de 15%. De même, si le S/C annuel baisse de 20% par an en moyenne, le S/C historique sera diminué de 15%.

Dans quelques cas extrêmes, la hausse peut atteindre des niveaux très élevés. On a donc choisi de la limiter à 30 %.

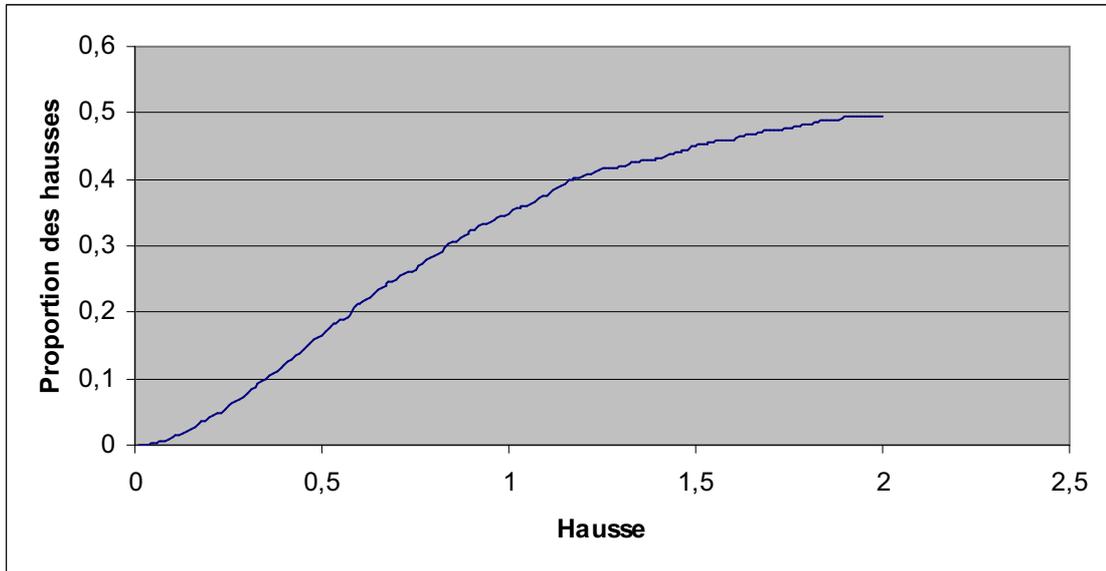
On donne ci contre la répartition des hausses et des baisses de S/C. On remarque que les hausses sont en moyenne plus importantes que les baisses. Elles concernent néanmoins une proportion plus faible de contrats.

Enfin, le S/C historique est crédibilisé pour obtenir le S/C final.

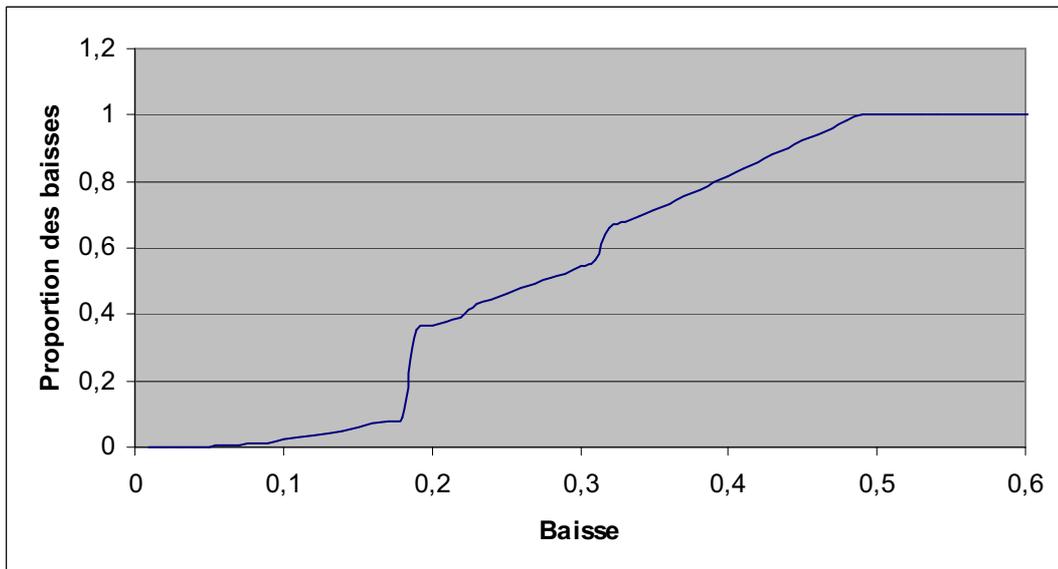
$$S/C_{\text{final}} = Z * S/C_{\text{contrat}} + (1 - Z) * \widehat{S/C}_{\text{catégorie}}$$

La méthode de crédibilité est celle de Jewell, et la crédibilité vaut $Z = \frac{\text{poids}}{\text{poids} + 6907}$. L'estimateur

$\widehat{S/C}_{\text{catégorie}}$ est un barycentre du S/C de la catégorie et de celui du portefeuille. Concrètement, il est presque égal au S/C de la catégorie.



Proportion des hausses inférieures ou égales à un niveau donné



Proportion des baisses inférieures ou égales à un niveau donné

3.3.3 Un exemple

Pour finir, donnons un exemple de calcul de ratio S/C annuel et un exemple de calcul de S/C final.

- Un calcul de S/C annuel :

On considère le cas d'un contrat M3T5 dont les caractéristiques pour l'année N-1 sont données ci contre. On détermine son S/C N-1.

Calcul des sinistres :

Sinistre 1 : $1200 * 1 * 1.01 = 1212$

Sinistre 2 : $50 * 0.95 * 1.01 = 48$

Sinistre 3 : $5000 * 1.2 * 1.05 + 5000 * 0.95 * 1.01 = 11097$

Le troisième sinistre est ramené à 3500 par écrêtement.

La cotisation est majorée de 10% et vaut donc 4400.

Le S/C du contrat pour N-1 est :

$$S/C = \frac{1212 + 48 + 3500}{4400} + \frac{320000}{1500000} + \frac{100000}{6000000} = 1.3.$$

- Un calcul de S/C historique :

On considère un contrat vérifiant les données ci contre.

Le S/C historique vaut $\frac{1000 * 0.5 + 1200 * 0.55 + 1200 * 0.6 + 1500 * 0.62}{1000 + 1200 + 1200 + 1500} = 0.57.$

On constate que le contrat s'est détérioré tous les ans, en moyenne de 7.4% par an. Son S/C devient donc $0.57 * (1 + 0.074 * 0.75) = 0.6.$

Le S/C final est donc $S/C = 0.7 * 0.6 + 0.3 * 0.57 = 0.59.$

DONNEES EXEMPLE 1

| Sinistres | Cout RC corp | Cout RC mat | Cout dommage |
|-----------|--------------|-------------|--------------|
| 1 | 0 | 0 | 1200 |
| 2 | 0 | 50 | 0 |
| 3 | 5000 | 5000 | 0 |

| | |
|----------------------------------|-----------|
| Prime année N-1 | 4000 |
| Majoration année N-1 | 10% |
| Seuil d'écrêtement | 3 500 |
| Prime totale de la catégorie | 1 500 000 |
| Cout mutualisé dans la catégorie | 320 000 |
| Prime totale du portefeuille | 6 000 000 |
| Cout mutualisé au niveau ptf | 100 000 |

DONNEES EXEMPLE 2

| Année | N-3 | N-2 | N-1 | N |
|-------|------|------|------|------|
| S/C | 0,5 | 0,55 | 0,6 | 0,62 |
| Poids | 1000 | 1200 | 1200 | 1500 |

| Crédibilité du contrat | S/C de la catégorie | Coefficient C pour la tendance |
|------------------------|---------------------|--------------------------------|
| 70% | 0,57 | 75% |

3.3.4 Choix du seuil d'équilibre

Un contrat est rentable s'il coûte moins qu'il ne rapporte. Cela semble indiquer que le ratio S/C d'un contrat doit être inférieur à 1. En pratique, ce ratio ne tient compte que des coûts de sinistres et pas d'autres sources de dépenses (frais généraux, commissions des agents généraux, frais de réassurance...). Le seuil d'équilibre d'un contrat se situe donc en dessous de 1. Un contrat est donc jugé comme rentable si son S/C est inférieur à ce seuil et déficitaire sinon.

Cette approche donne des résultats qui ne sont pas pleinement satisfaisants. En effet, les gros contrats ont un pouvoir de négociation supérieur aux petits et obtiennent souvent des rabais plus importants. Le S/C augmente nettement avec la taille. Par conséquent, le processus de majoration QP retient proportionnellement beaucoup plus de gros contrats que de petits. On donne ci contre les pourcentages obtenus. On constate que le taux de contrats retenus croît avec la taille (de façon encore plus nette pour les P3T5). Les contrats de taille 1 ne sont quasiment jamais retenus. Par contre ceux de taille 6 le sont très fréquemment (dans 40% des cas pour les M3T5 et près de 60 pour les P3T5). Cette situation n'est pas acceptable. On a donc décidé d'introduire un seuil d'équilibre variant avec la taille et le type de contrat. Concrètement cela revient à accepter que les petits contrats subventionnent les gros, ce qui est regrettable mais nécessaire si nous voulons conserver nos plus grosses affaires.

On donne aussi ci-contre la variance moyenne au cours du temps des contrats pour les différentes tailles. On constate que la stabilité du S/C augmente nettement avec la taille. Les gros contrats, bien que moins bons, offrent une plus grande sécurité ce qui justifie en partie que l'on soit moins exigeant avec eux.

Reste à déterminer les seuils de S/C pour chaque catégorie. On a graphiqué le S/C moyen en fonction de la taille pour les M3T5 et les P3T5. La progression est remarquablement linéaire (R^2 supérieur à 0.9 dans les deux cas). Les coefficients directeurs des droites de régression sont notés C_M et C_P respectivement. L'idée retenue est que le seuil d'équilibre doit augmenter de $\tau * C_M$ par taille de flotte pour les M3T5 et de $\tau * C_P$ pour les P3T5, pour un certain τ . Il faut donc choisir le coefficient τ . Prendre $\tau = 0$ revient à imposer le même seuil pour tous les contrats et on a vu que ce n'était pas souhaitable. Prendre $\tau = 1$ revient à ignorer totalement le fait que les gros contrats sont en moyenne moins bons ce qui est injuste pour les autres. On prend donc τ entre ces deux extrêmes. Il sera fixé à 75% par des arguments de pilotage.

Une fois que τ est choisi, les seuils sont déterminés à une constante additive près. Elle est choisie de façon à ce que le seuil moyen sur le portefeuille soit égal au niveau d'équilibre (au sens de la moyenne pondérée par les poids de 3.2). Ainsi, en moyenne le portefeuille a le niveau voulu.

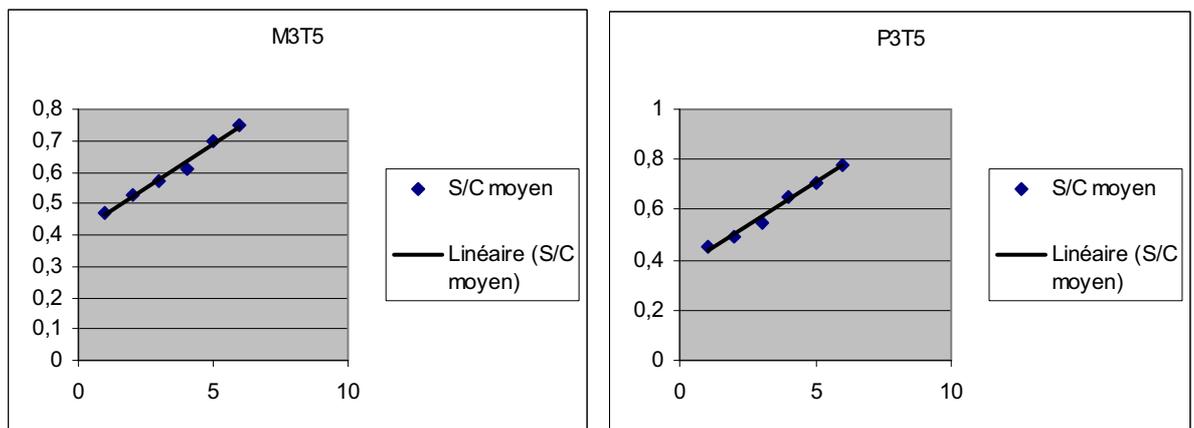
Enfin, on donne le pourcentage de contrats effectivement retenu. L'effet taille reste très net même s'il a été fortement diminué. Cela vient entre autre du fait que les petits contrats sont très protégés à la fois par la crédibilité et par l'écrêtement, contrairement aux gros.

| | M3T5 | | | | | | P3T5 | | | | | |
|-----------------|------|------|------|------|------|------|------|------|------|------|------|------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 |
| S/C moyen | 0,47 | 0,53 | 0,57 | 0,61 | 0,70 | 0,75 | 0,45 | 0,49 | 0,55 | 0,65 | 0,71 | 0,78 |
| variance du S/C | 0,34 | 0,15 | 0,13 | 0,07 | 0,04 | 0,10 | 0,27 | 0,14 | 0,11 | 0,09 | 0,08 | 0,07 |

Moyenne et variance du S/C en fonction de la taille de la flotte

| | M3T5 | | | | | | P3T5 | | | | | |
|---------------------|------|------|------|------|------|------|------|-----|------|------|------|------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 |
| Avec seuil uniforme | 1,1 | 7,2 | 15,0 | 22,2 | 39,1 | 40,7 | 0,6 | 4,8 | 12,8 | 27,5 | 41,3 | 57,1 |
| Avec seuil variable | 4,0 | 15,5 | 20,8 | 23,0 | 33,8 | 33,3 | 2,1 | 8,3 | 14,8 | 23,5 | 28,7 | 35,7 |

