ISUP

PROMOTION 2009

Mémoire présenté devant

l'Institut de Statistique
de l'Université Pierre et Marie Curie

Pour l'obtention du

Diplôme de Statisticien
Mention Actuariat

Assurance ■ ***Finance*** □

Par Mlle Noémie ROSE

Sujet : Provisionnement en assurance non-vie : Utilisation de modèles paramétriques censurés

Lieu du stage : WINTER & Associés

Responsables du stage : Laurent FAUCILLON / Michaël DONIO

Invité : Frédéric PLANCHET

CONFIDENTIEL □

Sommaire

Remerciements

En préambule de ce mémoire, je souhaitais adresser mes remerciements aux personnes qui m'ont apporté leur aide et contribué à l'élaboration de ce mémoire.

J'aimerais remercier tout d'abord le Cabinet WINTER & Associés de m'avoir accueilli et fait confiance durant mon stage et par la suite. J'adresse une attention particulière à Michaël DONIO ainsi qu'à toute son équipe pour leur aide et soutien.

Je remercie Laurent FAUCILLON et Abder OULIDI pour leur encadrement et le temps qu'ils m'ont consacré.

Je tiens également à remercier Jérôme BEAUVIR pour la grande disponibilité dont il a su faire preuve ainsi que Frédéric PLANCHET pour ses nombreux conseils avisés qui m'ont permis d'avancer tout au long de mon étude.

Résumé

Dans le cadre des assurances IARD, la complexité liée à l'estimation des provisions réglementaires résulte du choix des méthodes à employer.

Ces méthodes sont essentiellement statistiques et en fin de compte assez peu règlementées, le législateur s'assure principalement que le montant des provisions constituées couvre la sinistralité future. Si la grande majorité des assureurs ont recours à des techniques de mutualisation telles que les méthodes de *Chain-Ladder* introduites dans ce mémoire ; il apparait que ces méthodes sont moins bien adaptées dans le cas de portefeuilles de sinistres ayant une forte volatilité des fréquences et des montants et avec des cadences de règlements s'étalant sur de longues durées (comme c'est le cas des assurances de responsabilité civile ou les assurances de construction garantie décennale).

L'objectif de ce mémoire a donc été d'avoir recours à une méthode qui est généralement utilisée pour le développement de tables de mortalité, en l'occurrence les modèles censurés, et de voir dans quelle mesure l'application de ces modèles au monde de l'assurance non-vie permettrait d'appréhender avec plus de précisions l'estimation de ces provisions de sinistres dans les cas de figure les plus extrêmes.

L'application en tant que telle de ce type de modèle a nécessité un certain nombre de travaux et de retraitements permettant d'adapter le contexte de l'assurance non-vie à la théorie des modèles censurés.

Ce mémoire s'est donc intéressé à un portefeuille de construction garantie décennale et s'est décomposé selon les étapes suivantes :

- ✓ Le chapitre 1 décrit de manière générale la théorie des modèles censurés et en particulier leur application aux modèles IARD. Ces dispositifs nécessitent des retraitements au niveau des données pour la triangularisation ou les modèles de durée. L'application des modèles de durée introduit une notion de censure qui renseigne sur la clôture du sinistre ou son caractère encore en cours.
- ✓ Le chapitre 2 quant à lui décrit les modèles paramétriques qui ont été retenus en parallèle des données empiriques exploitées à l'aide de Kaplan-Meier pour la caractérisation des sinistres. Ces modèles classiquement retenus sont les modèles de Pareto et de Weibull.
- ✓ Le chapitre 3 s'est intéressé à la mise en œuvre pratique des modèles paramétriques et non paramétriques avec le choix des différentes lois et à l'application des indicateurs d'acceptation ou de refus des lois d'adéquation.

Abstract

In the IARD insurance, the choice of the adapted methods leads to complexity related to regulatory reserves.

The methods are mainly statistical and ultimately not much regulated, the legislator checks totalament that the amount of the collected reserves covers for the future total loss experience. If the great majority of the insurance companies employs mutualization techniques such as the *Chain-Ladder* methods introduced in this treatise, these methods appear to be less fitted to the case of loss portfolio with a high volatility of the frequencies and the amounts; and with long-term settlement rhythms (as it's the case with liability insurance or decennial guarantee).

This treatise originally presents a method generally used for mortality tables development, in this case censured models, and its purpose is to see to which extent these models application to the field of IARD insurance would allow us to comprehend more precisely the estimation of these casualty provisions in the most drastic eventualities.

In itself, this kind of model's application required quite an amount of works and reprocessings in order to adapt the IARD insurance context to the censured models' theory.

Thus, this treatise looked into a decennial guarantee portfolio and divided in several sections:

- ✓ Section 1 addresses the presentation of the censured models theory, especially their implementation to the IARD models. This induces data treatments for *Chain-Ladder* models or duration models. The application of the latter introduces a censure notion which informs whether a sinister is ended or still in process.
- ✓ Section 2 deals with parametric models retained as empirical data were used with the help of Kaplan-Meier for the sinisters' characterization. These referenced and common models are Pareto and Weibull models.
- ✓ Section 3 looks at applying the parametric and non-parametric models, with the choice of the different laws; and the acceptance or refusal of the appropriateness laws indicators.

Introduction

A. Contexte réglementaire

D’après le Code des Assurances, Article R 331 – 6, la provision pour sinistres à payer (PSAP) est la valeur estimative des dépenses en principal et en frais, tant internes qu'externes, nécessaires au règlement de tous les sinistres survenus et non payés, y compris les capitaux constitutifs des rentes non encore mises à la charge de l'entreprise.

Cette provision imposée aux assurances par la règlementation vise à protéger les assurés, souscripteurs et bénéficiaires des contrats. En effet, le cycle de production inversé présent dans les compagnies d’assurance conduit celles-ci à mettre en place un système d’estimation des charges futures le plus précis possible. Un trop faible provisionnement pourrait conduire les compagnies à la faillite et avec une trop forte exigence, la mise en provisionnement ne pourrait parfois pas être réalisée.

Ce mémoire porte sur le calcul de la provision pour sinistres à payer et plus précisément sur la provision à mettre en place face aux sinistres déjà survenus et déclarés mais dont la totalité du coût n’est pas encore réglée. Les IBNR sont donc écartés de l’étude.

L’étude de la provision pour sinistres à payer se fait au moment de l’inventaire donc au 31/12/n. Un sinistre contre lequel l’assuré est garanti pendant une certaine période intervient au cours de l’année n à une date dite de survenance. Ce sinistre est déclaré par l’assuré à la compagnie en date de déclaration. C’est à compter de cette date que l’assurance va commencer à rembourser le sinistre. Or, le paiement des sinistres ne s’effectue pas toujours en une fois. Il peut s’étaler au fil du temps et il est nécessaire de constituer des réserves pour pouvoir honorer les dettes futures qui tiennent compte de cet effet de cadencement. Or, dans la mesure où le montant qui sera finalement payé pour le sinistre est inconnu au départ, la somme à mettre en réserve est également inconnue et il faut l’estimer.

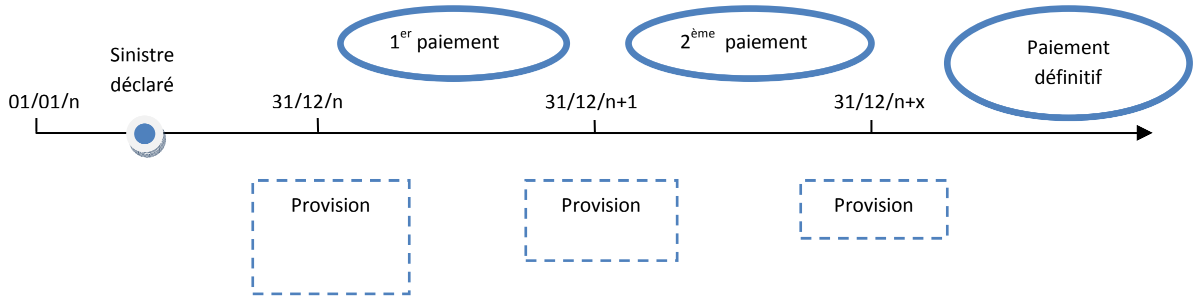


Schéma 1 : Paiements et Provisions d'un sinistre

Le schéma précédent représente l'évolution dans le temps pour un sinistre déclaré en année n de la provision à mettre en place chaque année ainsi que les montants des paiements réglés. La provision doit être réévaluée chaque année en fonction des nouvelles informations mais aussi des paiements effectués et à effectuer. Dans la plupart des cas, la provision est décroissante au fil du temps.