Proportion de contrats retenus en fonction de la taille de la flotte



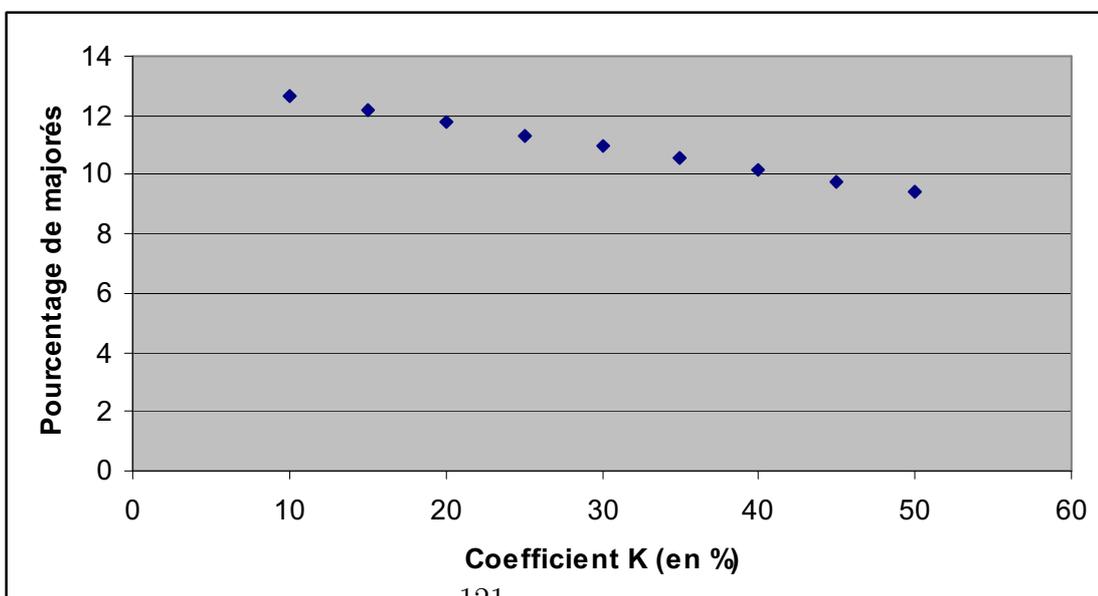
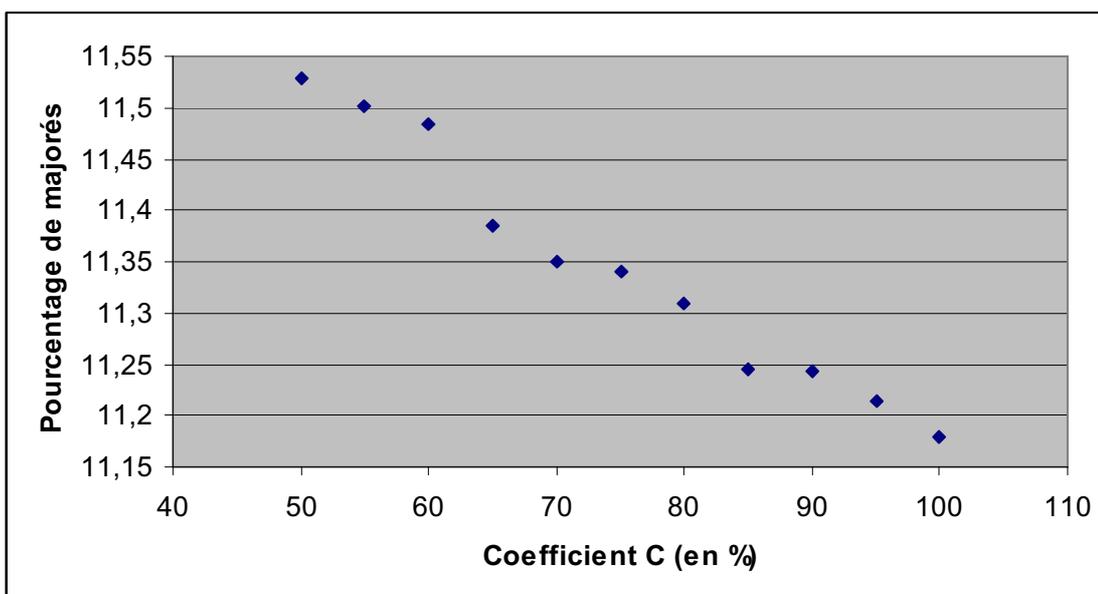
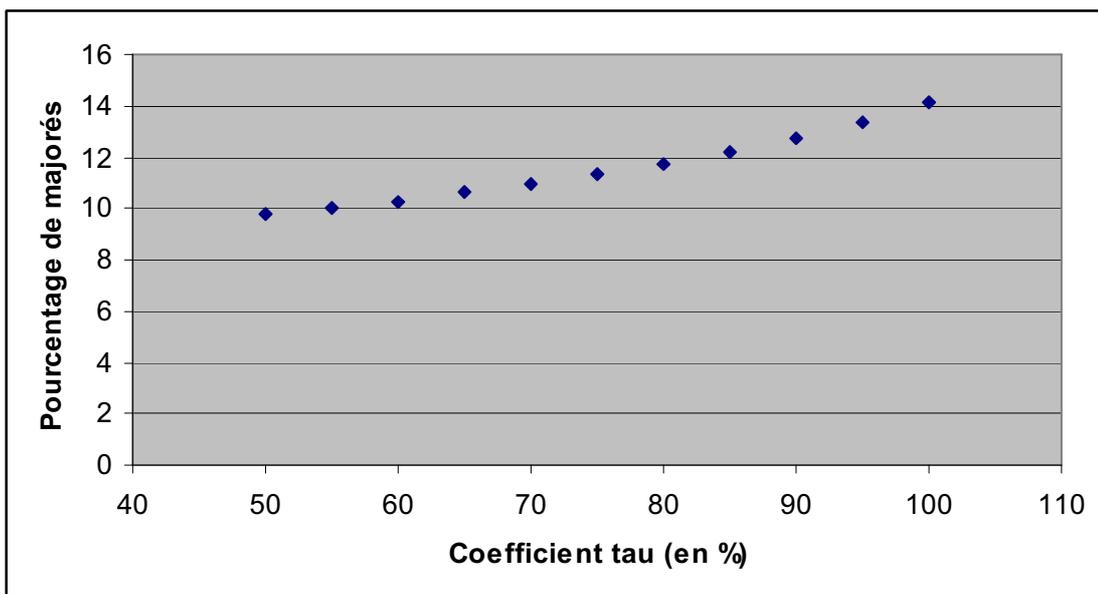
Régression des couts moyens sur les tailles de flottes

3.3.5 Pilotage de l'algorithme

Le but de la méthode décrite ci-dessus est de détecter les plus mauvais contrats et de les ramener à un niveau acceptable. Pour des raisons pratiques, il est préférable de ne pas repérer un trop grand nombre de contrats. Idéalement, on cherche à atteindre une proportion de 10 à 12 %. On veut de plus limiter le déséquilibre entre le taux de petits et de gros contrats repérés. Pour atteindre ces objectifs, on dispose de trois paramètres à fixer.

- La constante τ qui intervient dans le calcul du seuil d'équilibre (définie au 3.3.4). Plus τ est grand et plus le seuil d'équilibre varie en fonction de la taille. Quand on augmente τ , le taux de petits contrats retenus augmente et celui des gros diminue. Le taux de retenus sur l'ensemble du portefeuille augmente également (il y a beaucoup plus de petites flottes que de grosses).
- La constante C définie au 3.3.2 qui détermine à quel point la tendance (à la hausse ou à la baisse) du S/C doit être prise en compte par l'algorithme. Quand on augmente C , le taux de retenus décroît linéairement, mais de façon moins nette que pour les deux autres coefficients. L'effet est le même pour les petits et les gros contrats.
- La constante K définie au 3.2.2 qui mesure le surcroît de crédibilité accordé aux contrats dont le S/C est très stable. Augmenter K change peu de choses pour les gros contrats (qui sont de toutes façon déjà fortement crédibilisé). Par contre pour les petits contrats (dont le S/C est souvent instable), cela conduit en moyenne à une baisse du poids donc à une moindre crédibilité. Ainsi, le taux de retenus diminue (linéairement) avec K .

Au final, on a choisi de prendre $\tau = 0.75$, $C = 0.75$ et $K = 0.25$. Ce choix conduit à retenir 11% des contrats et à un effet taille moins net (mais encore très visible).



3.3.6 Niveau de majoration

Une fois que l'on a déterminé les contrats devant faire l'objet de majoration, reste à choisir le niveau de celle-ci. Pour chaque contrat, on a donné au 3.3.4 un seuil d'équilibre. Le contrat est jugé acceptable si son S/C est inférieur ou égal à ce seuil d'équilibre. Sinon, on le ramène au niveau acceptable en majorant de :

$$majoration = \frac{S/C_{\text{constaté}}}{S/C_{\text{équilibre}}}.$$

En somme, si toutes ces majorations avaient été appliquées il y a un an, aucun contrat n'aurait dépassé le seuil d'équilibre.

Dans la pratique, cela conduit à des résultats peu lisibles. Le pourcentage indiqué est donc arrondi suivant le barème donné ci contre. On a ainsi distingué sept niveaux de majoration. Les contrats correspondant seront appelés classe de majoration 1 à 7. La classe de majoration 0 regroupe les contrats qui sont sous le seuil d'équilibre (donc pour lesquels nous ne recommandons pas de modification).

On a choisi de ne pas majorer les contrats du premier niveau (dépassement de moins de 10%). Deux raisons à cela :

- le S/C que nous obtenons reste instable. Un contrat légèrement au dessus du seuil d'équilibre pourrait donc être en réalité rentable.
- pour des raisons pratiques, nous cherchons à retenir environ 10% des contrats. Ne pas majorer le premier niveau permet de se rapprocher de ce taux.

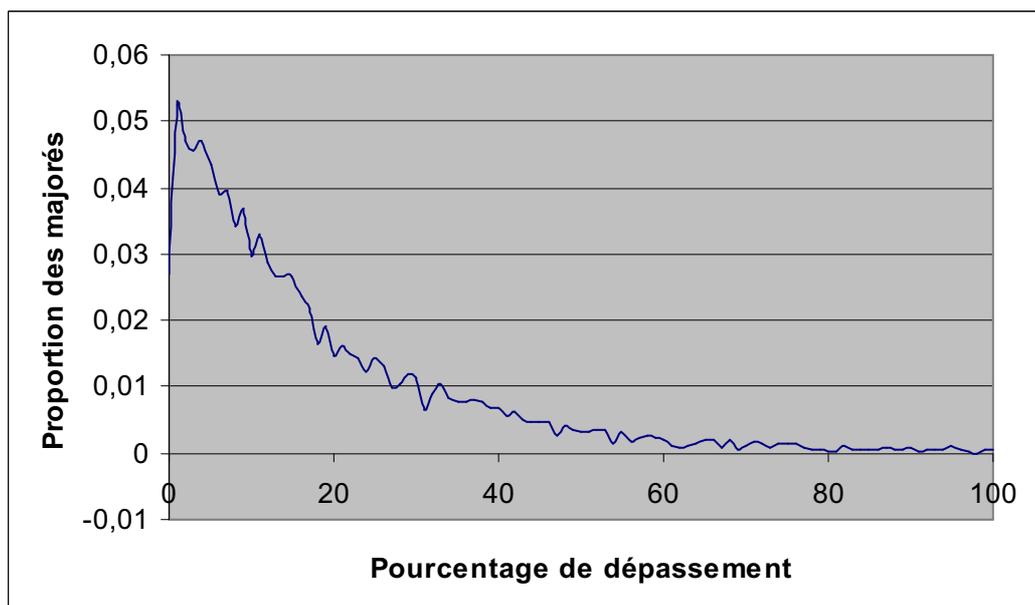
On donne enfin, parmi les contrats retenus, la proportion dans chacune de ces classes. On constate que dans la plupart des cas, le niveau indiqué est faible : dans plus d'un cas sur deux, il est de moins de 20%. On préconise peu de majorations élevées, ce qui est souhaitable. Les plus importantes (classe 6 et 7) ne représentent que 4% du total des majorés.

| | | | | | | | | |
|-----------------------|-----|-------|-------|-------|-------|-------|--------|-------|
| Dépassement | | 0 | 10% | 15% | 25% | 40% | 75% | |
| du seuil d'équilibre | <0% | à 10% | à 15% | à 25% | à 40% | à 75% | à 100% | >100% |
| Classe de majoration | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Majoration conseillée | 0% | 0% | 10% | 20% | 30% | 50% | 80% | 100% |

Majoration en fonction du dépassement du seuil d'équilibre

| | | | | | | |
|------------------------|--------|-------|--------|--------|-------|-------|
| Classe de majoration | 2 | 3 | 4 | 5 | 6 | 7 |
| Proportion des majorés | 24,94% | 30,17 | 24,65% | 16,38% | 2,67% | 1,19% |

Répartition des majorations entre les différents niveaux



Loi du dépassement du seuil d'équilibre

Synthèse de la troisième partie

On résume le processus de majoration QP d'un contrat.

- Calcul du S/C annuel du contrat :

Le S/C annuel d'un contrat est obtenu à partir des primes acquises pour cette année et des coûts de sinistres survenus cette année. Les primes sont actualisées, les sinistres sont actualisés, développés à l'ultime et écrêtés. Le quotient des deux donne le S/C annuel propre du contrat.

On obtient le S/C annuel en rajoutant la partie provenant de l'écrêtement (i.e. : en mutualisant les montants écrêtés lors des deux phases).

- Calcul du S/C propre du contrat :

Chaque observation est munie d'un poids qui dépend de sa cotisation et de son ancienneté. Le S/C propre est la moyenne pondérée des S/C annuels.

Pour les contrats ayant une tendance nette (à la hausse ou à la baisse), le S/C est modifié en conséquence.

- Calcul du S/C final :

Le S/C final est obtenu en crédibilisant le S/C propre. C'est un barycentre du S/C propre et de celui de la catégorie. La crédibilité d'un contrat augmente avec son poids.

- Calcul de la majoration à appliquer :

Le seuil d'équilibre à atteindre augmente avec la taille. Il est calculé de façon à ce que si tous les contrats étaient au seuil, le portefeuille ait un S/C à l'équilibre.

Les contrats en dessous du seuil ne sont pas majorés. Ceux au dessus font l'objet d'une majoration pour les ramener à l'équilibre.

Chapitre 4

Etude des résultats fournis par le processus QP

La méthode décrite au chapitre précédent a permis de calculer pour chaque contrat du portefeuille un S/C et si nécessaire un niveau de majoration. Reste à savoir si ces niveaux sont utilisables concrètement. Plusieurs questions peuvent se poser.