La PSAP représente donc la dette existante envers les assurés lorsque les sinistres ont été déclarés mais non encore réglés. Le calcul de cette provision est très important et de nombreuses méthodes sont développées autour de ce sujet.

B. Evolution des méthodes de calcul de la PSAP

Les méthodes traditionnellement employées dans le calcul de la provision pour sinistres à payer sont des méthodes dites statistiques qui reposent principalement sur les données historiques de la sinistralité. Elles sont de deux catégories :

- ✓ les méthodes déterministes ;
- ✓ les méthodes stochastiques.

Les quantités analysées sont des paiements de sinistres qui sont classés annuellement. Les sinistres sont rattachés à des années d'origine de sinistres survenus et les montants remboursés qui peuvent être exprimés annuellement ou en montants cumulés sont ensuite rangés par année de développement. Les années de développement correspondent aux années au cours desquelles les sinistres sont remboursés par rapport à leur année d'origine. Ces différentes données ainsi classées constituent des triangles de liquidation.

La méthode déterministe la plus employée permettant d'estimer les réserves de sinistres est la méthode de Chain-Ladder. Elle s'établit en deux parties. Dans un premier temps, des facteurs de développement correspondant au passage d'une année de développement à l'autre sont estimés. Dans un second temps, ces facteurs sont appliqués aux montants de sinistres déjà connus pour anticiper les règlements futurs, étant entendu qu'il est supposé que la cadence future de règlement sera à l'image de la cadence moyenne déterminée au cours de la première étape. Le triangle est ainsi complété pour former un rectangle de règlement. La différence entre le montant total à régler et le montant déjà remboursé constitue la réserve de sinistre ou provision.

Cette méthode est facile à mettre en place mais fait néanmoins apparaitre des insuffisances. De manière générale, les résultats peuvent être considérés comme inexacts dès lors que l'on observe dans les données :

- ✓ des irrégularités des cadences de règlement au sein par exemple des exercices plus récents;
- ✓ un nombre de contrats trop faible ;
- ✓ une fréquence de sinistre trop faible ;
- ✓ une dispersion des montants des sinistres élevés.

Par ailleurs, la méthode de Chain-Ladder ne permet pas d'intégrer certains éléments tels qu'une réassurance non proportionnelle (les sinistres sont agrégés par années d'origine) ou bien d'obtenir un intervalle de confiance (le modèle est déterministe).

Ce dernier point s'améliore avec la mise en place de méthodes stochastiques. Le modèle de Mack est un premier modèle stochastique pour la méthode Chain-Ladder. Il permet de dégager une estimation de la volatilité de l'estimateur des provisions techniques. Ce modèle stochastique est conditionnel, il détermine les espérances connaissant les réalisations des paiements présents dans le triangle de liquidation. Le modèle de Mack fournit donc des informations plus importantes sur les provisions à constituer.

Néanmoins, il se base toujours sur des montants de sinistres qui ont été agrégés entre eux par année d'origine et année de développement et d'après l'article R 331 – 15 du Code des Assurances, la réglementation prévoit que le montant de la provision pour sinistres à payer soit établie dossier par dossier.

C. Nouvelle méthode envisagée dans le cadre de ce mémoire

Ce mémoire propose une approche différente des modèles par triangulation. En effet, au lieu de fonder les estimations des coûts de sinistres à partir d'un historique de développement des paiements, l'objet de ce mémoire est d'apporter une méthode de calcul dossier par dossier, sinistre par sinistre dans l'évaluation de la provision.

Pour ce faire, il est nécessaire d'avoir des informations sur les sinistres considérés de façon individuelle. Pour chaque sinistre survenu durant la période d'étude, il faut en effet connaitre la cadence des règlements déjà réalisés pour chaque exercice comptable.

Les données des sinistres ne sont donc plus regroupées par année de survenance et année de développement mais chaque sinistre est considéré individuellement et tous les montants réglés pour un sinistre donné sont agrégés au moment de l'estimation.

La provision totale à mettre en place pour le portefeuille étudié est estimée en tant que somme des provisions individuelles correspondant à un sinistre. C'est seulement lorsque la provision de chacun est évaluée que les sinistres entre eux sont agrégés afin de connaitre la provision globale.