◇ Dans un premier temps, on va étudier la cohérence des résultats obtenus. Si le processus est bon, la sinistralité doit croître avec le S/C. De plus, les résultats doivent paraître logiques aux utilisateurs. On va d'abord vérifier que les contrats retenus ont bien une sinistralité plus importante que les autres. On va également comparer le S/C au S/C dossier dossier qui est utilisé par les souscripteurs.

◇ On va ensuite voir comment le processus se comporte dans l'espace. On cherche si certaines catégories de contrats sont plus souvent majorées que d'autres. Dans certains cas, c'est souhaitable (contrats en dégradation), dans d'autres cela vient du fait que certains types de contrats sont généralement moins bons (grosses flottes, TPM), enfin cela pourrait aussi indiquer un biais dans notre calcul.

◇ On va enfin voir comment le processus se comporte dans le temps. On sait que le S/C brut n'est guère stable. On peut craindre une forte variabilité dans le temps. C'est un problème classique avec ce genre d'algorithmes. On étudiera la stabilité dans le temps d'abord au niveau macro (évolution au cours du temps de la proportion de repérés ou de la majoration moyenne) puis au niveau micro (comparaison entre les S/C obtenus à deux dates différentes contrat par contrat).

4.1 Détection de la sinistralité

Le but du processus de majoration QP est de repérer les contrats les plus mauvais. On peut donc s'attendre à ce qu'ils aient une sinistralité plus élevée que les autres. Pour vérifier cela, on a graphiqué plusieurs quantités en fonction du rapport $\frac{S/C}{S/C_{\text{équilibre}}}$, donc en fonction du niveau de dépassement.

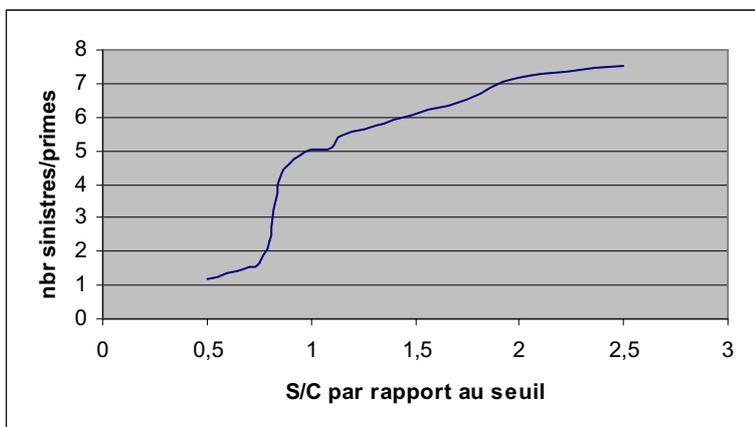
★ Un contrat pourrait être plus mauvais car il a eu plus de sinistres que la moyenne. Il faut bien sur rapporter le nombre de sinistres à la taille de la flotte. A nouveau, nous la matérialisons par la cotisation potentielle (ici, il aurait sûrement été préférable de raisonner par rapport au nombre de véhicules mais cette information n'est souvent pas disponible). On a donc calculé le ratio $\frac{\text{nbr sinistres}}{\text{primes}}$ en fonction du niveau de dépassement.

★ Un contrat pourrait être plus mauvais car ses sinistres ordinaires sont plus coûteux que la moyenne. Pour le vérifier, on a tracé le coût moyen des sinistres après écrêtement (pour évacuer l'effet sinistre exceptionnel) en fonction du niveau de dépassement.

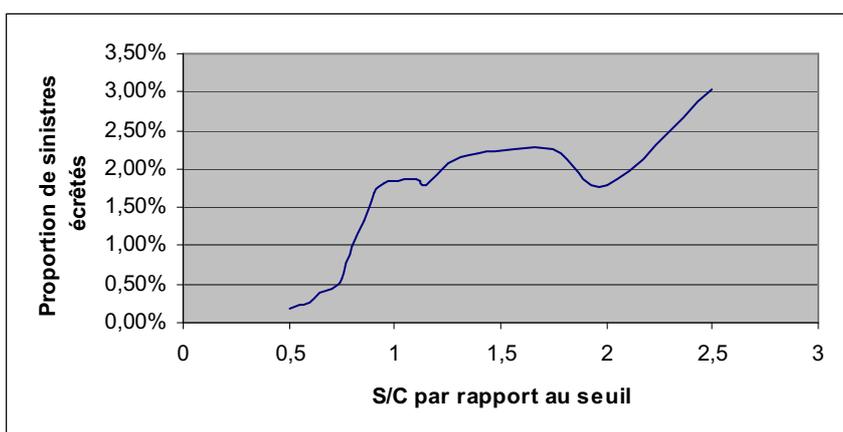
★ Enfin, un contrat pourrait être plus mauvais du fait de la présence de sinistres extrêmes. On a déterminé la proportion de sinistres étant écrêtés en fonction du niveau de dépassement.

Les graphes ci-dessus confirment ce que l'on pensait. La fréquence de la sinistralité et son intensité (aussi bien pour les sinistres ordinaires qu'exceptionnels) augmente avec le niveau de dépassement. Cela tend à confirmer la pertinence de l'algorithme. On remarque de plus un basculement assez net autour de 80% du seuil d'équilibre. Au delà de ce niveau, la sinistralité augmente nettement.

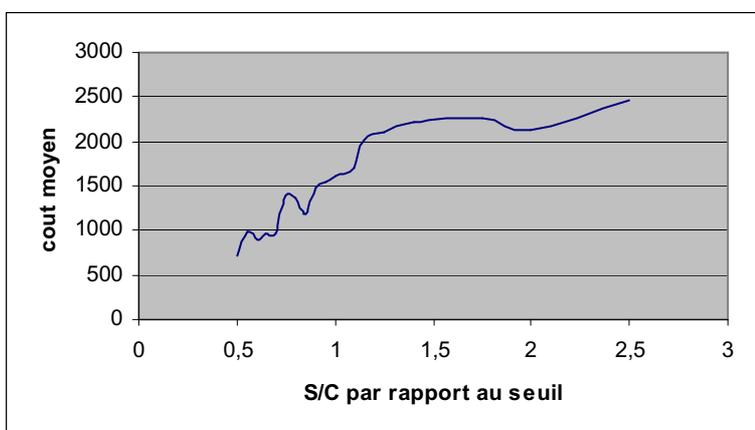
Ces résultats indiquent que l'algorithme semble cohérent et que notre ratio S/C capte bien une information de sinistralité. Il faut cela dit bien garder à l'esprit que ce ne sont que des valeurs moyennes et que des points aberrants peuvent exister.



Ration nombre de sinistres sur primes en fonction du niveau de dépassement



Proportion de sinistres écrêtés en fonction du niveau de dépassement



Cout moyen des sinistres en fonction du niveau de dépassement

4.2 Etudes statistiques des contrats retenus

Dans cette partie, on cherche à repérer d'éventuelles caractéristiques communes aux contrats retenus par le processus de majoration QP. On va déterminer suivant différents critères le taux de contrats repérés. On calculera ensuite les corrélations entre les différentes modalités et le fait d'être majoré. Enfin, on affinera l'étude par une analyse des correspondances multiples.

Cette étude doit permettre de repérer les types de police les plus susceptibles d'être repérées. Cela peut venir de données concrètes, c'est-à-dire que certaines classes contrats sont moins rentables que d'autres. On a par exemple vu au chapitre précédent que c'était le cas des grosses flottes. On avait limité cet effet en ajustant le seuil d'équilibre en fonction de la taille. Il pourrait être souhaitable de faire de même pour d'autres genres de polices. De plus, cela pourrait nous indiquer quelles classes de contrats surveiller plus attentivement. Enfin, cela nous permettra de rechercher d'éventuelles failles dans l'algorithme : une catégorie qui serait plus (ou moins) retenue que la moyenne sans que cela paraisse pertinent.

Une des difficultés vient du fait que la plupart des caractéristiques utilisées dans ce type d'étude sont liées aux véhicules ou au conducteur et non à la flotte (puissance moteur, âge du véhicule, bonus malus du conducteur..). Ces variables ne sont pas utilisables en automobile d'entreprise car les flottes comportent des véhicules très divers (et dans le cas des flottes ouvertes, on n'a même pas ces données à disposition). On dispose néanmoins des différentes segmentations du portefeuille (taille de la flotte, usage, poids) et on a ajouté d'autres critères.

4.2.1 Variables de l'étude

On donne la liste des variables. On en a retenu 12 pour 36 modalités. Les modalités ont été choisies de façon à avoir des classes de tailles proches, quitte à en regrouper certaines. On a fait en sorte qu'elles représentent toutes au moins 5% du portefeuille.

- Mode gestion : Distingue les flottes ouvertes (X) des flottes fermées (F). Comme les flottes ouvertes sont moins bien connues, on peut craindre que les tarifs soient moins précis et qu'il y ait plus de contrats repérés.
- Région : AXA a découpé le territoire français en 5 régions notées 64 à 68. Il peut y avoir des différences dans la sinistralité entre les zones géographiques. Par exemple, en région parisienne, le nombre de sinistres est en général plus élevé.
- Ancienneté : Nombre d'années depuis le début du contrat. En général les contrats plus anciens sont meilleurs.
- Type : Sépare les flottes entre plus et moins de 3.5 tonnes. On sait que les P3T5 ont souvent des sinistres plus coûteux que les M3T5.
- Usage : Sépare les flottes entre transport de marchandises (TPM) de voyageurs (TPV) et autres (AUT). Les TPM ont tendance à avoir des sinistres plus coûteux et un S/C moins bon, les TPV ont plus de sinistres extrêmes.
- Taille : La subdivision en classe de cotisation définie au 2.1.1. Comme les trois dernières classes contiennent peu de contrats, on les a regroupées en une seule (notée 4+). On a vu que le S/C se dégrade très rapidement avec la taille de la flotte.
- Amélioration : Indique si le S/C du contrat s'est amélioré ou s'est dégradé ces dernières années. On espère retenir moins de contrat en amélioration et plus en dégradation. Par défaut, les nouveaux contrats sont affectés à « sans tendance ».
- Stabilité : Indique si le S/C d'un contrat est ou non stable sur les dernières années. Par défaut, les nouveaux contrats sont affectés à « moyennement stable ».
- Développement : Indique si le contrat est en croissance, i.e. : si sa cotisation a augmenté ces dernières années.
- Toujours sinistré : Indique si le contrat a ou non été sinistré tous les ans depuis N-3 (sauf les années où il n'existait pas).
- Rabais : Indique le rabais accordé au contrat par rapport au tarif de référence. On peut craindre que les contrats bénéficiant de rabais soient moins bons que les autres.
- Majoration : Indique si le contrat a déjà subi une majoration (depuis l'année N-3).

| Variable | Modalité | Description | % des contrats |
|---------------|--------------------|---|----------------|
| Mode gestion | X | Flottes ouvertes | 20% |
| | F | Flottes fermées | 81% |
| Région | 64 | Ile de France | 20% |
| | 65 | Nord est | 21% |
| | 66 | Ouest | 23% |
| | 67 | Sud est | 20% |
| | 68 | Sud ouest | 16% |
| Ancienneté | 1-2 ans | | 23% |
| | 3-4 ans | | 22% |
| | 5-9 ans | | 24% |
| | 10 ans et + | | 31% |
| Type | M3T5 | Moins de 3,5 tonnes | 71% |
| | P3T5 | Plus de 3,5 tonnes | 29% |
| Usage | AUT | | 82% |
| | TPM | Transport de marchandises | 13% |
| | TPV | Transport de voyageurs | 5% |
| Taille | 1 | <5000 euros | 45% |
| | 2 | 5000 à 10000 euros | 24% |
| | 3 | 10000 à 25000 euros | 19% |
| | 4 | >25000 euros | 12% |
| Amélioration | Amélioration | S/C en baisse tous les ans | 17% |
| | Dégradation | S/C en hausse tous les ans | 31% |
| | Sans tendance | | 52% |
| Stabilité | Stable | Variance du s2 inférieure à 0,75*variance moyenne | 28% |
| | Moyenne | | 35% |
| | Instable | Variance du s2 supérieure à variance moyenne | 37% |
| Développement | En croissance | Cotisation en hausse tous les ans | 57% |
| | Pas en croissance | | 43% |
| Tjrs sinistré | Oui | Au moins un sinistre chaque année | 60% |
| | Non | | 40% |
| Rabais | Gros rabais | Rabais de 50% ou plus | 21% |
| | Rabais | Rabais de 10 à 50% | 30% |
| | Pas de rabais | | 42% |
| | Au dessus du tarif | Au dessus du tarif facial | 7% |
| Déjà majoré | Déjà majoré | A déjà fait l'objet d'une majoration depuis N-3 | 17% |
| | Non majoré | | 83% |

4.2.2 Liens entre les variables et les majorations

Pour chaque modalité, on a calculé la proportion de contrats repérés comme non rentable. On a donné également la corrélation entre les variables indicatrices $1_{\text{modalité}}$ et $1_{\text{contrat repéré}}$. Enfin, pour chaque variable, on donne le V de Cramer qui mesure son degré de dépendance avec la variable $1_{\text{contrat repéré}}$ (compris entre -1 et 1).