Le but est donc d'estimer dans un premier temps la loi du coût total d'un sinistre appartenant à un portefeuille particulier et ce à partir des sinistres historiques – donc clos – et des sinistres encore en cours à la date d'inventaire. Ainsi, les coûts totaux des sinistres en cours ainsi que le montant de la provision pour sinistres à payer sont déduits. Cette provision est constituée pour un sinistre individuel. Les provisions correspondantes à chaque sinistre sont ensuite cumulées.

D. Données censurées

La particularité du portefeuille de sinistres étudié réside dans la présence de sinistres clos et de sinistres en cours au moment de l'inventaire. Les notions de censure interviennent donc dans l'étude des montants de sinistres.

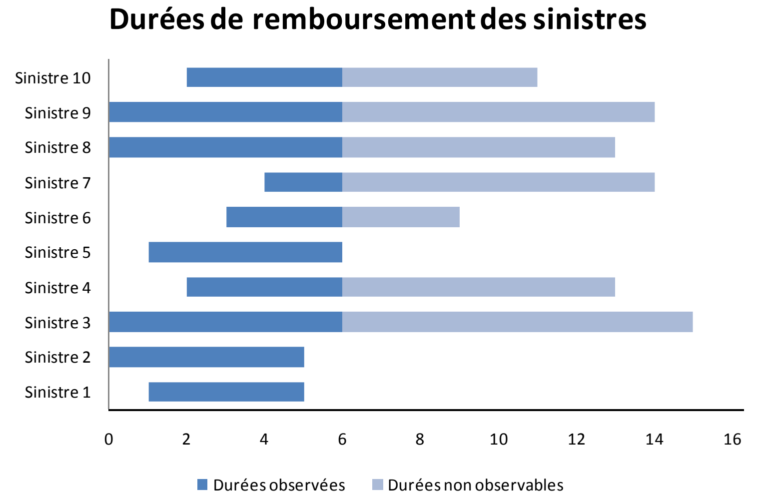


Schéma 2 : Représentation de la censure dans les durées de paiements

Le graphique ci-dessus représente les durées de remboursement en années de 10 sinistres d'un portefeuille construit à titre illustratif. Seule la durée est prise en considération. La durée de remboursement est comprise entre 0 et 15 ans. La date d'étude ou date d'inventaire correspond à l'année 6. Pour les sinistres qui n'ont pas totalement été réglés en année 6, il y a censure concernant la durée de dédommagement du sinistre. En parallèle, il y a censure sur le montant du sinistre. Dans l'exemple ici représenté, l'information « montant total de sinistres » n'est connu que pour les sinistres 1, 2 et 5. Pour les autres, il y a censure de cette information.

La première variable d'étude, notée X est le montant total d'un sinistre. Ce montant n'est pas connu pour chaque sinistre. En utilisant les données du portefeuille, la loi de cette variable X est estimée à partir des sinistres clos et des sinistres dont le montant est censuré. Ainsi, au lieu d'observer la réalisation de la variable d'étude dans son intégralité, seule la borne inférieure est connue dans le sens où si un sinistre est en cours, le montant de sinistre connu à la date d'inventaire est inférieur au montant total de celui-ci. Il y a donc censure de l'information « montant total de sinistre ».

E. Modèles utilisés

Pour évaluer la loi suivie par des données censurées, deux méthodes sont possibles :

- ✓ l'utilisation de modèles non paramétriques ;
- ✓ l'utilisation de modèles paramétriques.

Dans ce mémoire, l'attention se porte sur deux classes de modèles paramétriques.

L'estimation à l'aide des modèles non paramétriques de la fonction de survie permet d'obtenir une loi empirique du coût des sinistres. La méthode retenue pour cette estimation est l'estimateur de la fonction de survie de Kaplan-Meier. Des tests seront menés afin de comparer la fonction de survie issue des modèles paramétriques avec la fonction de survie empirique.

Dans le cadre de l'étude des modèles paramétriques, la sélection des différents modèles étudiés dans un premier temps correspond aux lois les plus utilisées dans la modélisation des montants des sinistres : il s'agit des lois de Pareto, Weibull. Cette étude consistera tout d'abord, en supposant que la loi suivie est une loi paramétrique, à en estimer les paramètres. Dans un second temps, une comparaison sera faite entre les modèles ainsi sélectionnés et la loi empirique. Enfin, en fonction du modèle le plus en adéquation avec les données et donc avec la fonction empirique, l'estimation de la provision sera réalisée.