Ci contre, les modalités ayant les plus forts taux de majorés sont données en rouge. Celles ayant le moins sont en bleu. Enfin, les variables les plus liées au fait d'être majoré sont indiquées en jaune. On peut faire plusieurs constatations :

- Le mode de gestion n'influe presque pas sur le taux de retenus. En fait, c'est la seule variable que l'on puisse considérer comme indépendante du fait d'être majoré. C'est assez inattendu. Les flottes ouvertes sont souvent considérées comme plus dangereuses pour l'assureur du fait que l'on ne contrôle pas sa composition. Dans la suite de l'étude, on ne tiendra donc plus compte de cette variable.
- Les régions n'ont pas non plus une grande influence. Le tarif (qui tient compte de la zone géographique du contrat) a permis d'égaliser les risques.
- Les contrats plus récents sont très peu repérés. C'est un défaut du modèle. En effet, les contrats les plus récents ont un poids faible (poids nul pour les années où ils n'existent pas). Ils sont donc peu crédibilisés, ce qui les protège. Plus le contrat devient ancien, plus ce biais diminue. Enfin, pour les contrats de plus de 4 ans, le facteur poids cesse de jouer et le taux de repéré diminue avec le temps comme on pouvait l'espérer.
- Type : L'incidence du type est très faible. Cela vient sans doute du fait que les contrats M3T5 et P3T5 sont traités de façon séparée tout au long de l'algorithme et que les P3T5 ont un seuil d'équilibre plus élevé. C'est aussi un biais du système.
- Usage : Contrairement au type, l'usage a un effet notable. Les TPM et dans une moindre mesure les TPV sont très fortement repérés. Cela correspond aux observations du portefeuille (les TPM ont souvent un S/C très élevé).
- Amélioration : Comme on le souhaitait, les contrats en amélioration sont moins touchés que la moyenne et ceux en dégradation plus. Cela est sans doute dû à la prise en compte de la tendance dans le calcul du S/C (comme décrit au 3.3.2).
- Développement : Les contrats en développement sont plus touchés que la moyenne, ce qui est regrettable.
- Stabilité : La stabilité du S/C est très liée au taux de contrats repérés. Ce lien est néanmoins assez tautologique : les contrats dont le S/C est très variable sont souvent ceux qui ont eu une année très mauvaise. On se doute qu'ils ont tendance à être plus repérés.
- Taille : Comme on l'a vu au paravent, les gros contrats sont nettement moins bons et donc plus souvent repérés que les petits, et ce même si leur seuil d'équilibre est plus haut.
- Toujours sinistré : Les contrats qui ont au moins une année sans sinistres sont beaucoup moins repérés. Cependant, ce sont surtout des contrats de petite taille. Cette variable double donc la précédente. Sur l'ACM, on remarquera d'ailleurs que les trajectoires des variables sont parallèles.
- Rabais : Contrairement à ce que l'on pouvait penser, plus un contrat a un rabais important, plus il est sûr. Cela correspond au « flair » des souscripteurs qui accordent plus de rabais aux contrats qu'ils devinent meilleurs.
- Déjà majoré : Les contrats qui ont déjà été majorés les années précédentes ont plus de chances de l'être à nouveau. En effet, les majorations suggérées ne sont pas toujours appliquées. Dans ce cas, le contrat risque de rester sous le seuil de rentabilité.

| Variable | Modalité | % de contrats repérés | Corrélation avec le fait d'être repéré |
|---------------|--------------------|-----------------------|--|
| Mode gestion | X | 10,6% | 0,9% |
| | F | 11,2% | |
| Région | 64 | 13,0% | 4,9% |
| | 65 | 10,5% | |
| | 66 | 9,6% | |
| | 67 | 12,5% | |
| | 68 | 9,9% | |
| Ancienneté | 1-2 ans | 6,8% | 12,1% |
| | 3-4 ans | 8,2% | |
| | 5-9 ans | 15,6% | |
| | 10 ans et + | 13,0% | |
| Type | M3T5 | 10,5% | 3,5% |
| | P3T5 | 12,7% | |
| Usage | AUT | 9,7% | 11,3% |
| | TPM | 19,2% | |
| | TPV | 14,7% | |
| Taille | 1 | 3,7% | 28,9% |
| | 2 | 13,8% | |
| | 3 | 18,5% | |
| | 4 | 25,0% | |
| Amélioration | Amélioration | 7,1% | 9,4% |
| | Dégradation | 14,8% | |
| | Sans tendance | 10,2% | |
| Stabilité | Stable | 1,5% | 37,8% |
| | Moyenne | 8,0% | |
| | Instable | 28,1% | |
| Développement | En croissance | 13,4% | 10,9% |
| | Pas en croissance | 7,6% | |
| Tjrs sinistré | Oui | 16,8% | 24,1% |
| | Non | 2,6% | |
| Rabais | Gros rabais | 10,2% | 3,5% |
| | Rabais | 11,0% | |
| | Pas de rabais | 11,2% | |
| | Au dessus du tarif | 14,5% | |
| Déjà majoré | Déjà majoré | 26,0% | 23,5% |
| | Non majoré | 8,0% | |

Lien entre les modalités et le fait d'être majoré.

En bleu les modalités les moins majorées, en rouge les plus touchées et en jaune les variables les plus corrélées au fait d'être détecté.

4.2.3 Analyse des correspondances multiples

Pour affiner l'étude précédente, on va effectuer une ACM sur les variables décrites plus haut (sauf le mode de gestion dont on a vu qu'il n'intervenait pas). On va faire intervenir les majorations en tant que variables supplémentaires, d'abord en projetant la variable 1_{le contrat est repéré} puis en projetant les différents niveaux de majoration.

4.2.3.1 Rappel sur l'ACM

◇ Généralités :

L'ACM est une méthode permettant de visualiser les liens entre les modalités de plusieurs variables qualitatives. On en rappelle le principe et les principales propriétés.

On se donne m variables, la variable i ayant m_i modalités. L'espace engendré par les modalités a pour dimension $N = \sum_i (m_i - 1) = \sum m_i - m$. On note n le nombre d'individu.

On considère le tableau de Burt : $B = (b_{ij})$ où b_{ij} est le nombre d'observations ayant les modalités i et j . On veut effectuer une ACP de ce tableau. On cherche donc des variables combinaisons linéaires des modalités : $A_n = \sum_k \lambda_k 1_{\text{modalité } k}$. On choisit ces variables de façon à ce que le nuage des observations projeté sur l'espace $\text{vect}(A_1..A_k)$ ait une inertie maximale parmi les sous espaces de dimension k . La solution consiste à prendre comme variables (A_i) une base de vecteurs propres de VM où V est la matrice de variance covariance du tableau B et M la matrice des poids (le poids d'une modalité est la proportion d'individus ayant cette modalité).

Une fois que l'on a déterminé les A_i , on peut projeter le nuage sur $\text{vect}(A_1..A_k)$ avec k bien choisi. On obtient ainsi une représentation du nuage des observations dans un espace de dimension plus faible. Cela rend l'étude plus lisible et permet de mieux visualiser les liens entre les différentes modalités.

◇ Choix du nombre d'axes :

Reste à choisir le nombre d'axe. Un critère classique consiste à faire en sorte que le nuage projeté conserve un pourcentage suffisamment élevé de l'inertie du nuage initial. Dans le cas de l'ACM, cette méthode ne peut pas fonctionner. En effet, on a un très grand nombre d'axes, chacun contribuant peu à l'inertie du nuage. Il faudrait donc conserver beaucoup d'axes, ce qui rendrait l'analyse illisible.

Une autre approche a été proposée par JP Benzecri. L'idée est de ne conserver que les axes correspondants à des valeurs propres μ supérieures $\frac{1}{m}$. On recalcule ensuite l'inertie à partir de ces seuls axes. L'inertie associée à l'axe k devient $(\mu - \frac{1}{m})^2$.

On choisit ensuite un nombre d'axe de façon à conserver un pourcentage assez élevé de l'inertie au sens de Benzecri. En pratique, il y a généralement un seuil au delà duquel cette inertie baisse nettement. Une solution est de prendre tous les axes jusqu'à ce seuil.

◇ Interprétation des axes :

Reste ensuite à interpréter les axes obtenus. On regarde pour cela quelles sont les variables apportant le plus d'inertie à l'axe. L'inertie que le point i du nuage apporte à l'axe j , se note $ctr(i, j)$. Plus cette quantité est grande et plus le point joue un rôle important pour interpréter l'axe. Néanmoins, plus un point a un poids important (plus sa fréquence est élevée), plus il a d'inertie et donc plus ses contributions seront grandes. Pour pouvoir comparer l'apport de deux variables, il faut donc tenir compte de leur poids. On dira par la suite qu'une variable contribue à un axe si sa contribution excède son poids, qu'elle contribue assez fortement si sa contribution excède deux fois son poids et qu'elle contribue très fortement si sa contribution excède trois fois son poids.

◇ Qualité de la représentation d'un point

On peut aussi chercher à savoir si un point du nuage est bien représenté par sa projection sur un axe. Le critère consiste à voir quelle proportion d'inertie est conservée après la projection. Cela revient à calculer le cosinus de l'angle entre le point et l'espace de projection.

Pour la modalité i et l'axe j , cela donne

$$\cos^2(i, j) = \frac{\text{inertie du projeté de } i \text{ sur l'axe } j}{\text{inertie totale de } i}.$$

En pratique, un cos carré supérieur à 0.1 indique que le point est bien représenté par sa projection sur l'axe.

On peut aussi voir si un point est bien représenté par son projeté sur l'espace engendré par les k premiers axes. Il suffit d'additionner les cos carrés correspondants à chacun des axes (cela revient encore à faire le quotient $qual(i) = \frac{\text{inertie du projeté de } i \text{ sur l'espace de projection}}{\text{inertie totale de } i}$).

◇ Remarques diverses :

Plus une modalité a une fréquence (donc un poids) faible, plus ses contributions sont fortes. Les modalités à très faibles effectifs peuvent donc impacter fortement le résultat de l'ACM.

De même, la contribution totale d'une variable augmente avec le nombre de ses modalités.

Au final, on fera donc en sorte de ne pas avoir de modalités représentant moins de 5% des données et pas de variables de plus de six modalités.

4.2.3.2 Application aux données

On a effectué l'ACM à partir des variables données ci-dessus. Comme on a 12 variables, un axe sera jugé significatif si son inertie dépasse $\frac{1}{12}$. En l'occurrence, cela donne 10 axes. On donne aux pages suivantes les inerties au sens classique et au sens de Benzecri de chaque axe. Au final, il y a une baisse nette d'inertie à partir du deuxième axe et l'inertie cumulée vaut 84%. On retient donc deux axes et la représentation est jugée de bonne qualité.

Remarque : le troisième axe a une inertie assez notable, on aurait donc pu en retenir trois. Néanmoins cela aurait compliqué la représentation graphique.

Le premier axe oppose des gros contrats et des contrats ayant des gros sinistres (taille 3 et 4+, TPM P3T5) à des petits contrats (taille1, une année sans aucun sinistre).

Le deuxième axe oppose des gros contrats, des contrats à gros sinistres et des contrats stables (10 ans et +, pas en croissance) à des contrats en évolution (S/C en hausse, en baisse et instable).

On donne ci contre le résultat de l'ACM. On observe cinq grands groupes de points.

◇ Un premier groupe en haut à droite comprend les taille3 et 4+, les TPM, les TPV et les P3T5 ainsi que les contrats ayant au moins un sinistre chaque année et ceux dont le S/C est moyennement stable. Ce groupe représente les gros contrats et ceux dont les sinistres sont plus élevés que la moyenne.

◇ Un deuxième groupe en bas à droite comprend les contrats dont le S/C n'est pas stable, en hausse, en baisse, dont les primes augmentent et qui ont déjà fait l'objet d'une majoration. Il contient aussi les contrats d'âge intermédiaire et ceux dont la prime est supérieure au tarif. Les contrats de ce groupe sont en évolution, sous une forme ou une autre.

◇ Un troisième groupe en bas à gauche regroupe les contrats de petite taille, les M3T5 et les AUT ainsi que ceux ayant eu une année sans sinistre. On peut voir ce groupe comme l'opposé du premier. Il correspond aux petits contrats et à ceux ayant peu de sinistres coûteux.

◇ Un quatrième groupe en haut à gauche regroupe les contrats anciens, ceux au S/C stable et n'ayant pas de tendance et ceux bénéficiant d'un gros rabais. C'est l'opposé du groupe 3. Il s'agit des contrats les plus stables, qui n'évoluent pas.

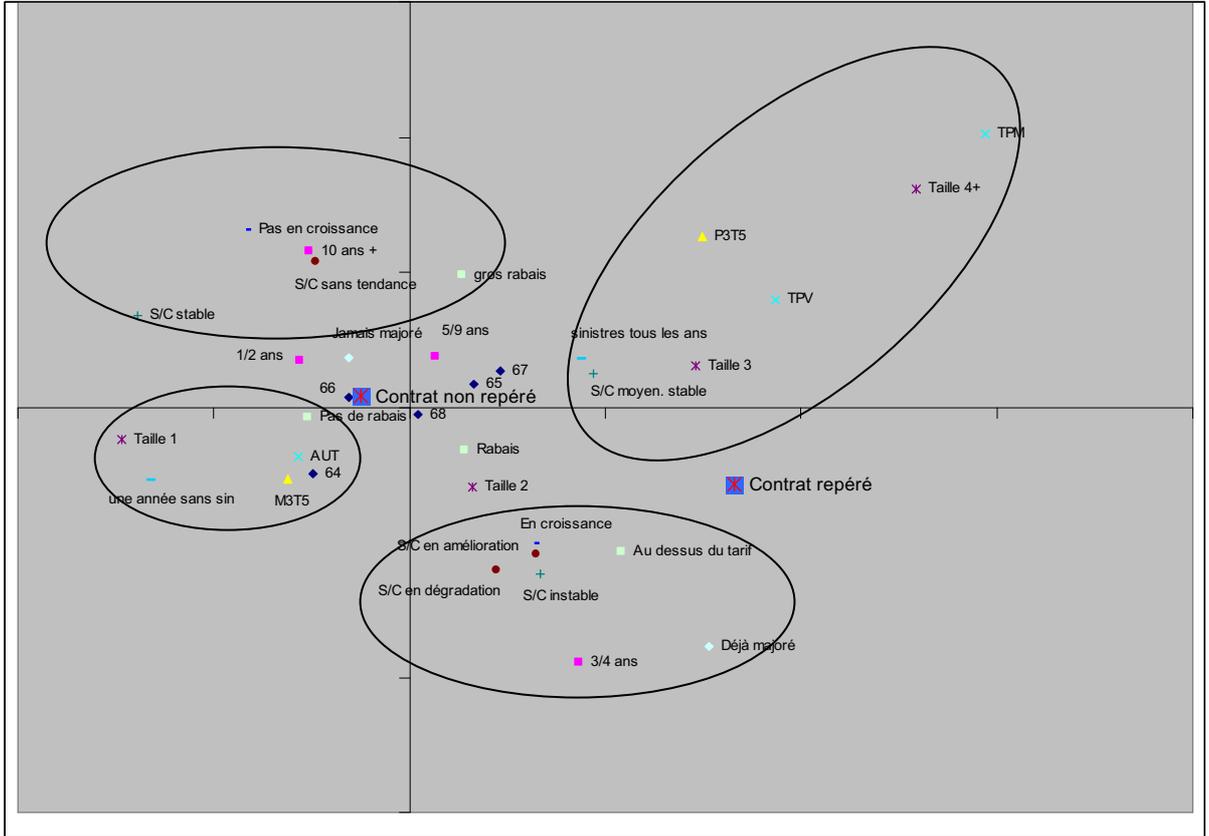
◇ Un dernier groupe au centre regroupe des variables qui sont mal représentées dans le plan factoriel (presque tous les points ayant une qualité inférieure à 10% s'y trouvent). Il s'agit surtout des régions et des différents niveaux de rabais ainsi que de certaines anciennetés de contrats.

On a également projeté la variable 1^{er} contrat est repéré. Les contrats repérés sont situés nettement sur la droite du graphique (\cos^2 de 0.1) donc du côté des gros contrats et des contrats en évolution. Ils se projettent moins bien sur le deuxième axe (\cos^2 de 0.01) du côté inférieur ce qui indiquerait que l'instabilité joue un rôle encore plus grand que la nature et la taille du contrat. A l'inverse, les contrats non repérés sont situés près des contrats petits, stables et ayant peu de gros sinistres.

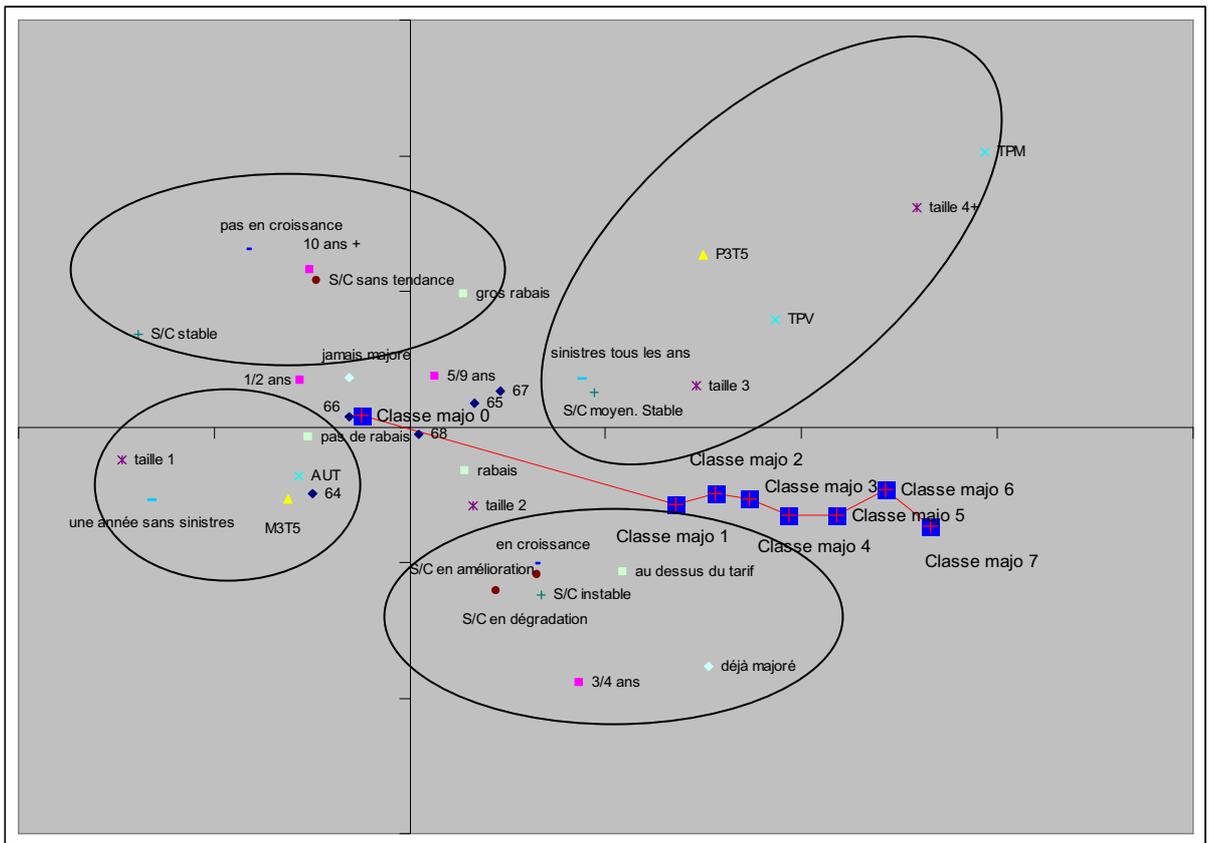
En somme les facteurs qui augmentent le plus le risque d'être majoré sont la taille, la probabilité d'avoir des gros sinistres (P3T5, TPM, TPV) et le fait d'être instable (instabilité du S/C ou de la prime). L'ACM confirme donc les observations obtenues au paragraphe précédent.

Enfin on donne à la page suivante l'ACM avec projection des différents niveaux de majorations (non retenu, majoré classe 1 à 7, plus la classe est grande plus le contrat doit être majoré). On constate à nouveau le rôle essentiel du premier axe : plus le niveau de majoration est grand et plus le contrat se situe sur la droite de l'axe. A l'inverse le deuxième axe ne semble pas jouer un grand rôle. Cette impression est confirmée par les qualités des points (de l'ordre de 10 % pour l'axe 1 et 2 à 0.1% pour l'axe 2).

A nouveau on remarque que le niveau de majoration augmente avec la taille du contrat, son risque de sinistres graves et son instabilité.



ACM avec le statut repéré/non repéré en variable supplémentaire.



ACM avec le niveau de majoration en variable supplémentaire

| Axe | Inertie | d_k | Pourcentage d'inertie | Inertie Benzecri cumulée |
|-----|---------|------------|-----------------------|--------------------------|
| 1 | 0,20294 | 0,01255092 | 54,6120% | 54,6120% |
| 2 | 0,16796 | 0,00593684 | 25,8326% | 80,4445% |
| 3 | 0,14179 | 0,00258887 | 11,2648% | 91,7093% |
| 4 | 0,12511 | 0,0011697 | 5,0896% | 96,7989% |
| 5 | 0,1146 | 0,00056126 | 2,4422% | 99,2411% |
| 6 | 0,10138 | 0,00010964 | 0,4771% | 99,7182% |
| 7 | 0,09827 | 5,4183E-05 | 0,2358% | 99,9539% |
| 8 | 0,09404 | 9,8026E-06 | 0,0427% | 99,9966% |
| 9 | 0,09179 | 7,76E-07 | 0,0034% | 100,0000% |
| 10 | 0,091 | 8,2645E-09 | 0,0000% | 100,0000% |
| 11 | 0,08988 | | | |
| 12 | 0,08882 | | | |
| 13 | 0,08557 | | | |
| 14 | 0,08251 | | | |
| 15 | 0,08114 | | | |
| 16 | 0,07516 | | | |
| 17 | 0,07212 | | | |
| 18 | 0,06593 | | | |
| 19 | 0,05692 | | | |
| 20 | 0,05515 | | | |
| 21 | 0,0437 | | | |
| 22 | 0,03898 | | | |
| 23 | 0,02615 | | | |

Inertie des différents axes.

| | Contribution positive | | Contribution négative | |
|-------|-----------------------|-------------------|-------------------------|--------------|
| | forte | assez forte | forte | assez forte |
| Axe 1 | P3T5 | taille 3 | Une année sans sinistre | taille 1 |
| | TPM | | | S/C stable |
| | TPV | | | |
| | taille 4+ | | | |
| | déjà majoré | | | |
| Axe 2 | TPM | P3T5 | S/C en hausse | 3-4 ans |
| | taille 4+ | 10 ans et + | S/C en baisse | déjà majorés |
| | | pas en croissance | S/C instable | |

Explication des axes par les variables.

| Variable | Modalité | Qualité | CTR axe 1 | CTR axe 2 | poids | Cos2 axe1 | Cos2 axe2 |
|-------------|-----------------|---------|-----------|-----------|--------|-----------|-----------|
| Région | 64 | 0,0303 | 0,55% | 0,65% | 0,0166 | 0,0153 | 0,0151 |
| | 65 | 0,0095 | 0,26% | 0,09% | 0,0181 | 0,0075 | 0,002 |
| | 66 | 0,0075 | 0,25% | 0,02% | 0,0188 | 0,0071 | 0,0004 |
| | 67 | 0,0174 | 0,47% | 0,19% | 0,0163 | 0,013 | 0,0044 |
| | 68 | 0,0002 | 0,00% | 0,01% | 0,0135 | 0,0001 | 0,0001 |
| Ancienneté | 1-2 | 0,0337 | 0,83% | 0,39% | 0,0198 | 0,0243 | 0,0094 |
| | 10+ | 0,1115 | 0,64% | 3,95% | 0,0181 | 0,0184 | 0,0931 |
| | 3-4 | 0,3355 | 1,97% | 11,46% | 0,0198 | 0,0578 | 0,2777 |
| | 5-9 | 0,0172 | 0,06% | 0,58% | 0,0256 | 0,0019 | 0,0154 |
| Type | M3T5 | 0,4006 | 3,08% | 2,66% | 0,0587 | 0,2335 | 0,1671 |
| | P3T5 | 0,4006 | 7,38% | 6,38% | 0,0245 | 0,2335 | 0,1671 |
| Usage | AUT | 0,5242 | 2,99% | 1,45% | 0,0684 | 0,3739 | 0,1503 |
| | TPM | 0,4596 | 12,20% | 7,00% | 0,0105 | 0,3117 | 0,1479 |
| | TPV | 0,0567 | 2,04% | 0,45% | 0,0043 | 0,048 | 0,0087 |
| Taille | Taille 1 | 0,4565 | 10,92% | 0,36% | 0,0375 | 0,4444 | 0,0121 |
| | Taille 2 | 0,0349 | 0,28% | 1,11% | 0,0198 | 0,0082 | 0,0268 |
| | Taille 3 | 0,1345 | 4,65% | 0,25% | 0,0161 | 0,1288 | 0,0057 |
| | Taille 4 | 0,3084 | 8,77% | 4,14% | 0,0097 | 0,2218 | 0,0866 |
| Tendance | AMELIORATION | 0,0818 | 0,80% | 2,71% | 0,0142 | 0,0214 | 0,0604 |
| | DEGRADATION | 0,1868 | 0,69% | 6,14% | 0,0259 | 0,0222 | 0,1646 |
| | SANS AMEL. | 0,3758 | 1,33% | 8,18% | 0,0432 | 0,0616 | 0,3142 |
| Stabilité | INSTABLE | 0,1891 | 1,39% | 5,72% | 0,0231 | 0,043 | 0,1461 |
| | MOY.STABLE | 0,1282 | 3,48% | 0,30% | 0,0292 | 0,1197 | 0,0086 |
| | STABLE | 0,3535 | 8,00% | 2,34% | 0,0310 | 0,2846 | 0,0689 |
| Croissance | CROISSANCE | 0,4653 | 2,60% | 7,80% | 0,0471 | 0,1337 | 0,3315 |
| | PAS DE CROIS. | 0,4653 | 3,39% | 10,14% | 0,0362 | 0,1337 | 0,3315 |
| Sin tjrs | AUTOMATIQUE | 0,338 | 5,18% | 1,04% | 0,0500 | 0,29 | 0,048 |
| | PAS TJRS | 0,338 | 7,81% | 1,56% | 0,0332 | 0,29 | 0,048 |
| Déjà majoré | MAJORE | 0,2793 | 4,43% | 7,19% | 0,0141 | 0,1193 | 0,1601 |
| | NON MAJORE | 0,2793 | 0,91% | 1,47% | 0,0691 | 0,1193 | 0,1601 |
| Rabais | RABAIS | 0,0191 | 0,26% | 0,42% | 0,0247 | 0,0082 | 0,011 |
| | GROS RABAIS | 0,0685 | 0,17% | 2,72% | 0,0175 | 0,0049 | 0,0636 |
| | PAS DE RABAIS | 0,0499 | 1,27% | 0,03% | 0,0350 | 0,0488 | 0,0011 |
| | SUR TARIF | 0,0453 | 0,95% | 1,12% | 0,0060 | 0,0229 | 0,0223 |
| Retenu | Non retenu | 0,1155 | | | | 0,1032 | 0,0123 |
| | Retenu | 0,1155 | | | | 0,1032 | 0,0123 |
| Majoration | Retenu classe 1 | 0,1022 | | | | 0,0975 | 0,0047 |
| | Retenu classe 2 | 0,1251 | | | | 0,111 | 0,0141 |
| | Retenu classe 3 | 0,0932 | | | | 0,0902 | 0,003 |
| | Retenu classe 4 | 0,1055 | | | | 0,0905 | 0,015 |
| | Retenu classe 5 | 0,0911 | | | | 0,0843 | 0,0068 |
| | Retenu classe 6 | 0,0903 | | | | 0,0881 | 0,0022 |
| | Retenu classe 7 | 0,1121 | | | | 0,0775 | 0,0346 |

Comportements des modalités dans l'ACM

4.3 Comparaison avec le S/C dossier-dossier

Un autre indicateur important de la qualité d'un contrat est le S/C dossier-dossier ou SCdd. Cet indicateur est le quotient des sinistres d'un contrat sur une période par ses primes acquises. Il s'agit en somme du S/C sur lequel on n'a effectué aucun traitement. Il présente donc les défauts décrits au paravent : il est extrêmement volatile, surtout pour les petits contrats et peu robuste. En cas de sinistres extrêmes, le S/Cdd peut prendre des valeurs très élevées. A titre d'exemple, sur le portefeuille d'AXA pour la période 2005-2008, le plus mauvais S/C est 4, et moins de 5 contrats dépassent 2. Par comparaison, le pire S/Cdd vaut 130. Le S/Cdd est moins fiable et conduirait de plus à retenir une plus grosse proportion de contrats (20% contre 11). De plus, il est moins précis : les sinistres ne sont pas développés à l'ultime et les primes pas mise sous forme as if. Il paraît donc dangereux de se baser sur cet indicateur pour déterminer les majorations. Néanmoins, on ne peut pas le négliger entièrement. En effet, la direction technique ne fait que suggérer les majorations. C'est ensuite aux souscripteurs de choisir de les appliquer ou non. Or ils se basent essentiellement sur le S/Cdd qui est à la fois plus lisible et plus simple à obtenir. Il est donc souhaitable que le S/C que nous obtenons reste cohérent avec le S/Cdd.

Pour comparer le S/C et le S/Cdd, on va déterminer leur corrélation. Elle vaut 0.25 ce qui paraît assez faible. Néanmoins, une des causes est la présence de S/Cdd extrêmes. On a vu que le S/Cdd pouvait monter jusqu'à 130 alors que le S/C ne dépasse presque jamais 2. En pratique, le plus haut niveau de majoration est atteint dès qu'un contrat dépasse le double du seuil d'équilibre. Avoir un S/C de 2 ou de 130 revient donc concrètement au même. On a ainsi recalculé la corrélation en limitant le S/Cdd à 3 puis à 2. On constate que la corrélation augmente très nettement (respectivement 0.63 et 0.68).

On cherche ensuite à déterminer les facteurs qui génèrent le plus de différences entre les deux indicateurs. On a recalculé les S/C en supprimant certaines étapes de l'algorithme pour voir l'influence sur la corrélation. Le résultat est donné dans le tableau ci contre. Deux facteurs génèrent des différences importantes : la crédibilité et surtout l'écrêtement. Les autres n'influent presque pas.

◇ L'écrêtement influe doublement sur le S/C. Tout d'abord, il évite la présence de certains très gros S/C (petits contrats ayant eu un très gros sinistre par exemple), et on a vu que ces valeurs très élevées diminuaient nettement la corrélation. Ensuite, il modifie aussi le S/C de tous les contrats via l'étape de mutualisation. Typiquement, beaucoup de contrats, (surtout les plus petits) n'ont pas de sinistres sur la période 2005-2008. Leur S/Cdd vaut donc 0. Par contre, à cause de la mutualisation des gros sinistres, leur S/C vaudra entre 0.15 et 0.20. Il y a donc une différence très nette.

◇ La crédibilité agit elle aussi nettement sur le S/C des contrats, surtout pour les plus petits. Elle contribue à « niveler par le milieu » les S/C des contrats.

Les autres facteurs n'ont guère d'effet.

◇ Le développement à l'ultime joue très peu. En effet, les coefficients de développement sont presque tous proche de 1 (sauf pour les sinistres corporels de moins de 2 ans).

◇ Le taux d'inflation retenu est assez faible donc peu influent vu le nombre limité d'années.

◇ La prise en compte de la tendance (bonification/pénalisation des contrats en amélioration/dégradation) joue peu car la plupart des contrats ont un comportement assez erratique et ne sont donc pas concernés.

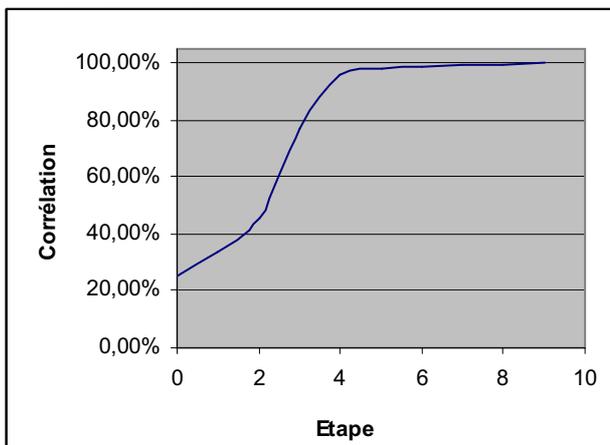
◇ Une dernière différence vient du calcul de la moyenne des 4 observations : dans le cas du S/C, la moyenne est pondérée par les cotisations potentielles. Le S/Cdd est la somme des sinistres sur la somme des cotisations, ce qui revient à pondérer les années par les primes acquises. Dans la pratique, les cotisations potentielles sont généralement assez proches des primes acquises, mais ce n'est pas toujours le cas (par exemple pour des contrats s'étant développés plus vite que prévu). Enfin, on a cherché à transformer par étape le S/C pour approcher le S/Cdd. Le résultat final est donné ci contre. A nouveau on voit le rôle majeur de l'écrêtement. On retrouve à la fin le même résultat.

| | Corrélation entre le S/C et les S/Cdd |
|---------------------------------------|---------------------------------------|
| Sans changement | 0,2546 |
| Avec un S/Cdd plafonné à 3 | 0,6326 |
| Avec un S/Cdd plafonné à 2 | 0,6816 |
| Sans crédibilité | 0,414 |
| Sans écrêtement | 0,773 |
| Sans développement | 0,2527 |
| Sans tendance | 0,263 |
| Sans mise sous forme as if des primes | 0,2543 |
| Sans inflation | 0,2548 |

Influence des différents facteurs sur la corrélation entre le S/C et le S/Cdd.

| Etape | Changement | Corrélation |
|-------|-------------------------------------|-------------|
| 0 | | 25,46% |
| 1 | Sans mutualisation | 33,75% |
| 2 | Sans écrêtement par classe | 45,25% |
| 3 | Sans écrêtement global | 77,30% |
| 4 | Sans crédibilité | 95,89% |
| 5 | Avec poids par les primes | 97,89% |
| 6 | Sans prise en compte de la tendance | 98,40% |
| 7 | Sans développement des sinistres | 99,10% |
| 8 | Sans mise en as if des primes | 99,60% |
| 9 | Sans inflation | 100,00% |

Passage par étape du S/C au S/Cdd



Dans la pratique, le facteur à déterminer n'est pas le S/C mais la majoration à effectuer. On doit donc comparer les majorations obtenues en utilisant le S/C et le S/Cdd. Pour des raisons de lisibilité, on a arrondi la majoration indiquée. Cela a conduit à une segmentation en 7 groupes des contrats repérés (le groupe 1 n'étant pas majoré quand le groupe 7 l'est de 100%). On les nomme classe de majoration 1 à 7. On regroupe dans la classe de majoration 0 les contrats en dessous du seuil d'équilibre (pour lesquels nous ne recommandons pas de majoration). Le détail de ces classes est donné au 3.3.5. On va donc comparer les résultats obtenus par les deux méthodes, S/C et S/Cdd.

Si on regarde juste le fait de majorer ou non un contrat, les deux algorithmes coïncident dans 78% des cas. Ils retiennent effectivement les mêmes contrats. Par contre, les niveaux de majorations préconisés sont assez différents. La méthode dossier/dossier tend à suggérer des majorations beaucoup plus élevées. En particulier, 2600 contrats dépassent le double du seuil d'équilibre (classe 7) contre seulement 53 avec l'algorithme. Ces contrats catastrophiques en dossier/dossier mais juste légèrement mauvais pour nous sont principalement de deux types :

★ des contrats ayant eu un sinistre extrême. Les deux tiers de nos contrats ont moins de 10000 euros de primes annuels, donc un sinistre de 50000 euros leurs assurent la classe 7.

Dans notre méthode, ces contrats sont protégés par l'écrêtement.

★ des petits contrats ayant eu un nombre de sinistres plus élevé que la moyenne. Pour un contrat de 1000 euros ou moins (20% du portefeuille), il suffit de deux sinistres assez élevés pour atteindre la classe 7.

Dans notre méthode, ces contrats sont protégés par la crédibilité.

Au final, dans ces deux cas, il semble préférable de pratiquer des majorations plus faibles. Le premier cas (un sinistre exceptionnel) relève généralement du hasard. Le second incite aussi à la prudence. On rappelle que la classe de majoration 7 correspond à un doublement de la prime donc la plupart du temps à une résiliation du contrat.

Dans la grande majorité des cas, les contrats repérés sont plus majorés par la méthode dossier/dossier que par l'algorithme. Néanmoins pour environ pour 2000, c'est le contraire. Deux cas de figure principaux :

★ des contrats dont le S/C est en augmentation depuis plusieurs années. L'estimation du S/C fait l'objet d'une hausse pouvant aller jusqu'à 30%.

★ des contrats qui n'ont pas eu de gros sinistres donc ne profitent pas de l'écrêtement, mais voient leur S/C augmenter du fait de la mutualisation.

Enfin, on a calculé pour chaque contrat la quantité classemajoration - classemajorationdd pour mesurer l'écart entre les deux méthodes. Sur l'ensemble des contrats, 75% sont identiques. Si on se limite aux contrats repérés, près des deux tiers ont au plus 2 classes d'écart. Ainsi, les deux méthodes semblent donner des résultats assez cohérents, ce qui est rassurant.

| S/C | S/Cdd | | | | | | | | | |
|-------|-------|-------|-----|-----|-----|------|-----|------|-------|-------|
| | | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Total |
| | 0 | 25037 | 479 | 170 | 303 | 313 | 406 | 185 | 672 | 27565 |
| | 1 | 1003 | 317 | 122 | 216 | 209 | 311 | 130 | 543 | 2851 |
| | 2 | 169 | 100 | 51 | 77 | 111 | 139 | 71 | 220 | 938 |
| | 3 | 117 | 102 | 66 | 123 | 128 | 221 | 110 | 353 | 1220 |
| | 4 | 44 | 42 | 30 | 75 | 104 | 201 | 73 | 336 | 905 |
| | 5 | 15 | 5 | 13 | 17 | 49 | 139 | 83 | 348 | 669 |
| | 6 | 1 | 0 | 1 | 3 | 3 | 9 | 16 | 75 | 108 |
| | 7 | 1 | 0 | 0 | 0 | 0 | 2 | 3 | 47 | 53 |
| Total | 26387 | 1045 | 453 | 814 | 917 | 1428 | 671 | 2594 | 34309 | |

Répartition des contrats par niveaux de majoration selon l'algorithme et selon la méthode dossier dossier

| | | S/Cdd | |
|-----|------------|-------------|------------|
| | | Non retenu | Retenu |
| S/C | Non retenu | 0,782185432 | 0,1043458 |
| | Retenu | 0,017371535 | 0,09609723 |

Proportion de contrat retenus suivant les deux méthodes

| Classemajo-classemajodd | Ensemble | Contrats repérés |
|-------------------------|----------|------------------|
| -7 | 672 | |
| -6 | 728 | 543 |
| -5 | 756 | 350 |
| -4 | 1048 | 735 |
| -3 | 1097 | 794 |
| -2 | 1139 | 969 |
| -1 | 1165 | 686 |
| 0 | 25834 | 797 |
| 1 | 1305 | 1305 |
| 2 | 323 | 323 |
| 3 | 175 | 175 |
| 4 | 50 | 50 |
| 5 | 15 | 15 |
| 6 | 1 | 1 |
| 7 | 1 | 1 |

| Ecart | Ensemble | Contrats repérés |
|-------|------------|------------------|
| 0 | 0,75298027 | 0,118179122 |
| 1 | 0,07199277 | 0,295225386 |
| 2 | 0,04261273 | 0,191577699 |
| 3 | 0,03707482 | 0,143683274 |
| 4 | 0,03200326 | 0,116399763 |
| 5 | 0,02247224 | 0,054122183 |
| 6 | 0,02124807 | 0,080664294 |
| 7 | 0,01961584 | 0,00014828 |

Ecart entre les majorations par les deux méthodes

4.4 Etude de la stabilité dans le temps

Un problème classique avec les calculs de S/C est la stabilité dans le temps. En effet, le S/C d'un contrat possède une volatilité très importante, en particulier pour les petits contrats qui représentent l'essentiel du portefeuille (typiquement, pour un contrat de 2000 euros de primes, un sinistre IDA fait passer le S/C de 0.20 à 0.80). A titre d'illustration, la corrélation entre le S/C de 2008 (après crédibilité) et celui de 2007 n'est que de 0.4 (et de 0.6 avec celui de 2006). Il est donc important de rendre le modèle le plus robuste possible. Plusieurs étapes de la méthode permettent de limiter les fluctuations du résultat :

- Le S/C est obtenu comme moyenne des 4 derniers S/C annuels, ce qui limite l'influence d'une année anormale.
- L'écrêtement évite d'avoir des valeurs exceptionnellement élevées (seul deux S/C sont supérieurs à 2, et 2% des S/C annuels).
- La crédibilité fait intervenir le résultat moyen de la classe, qui est beaucoup moins volatil que celui des contrats individuels).

Dans cette partie, on cherche à vérifier la stabilité dans le temps des résultats obtenus. On va d'abord comparer les indices globaux (taux de contrats repérés, proportion des différents niveaux de majoration) à divers dates par la méthode que nous avons développé et par la méthode dossier/dossier. Ensuite, on comparera contrat par contrat les S/C obtenus à un an d'écart.

4.4.1 Evolution des résultats au cours du temps

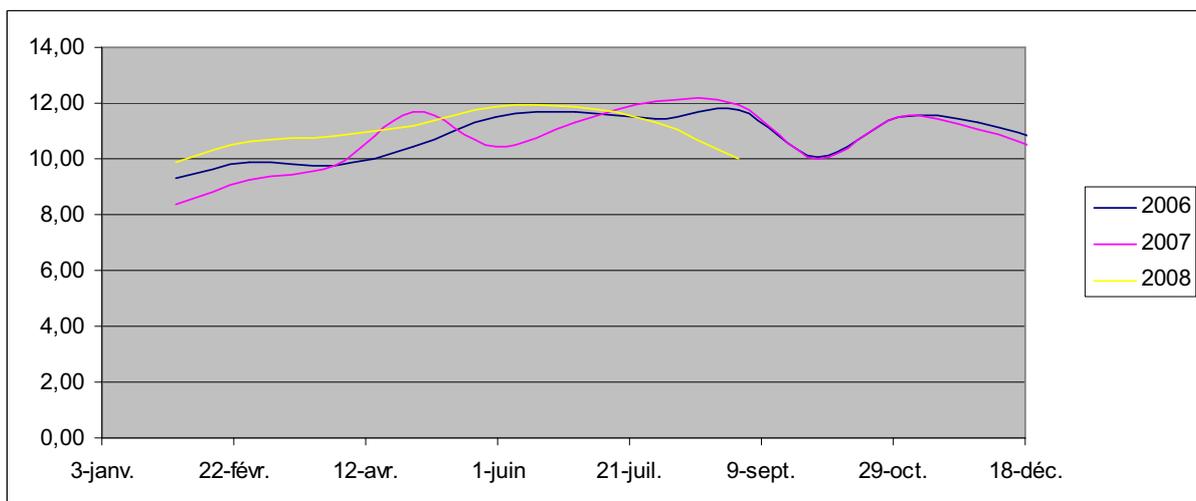
On va étudier la stabilité des résultats fournis par la méthode QP au cours du temps. On s'intéresse d'abord aux résultats globaux. On étudie l'évolution au cours du temps de la proportion de contrats retenus, c'est-à-dire devant faire l'objet d'une majoration. Pour cela, on a appliqué l'algorithme QP à différentes dates d'observation : toutes les fins de mois depuis le début de 2006. A titre de comparaison, on a également déterminé le taux de contrats repérés par le S/C dossier dossier à chacune de ces dates. On constate une variabilité faible mais non négligeable, les résultats semblent osciller autour d'une valeur de 10.5%. De plus, les évolutions des taux de repérés QP et dossier dossier sont parallèles, les variations semblent donc venir de données intrinsèques (en particulier des primes et/ou des sinistres). Il ne semble pas y avoir de biais dans l'algorithme QP lui-même. On remarque que le taux de repérés par notre méthode est plus volatil que celui de l'approche dossier dossier.

Globalement, les résultats des différentes années semblent cohérents entre eux. Au sein de chaque année, on observe une oscillation du taux de repérés mais son amplitude est faible.

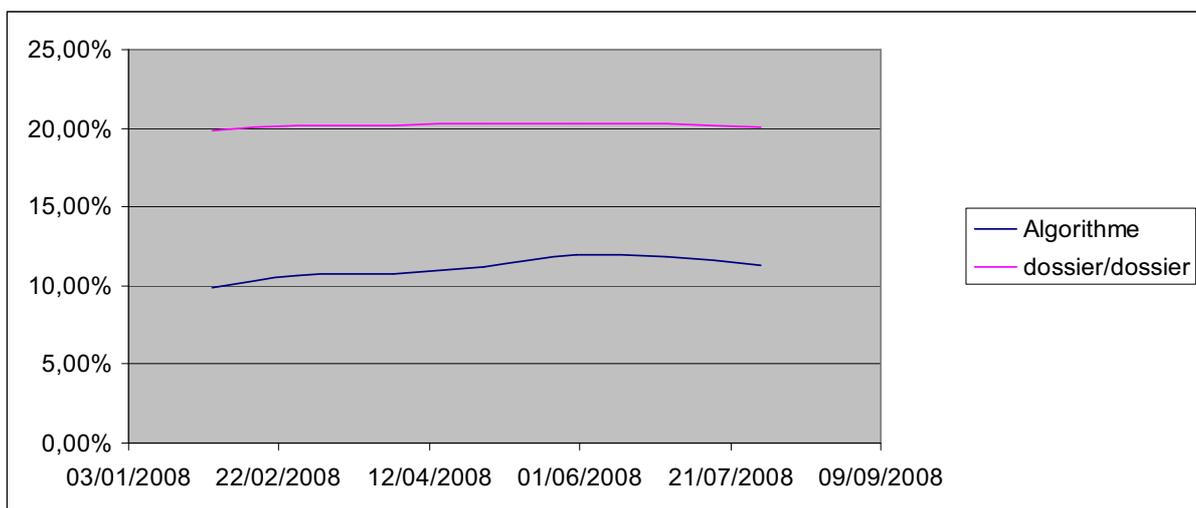
On peut se demander si certains biais n'apparaissent pas quand on examine ce qui se passe au niveau des différentes classes de flottes. On a calculé le taux de contrats repérés chaque mois pour les différentes classes de flottes, les différents usages et les différents poids. Pour les catégories regroupant peu de contrats (TPV, flottes de taille 5 ou 6), la volatilité est nettement plus grande, ce qui était prévisible. Par contre, pour les autres, on ne constate rien de particulier. Les taux se comportent essentiellement de la même façon que pour le portefeuille entier.

On a étudié la proportion de contrats devant être majorés. Un autre facteur essentiel est le niveau de la majoration suggérée. On donne ci contre la proportion de chaque niveau de majoration. Les valeurs obtenues sont extrêmement stables, bien plus que le taux de retenus. Par exemple, les contrats pour lesquels nous recommandons une majoration de 10% représentent toujours entre 23 et 25 % des contrats retenus, ceux devant faire l'objet d'une majoration de 20 % entre 19 et 23%. Seuls les très hauts niveaux de majoration (70 et 100 %) fluctuent, mais cela vient sans doute du fait qu'ils représentent des effectifs beaucoup plus faibles.

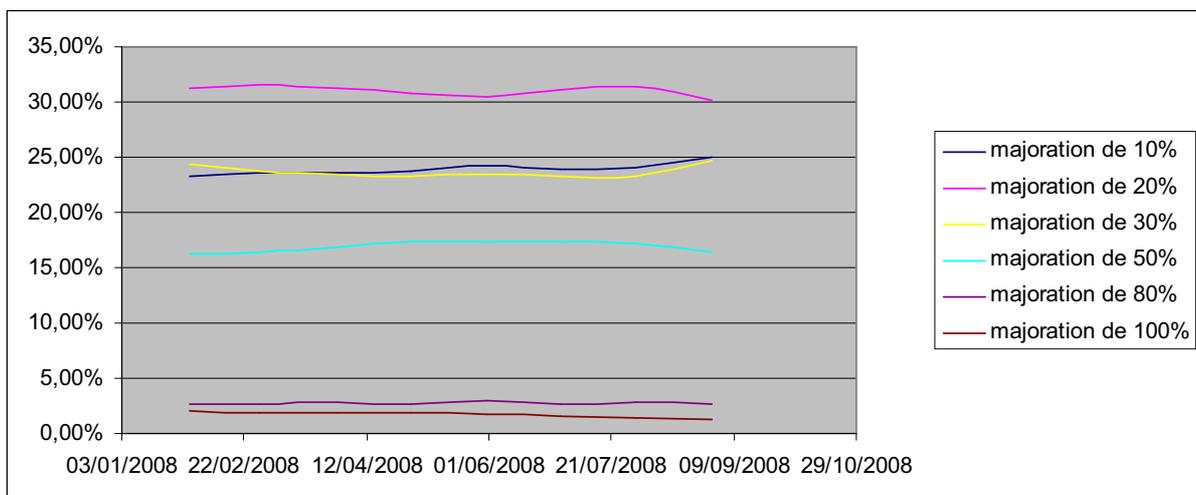
Au final, la structure global du portefeuille semble stable dans le temps, les oscillations autour de la moyenne sont faibles et paraissent peu gênantes.



Proportion de contrats retenus selon la date d'observation



Comparaison entre les taux de repérés avec les méthodes QP et SCdd



Proportion des différents niveaux de majoration recommandées selon la date d'observation

4.4.2 Variation du S/C d'un contrat au cours du temps

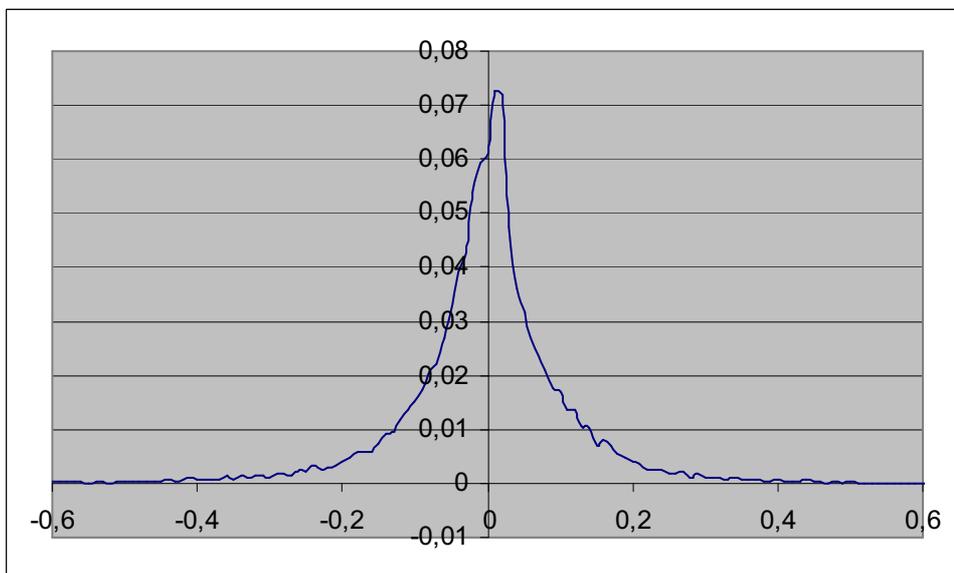
Au paragraphe précédent, on a étudié l'évolution du S/C de façon globale, c'est-à-dire en considérant le portefeuille dans son ensemble. A présent, on va comparer les S/C à deux dates différentes, contrat par contrat. Les S/C annuels que nous avons calculés fluctuent énormément au cours du temps. C'est pour cela que nous avons utilisé une moyenne des quatre dernières années. On peut se demander si cela a suffi à limiter ces variations. On a donc calculé les S/C des différents contrats du portefeuille à 1 an d'écart (31 juillet 2007 et 31 juillet 2008). On peut ainsi voir comment ces ratios ont pu évoluer. Logiquement, sauf dans quelques cas exceptionnels, il devrait y avoir eu peu de variations.

Un problème se pose : entre ces deux dates, certains contrats ont fait l'objet de majorations, et il est logique que leur S/C ait évolué. On a donc mis les S/C 2007 sous forme as if. Comme pour l'algorithme QP, on a utilisé les coefficients techniques CT qui mesurent l'écart entre tarif de référence et tarif appliqué. La formule est :

$$SC_{\text{as if}} = SC_{2007} * \frac{CT_{2007}}{CT_{2008}}.$$

L'étude ne porte donc que sur les contrats pour lesquels ces coefficients sont disponibles, c'est-à-dire sur les flottes fermées. Au total, on dispose de 21000 contrats existant depuis au moins 2007 et pour lesquels le SC as if a pu être calculé. On donne la loi de SC2008-SC2007 as-if. Comme on pouvait l'espérer, la loi est concentrée autour de 0. La dispersion est faible mais non négligeable. Dans 75% des cas, les S/C diffèrent de moins de 0.1 et dans 54% des cas de moins de 0.05. Un point positif est qu'il ne semble pas y avoir de biais systématique. La moyenne est quasi nulle. Il n'y a donc pas de différence structurelle entre les deux années. La méthode semble donc plutôt cohérente dans le temps.

On s'intéresse aussi au niveau de majoration recommandé par l'algorithme sur la base des S/C 2008 et 2007 as if. Comme précédemment, on attribue la valeur 0 aux contrats non retenus, et des valeurs 1 à 7 pour les contrats retenus (plus la valeur est grande, plus la majoration recommandée est importante). Logiquement, dans la plupart des cas on doit trouver la même classe de majoration pour les deux années. Un écart de 1 indique un contrat ayant légèrement évolué, par exemple étant de peu sous le seuil d'équilibre en 2007 et qui a basculé au dessus en 2008. Les écarts de 3 ou plus indiquent des changements plus brutaux du S/C. Une trop forte proportion de gros écarts serait donc un signe d'instabilité numérique. Au final, les résultats sont rassurants. Dans 76% des cas, les majorations obtenues les deux années coïncident. Les écarts de 3 classes ou plus ne représentent que 5% des contrats, ceux de 4 classes ou plus moins de 2%. Ces cas extrêmes sont essentiellement des contrats dont le mode de fonctionnement a évolué. Le coefficient technique CT ne permet pas dans ce cas de bien évaluer la majoration entre 2007 et 2008 et le S/C 2007 as if est faux. Enfin, il y a bien sur quelques cas où la charge des sinistres en 2004 ou 2008 est particulièrement élevée (d'où un écart entre les deux S/C).



Densité de SC2008-SC2007 as if

| | | 2008 | | | | | | | |
|------------|---|-------|-----|-----|-----|-----|----|----|---|
| | | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 2007 as if | 0 | 15016 | 987 | 269 | 254 | 157 | 58 | 3 | 1 |
| | 1 | 950 | 413 | 111 | 164 | 106 | 58 | 3 | 0 |
| | 2 | 203 | 156 | 52 | 72 | 64 | 24 | 1 | 1 |
| | 3 | 187 | 213 | 96 | 128 | 75 | 37 | 5 | 2 |
| | 4 | 78 | 100 | 58 | 105 | 104 | 71 | 6 | 1 |
| | 5 | 22 | 22 | 16 | 68 | 69 | 85 | 15 | 1 |
| | 6 | 0 | 0 | 0 | 4 | 6 | 22 | 9 | 1 |
| | 7 | 0 | 0 | 1 | 2 | 2 | 5 | 3 | 7 |

Répartition des contrats selon les niveaux recommandés en 2008 et en 2007 as if (de 0=pas de majoration à 7=100% de majoration)

| | |
|---|--------|
| 0 | 76,22% |
| 1 | 13,17% |
| 2 | 5,27% |
| 3 | 3,37% |
| 4 | 1,54% |
| 5 | 0,41% |
| 6 | 0,01% |
| 7 | 0,00% |

Ecart entre la classe de majoration en 2008 et en 2007 as if

4.4.3 Bilan

On dresse ici un bilan des études sur la stabilité dans le temps du modèle. Globalement, l'algorithme paraît stable dans le temps au niveau macroscopique mais plus volatil si l'on regarde contrat par contrat.

◇ La structure global du ratio S/C varie très peu dans le temps. La forme de la courbe de S/C contrat/ S/C équilibre reste remarquablement stable, même si elle subit des translations au cours du temps. En particulier, la proportion des différents niveaux de majoration recommandés ne varie quasiment pas.

◇ Le S/C moyen semble osciller autour de sa moyenne, mais l'amplitude est faible et il est dur de dire s'il s'agit du hasard ou d'un phénomène récurrent.

◇ Pour conclure sur l'étude au niveau macro, il ne semble pas y avoir de biais systématique.

◇ Pour les contrats pris individuellement, l'étude faite au paragraphe 3.3.2 montre une dispersion non négligeable. Les valeurs du S/C peuvent évoluer de façon assez nette d'une année sur l'autre. Néanmoins, dans la majorité des cas, ces variations restent raisonnables.

Synthèse du quatrième chapitre

On a analysé les résultats obtenus par l'algorithme de majoration QP. Au final, les conclusions sont plutôt satisfaisantes.

○ Les S/C obtenus semblent assez cohérents. La sinistralité augmente de façon nette avec le niveau de dépassement du seuil d'équilibre. De plus, les S/C que nous avons obtenus restent plutôt proches des ceux utilisés par les souscripteurs (sauf pour les valeurs extrêmement élevées mais qui ne jouent aucun rôle en pratique). De ce point de vue, l'algorithme est donc satisfaisant.

○ La proportion de contrats retenus varie de façon nette selon les différentes variables considérées. Dans la plupart des cas, cela correspond à la réalité du marché. Par exemple, certaines catégories de contrats sont sensiblement moins rentables (grosses flottes, TPM). Quelques biais apparaissent. Dans certains groupes, nous avons du mal à distinguer les mauvais contrats des autres, et le taux de contrats retenus est trop faible. On pense en particulier aux petites flottes et aux flottes récentes. Enfin, certaines variables de tarification comme les régions n'influent pas. Cela correspond au fait que le tarif est bien conçu, ce qui est rassurant.

○ La stabilité dans le temps est souvent le point le plus problématique des méthodes de QP. Notre algorithme n'échappe pas à la règle. Les résultats peuvent varier de façon non négligeable d'un mois à l'autre. Néanmoins, la structure globale du ratio S/C reste assez stable. Au niveau contrat par contrat, la dispersion est plus grande mais paraît acceptable.

Conclusion

Dans ce mémoire, nous avons défini un processus permettant de surveiller les contrats en portefeuille, de détecter ceux qui sont déficitaires et de déterminer la majoration permettant de les ramener à l'équilibre. Pour ce faire, nous avons aussi modélisé la loi de la charge des sinistres en fonction de la taille de la flotte et du type de véhicules.

En ce qui concerne la partie modélisation, on est parvenu à un modèle traitant l'ensemble du portefeuille. La fréquence des sinistres suit une loi binomiale négative, les coûts en dessous de 15000 euros une loi log normale et ceux au dessus une loi de Pareto. Les paramètres de la fréquence dépendent de la taille de la flotte, ceux du coût du type de véhicules. Les adéquations des lois varient mais restent globalement bonnes. Pour conserver des classes suffisamment grandes, on n'a pas tenu compte des différents usages de véhicules ce qui est regrettable.

Pour le calcul du S/C, on écrete d'abord les coûts de sinistres. Le seuil est défini comme un quantile de la loi de coût total et varie selon la taille et les types de flottes. On en déduit les S/C annuels des contrats pour l'année en cours et les 3 années précédentes. Le S/C propre est la moyenne de ces ratios (pondérés par les cotisations potentielles). Une correction est effectuée de façon à tenir compte d'une évolution dans le temps des résultats. Enfin, le S/C est crédibilisé suivant une méthode de Jewell à trois niveaux, les classes étant composées de flottes de taille et de type de véhicule équivalents.

La méthode que nous avons développée a débouché sur un programme SAS. Il permet à n'importe quelle date de déterminer le S/C des contrats et de repérer ceux posant problème. Le temps de traitement reste assez court : de l'ordre de 3 à 5 minutes. De plus, les paramètres ont été ajustés de façon à ce que le taux de contrats repérés soit raisonnable, pour permettre un traitement rapide par les souscripteurs. Les résultats obtenus semblent être cohérents, conformes à nos intuitions et en phase avec les indicateurs utilisés par les agents généraux (en particulier le S/C dossier dossier).

Quelques problèmes se posent néanmoins. Un effet poids apparait. Les petits contrats sont très peu crédibilisés et donc trop protégés. Nous en repérons sans doute trop peu. Un autre problème (classique pour ce genre de méthode) est l'instabilité dans le temps. La crédibilité et l'écrêtement la limite mais elle reste notable. Enfin, on se sert de données anciennes (parfois vieilles de 3 ans) et qui peuvent ne plus être pertinentes (le contrat a pu évoluer depuis).

Pistes de développement

Certains aspects du modèle de sinistralité et de la méthode de majoration QP posent problème et pourraient être améliorés. Nous en présentons plusieurs ainsi le cas échéant que des pistes que nous avons suivies.

• **Prise en compte de l'usage**

Nous avons montré au paragraphe 2.2 que la loi des coûts dépendait de l'usage des véhicules. Cependant, pour des raisons de tailles de classes, nous n'en n'avons pas tenu compte (nous avons privilégié le critère du type de véhicules). Il serait peut être souhaitable de développer un modèle croisant ces deux critères.

★ Pour la crédibilité, on pourrait imaginer un modèle hiérarchique à trois niveaux intermédiaires : le portefeuille est divisés en six classes selon les tailles des contrats, elles même divisées en deux catégories (M3T5 et P3T5) qui seraient découpées en trois sous-catégories (AUT TPM TPV). Cependant, on a vu que le deuxième niveau n'apportait presque rien dans la pratique. On peut donc s'attendre à ce que le troisième ne change pas grand-chose non plus.

★ Pour l'écrêtement, on pourrait définir un seuil en fonction à la fois du type de véhicules et de l'usage, soit six cas possibles, avant de croiser avec les tailles de flottes. Le problème vient surtout de la mutualisation. Certaines sous-catégories comprendraient peu de contrats. Si l'un d'entre eux avait un sinistre majeur, cela impacterait de façon trop nette le résultat des autres. On pénaliserait donc fortement des contrats pour des sinistres qu'ils n'ont pas eus. Une solution envisagée était de faire dépendre le seuil d'écrêtement des trois facteurs : taille, type et usage mais au moment de la mutualisation de ne plus tenir compte de l'usage. Cette solution est néanmoins complexe, peu lisible et discutable sur le plan mathématique.

• **Prise en compte de la modification des contrats**

Dans notre calcul de moyenne, nous utilisons les résultats des trois années précédentes. Or les données les plus anciennes ne sont pas toujours pertinentes. Par exemple, des véhicules peuvent être plus anciens et donc moins surs. A l'inverse, un contrat peut avoir eu de mauvais S/C du fait de véhicules vétustes et les avoir ensuite changer. Ou avoir amélioré son système d'entretien. Dans ce cas, les données passées ne devraient plus être considérées, ou alors avec un poids moindre. Un autre problème du même type peut se poser. Si un contrat a été mauvais, il est majoré. La plupart du temps, cette majoration se traduit par une hausse de la prime et nous en tenons compte dans la mise sous forme as if. Parfois, l'assuré peut préférer une autre solution comme la hausse de ses franchises. Or nous ne disposons pas toujours de ses données. Cela ne se traduit pas par une hausse du CT (qui ne dépend que de la prime) et ne permet donc pas une mise sous forme as if des années antérieures. Dans ce cas, nous ne prenons donc pas en compte la majoration passé, et nous conservons des S/C anciens qui sont anormalement mauvais.

Nous n'avons pas trouvé de solutions simples à ces problèmes. Dans la pratique, l'information dont nous disposons est souvent insuffisante. De plus, chaque contrat est un cas particulier, mais nous devons conserver un traitement standardisé. Au final, ce sont sans doute les souscripteurs qui

sont les plus à même de gérer ces questions, car ils sont en prise directe avec les clients.

• **Prise en compte des petits poids :**

Un défaut de l'algorithme est qu'il est trop clément vis-à-vis des contrats de poids faibles. Ceux-ci sont en effet très peu crédibilisés, et donc leur expérience propre compte peu. Cela touche deux catégories de contrats : les petites flottes et les contrats récents.

★ Les petites flottes sont généralement bien meilleures que la moyenne, mais on peut se demander si on ne devrait tout de même pas en repérer plus. Au final, environ 3% des contrats taille 1 dépasse le seuil d'équilibre contre 15% si l'on se fie au S/Cdd. Cela vient en partie de la trop grande volatilité du S/Cdd mais aussi d'une protection excessive due à la crédibilité.

★ Les contrats nouveaux ont un poids faible, car leur poids pour les années passées est nul. Ils sont donc peu crédibilisés et moins repérés. Cela pose un problème car en pratique, les affaires nouvelles sont en moyenne moins bonnes que les contrats anciens. En effet, nous n'avons pas encore pu trier les bons contrats et traiter les autres. Néanmoins, il paraît difficile de juger un contrat sur une seule année d'observation, et ce phénomène est sans doute un mal nécessaire.

Une solution envisagée est d'augmenter le poids pour les petites flottes. Au lieu de pondérer par les cotisations, on considérerait $\psi(C)$ où ψ serait une fonction croissant, supérieure à l'identité sur $[0; 5000]$ (donc pour les classes 1), et équivalente à l'identité en $+\infty$. Plusieurs fonctions ont été testées. La plus intéressante est :

$$\psi(t) = ((t - 5000)^3 + 5000^3)^{\frac{1}{3}} \text{ si } t \leq 5000 \text{ et } \psi(t) = t \text{ sinon.}$$

Néanmoins, cette approche rend le modèle moins lisible, et augment nettement le taux de contrats retenus, ce qui n'est pas souhaitable.

Une autre solution est de modifier le calcul de la crédibilité pour les petits contrats, en diminuant la constante K pour les contrats de classe 1. Il s'agit d'une solution ad-hoc, sans fondements mathématiques. De plus, dans ce cas la crédibilité n'est plus une fonction croissante du poids.

Une dernière solution serait de recourir pour certains types de contrats, nouveau ou de petite taille, à un procédé reposant uniquement sur la fréquence. Celle-ci est en effet beaucoup plus stable que la charge totale et la crédibilité serait donc plus forte. Cette approche présente néanmoins plusieurs défauts. Tout d'abord, si on ne tient compte que de la fréquence, on oublie le coût total alors qu'il s'agit d'une grandeur réelle, que l'assureur devra verser. De plus, pour les plus petites flottes, cette approche atteint aussi ces limites. Pour une flotte de moins de 5000 euros, le nombre moyen de sinistre par an est inférieur à un (de l'ordre de 0.3). Une année à un sinistre n'est guère inquiétante. Une année à deux sinistres ne représente qu'un événement de plus que précédemment, et relève sans doute du hasard. On ne devrait donc à priori s'inquiéter qu'à partir de trois sinistres dans une même année ce qui reste très rare.

• **Stabilité dans le temps**

Le S/C que nous calculons reste moyennement stable dans le temps. D'une année sur l'autre le S/C d'un contrat peut varier de façon nette, même si cela reste rare (dans 1.5 % des cas, le S/C a évolué de plus de 0.5 en un an). Cela vient sans doute en partie de la faible pondération des affaires nouvelles, dont le S/C varie nettement les premières années. Une solution serait de mesurer plus finement leur risque, peut être en augmentant leur poids. Pour les autres contrats, on pourrait stabiliser le S/C en faisant intervenir un plus grand nombre d'années, mais les données trop anciennes risquent de ne pas être cohérentes.

- **Prise en compte du risque de résiliation**

Quand nous choisissons de majorer une police, et à fortiori de la majorer de façon importante, le client n'est pas obligé d'accepter. Il peut préférer résilier son contrat. Nous devons tenir compte de ce risque. Il serait donc intéressant de voir dans quelle mesure les majorations QP influent sur les résiliations et si certains types de contrat les acceptent plus facilement que d'autres. A partir de là, on pourrait éventuellement ajuster nos suggestions de façon à tenir compte des objectifs de chiffre d'affaire.

Bibliographie

1) Mémoires

- *Surveillance du portefeuille de flottes automobiles d'AXA* Marpillat Mémoire ISUP 2005
- *Mise en place d'une nouvelle tarification en automobile entreprise* Heulin Mémoire IMA 2000
- *Prédiction de l'acte de résiliation de l'assuré et optimisation de la performance en assurance automobile particulier* Solange Hamel ENSAE 2007

2) Ouvrages

- *Assurance non vie* Partrat Besson Economica 2005
- *Mathématiques de l'assurance non vie* Charpentier Denuit Economica
- *A course in credibility theory and its applications* Buhlmann Gisler Springer 2006
- *Méthodes statistiques* Tassi Economica 1985

3) Articles

- *Rate making and large claims* Gigante Picech Sigalotti Astin bulletin 2001
- *Estimating the tail of loss severity distribution using extreme value theory* Mc Neil Astin bulletin 1997 bulletin
- *Optimum trimming of data in the credibility model* Gisler Journal des actuaires suisses 1980
- *Excess claims and data trimming in the context of credibility rating procedure* Buhlman Gisler Journal des actuaires suisses 1980
- *A comparison of three credibility formulae using multidimensional techniques* Buhlmann Astin bulletin 1974

4) Documents

- *Cours de crédibilité* Marie-Jeanne
- *Cours de SAS* Preud'homme

Liste des abréviations

S/C : Ratio entre le coût des sinistres d'un contrat survenue et ses primes acquises (on parle aussi souvent de S/P).

M3T5 : véhicule (ou flotte de véhicules) de moins de 3.5 tonnes.

P3T5 : véhicule (ou flotte de véhicules) de plus de 3.5 tonnes.

TPM : véhicule (ou flotte de véhicules) servant à transporter des marchandises n'appartenant pas à l'assuré (transport public de marchandise)

TPV : véhicule (ou flotte de véhicules) servant à transporter des voyageurs qui ne sont pas salariés de l'assuré (transport public de voyageur).

TPC : véhicule (ou flotte de véhicules) servant à transporter des salariés de l'assuré ou des marchandises lui appartenant (transport pour propre compte).

AUT : Flotte qui ne sont ni TPM ni TPV

BDG : bris de glace

RCCORP : sinistres en responsabilité civile faisant intervenir des dégâts corporels

RCMAT : sinistres en responsabilité civile faisant intervenir des dégâts matériels

DOMM : sinistres impliquant des dégâts sur le véhicule de l'assuré

IDA : indemnisation directe de l'assuré. Convention entre les assureurs automobiles visant à ce que chacun indemnise ses propres assurés (moyennant une indemnité forfaitaire). Ne s'applique qu'aux petits sinistres.

EVT : théorie des valeurs extrêmes (extreme value théorie)

POT : méthode consistant à ajuster une loi de Pareto généralisée sur les coûts en excédent d'un certain seuil (peak over threshold).

DAM : domaine d'attraction maximale d'une loi. Ensemble des variables dont le maximum empirique (après normalisation) converge vers la loi donnée.

GPD : loi de Pareto généralisée (generalized Pareto distribution).

QP : ensemble des méthodes visant à améliorer la qualité du portefeuille. Dans ce mémoire, majoration des contrats déficitaires pour les ramener à l'équilibre.

BS : Buhlmann Straub

S/Cdd : S/C dossier-dossier, c'est à dire sans traitement