En parallèle de cette approche sinistre par sinistre, le modèle de Chain-Ladder sera mis en place afin de pouvoir comparer la nouvelle méthode envisagée à une méthode plus classique et activement employée dans le monde de l'assurance.

Le chapitre 1 présentera de façon globale les différentes notions nécessaires à cette étude. Ces modèles seront détaillés et appliqués aux modèles estimant les montants des sinistres dans le chapitre 2 avec une méthode proposée pour choisir le meilleur modèle en fonction des données. Enfin dans un troisième chapitre, seront appliquées ces études aux données disponibles. Il s'agit ici d'une mise en œuvre pratique.

Le travail est effectué sur des données de garanties construction décennale.

1. Chapitre 1 - Préambule

1.1. *Présentation des données*

Les données d'étude sont constituées des montants réglés à chaque date de paiement pour chacun des sinistres pris individuellement. Selon la méthode employée pour calculer la provision, il est donc possible de regrouper les sinistres par année de survenance et par année de développement ou bien de considérer chaque sinistre séparément mais de cumuler les montants réglés afin de connaitre le coût total d'un sinistre.

Les données peuvent donc être mises en forme de deux façons :

- ✓ somme des sinistres ;
- ✓ somme des coûts pour chaque sinistre.

1.1.1. Triangulation des coûts de sinistres [1]

Les quantités analysées sont des paiements de sinistres survenus. Les sinistres sont rapportés à des périodes annuelles. L'année récurrente n se déroule du 01/01/n au 31/12/n. Le 31/12 étant la date d'inventaire ou date de fin d'exercice.

Les sinistres sont mis sous forme de triangles, c'est à dire qu'ils sont rattachés à des années d'origine qui peuvent être :

- ✓ l'année de survenance ;
- ✓ l'année de souscription ;
- ✓ l'année de déclaration.

En considérant une branche dont les sinistres se déroulent sur $n+1$ années, il est noté :

- ✓ i l'année d'origine où $i \in [0,...,n]$.
- ✓ k le délai de développement où $k \in [0,...,n-i]$.
- ✓ $Z_{i,k}$ la charge de sinistres correspondant à l'année d'origine i et au délai de développement k.

---

[1] « Mathématiques de l'assurance non-vie » de M.DENUIT & A.CHARPENTIER – Economica 2005

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

Au 31/12/n, les paiements de sinistres réglés antérieurement à cette date sont classiquement mis sous la forme d'un triangle de liquidation (run-off triangle) des montants non cumulés ou cumulés. La méthode de Chain-Ladder est mise en place sur les montants cumulés de sinistres.

S'il est supposé que chaque sinistre sera totalement réglé après $n+1$ années de développement, soit dans les n années civiles suivant la date d'origine, cette durée est alors une borne maximale.

Chaque $Z_{i,k}$ correspond au paiement de l'année de développement k pour des sinistres de l'année d'origine i. $Z_{i,k}$ correspond donc à un paiement effectué dans l'année civile $i+k$. Les $Z_{i,k}$ sont supposés observables pour $i+k \leq n$ et pas encore observés pour $i+k > n$. Les montants de sinistres sont donc regroupés dans des triangles de développement de la forme :

Année d'origine	Année de développement 0	1	…	k	…	n-i	…	n
0	$Z_{0,0}$	$Z_{0,1}$	…	$Z_{0,k}$	…	$Z_{0,n-i}$	…	$Z_{0,n}$
…	…	…	…	…	…	…	…	
i	$Z_{1,0}$	$Z_{1,1}$	…	$Z_{i,k}$	…	$Z_{i,n-1}$		
…	…	…	…	…	…			
n-k	$Z_{n-k,0}$	$Z_{n-k,1}$	…	$Z_{n-k,k}$				
…	…	…	…					
n-1	$Z_{n-1,0}$	$Z_{n-1,1}$						
n	$Z_{n,0}$							

Schéma 3 : Représentation d'un triangle de développement

Les montants $S_{i,k} = \sum_{l=0}^{k} Z_{i,l}$ correspondent quant à eux aux paiements cumulés des sinistres. Ces montants de coûts totaux sont eux aussi observables pour $i+k \leq n$. Un coût $S_{i,n}$ est le coût total de sinistres ayant pour année d'origine, l'année i et la différence $R_i = S_{i,n} - S_{i,n-i}$ constitue la réserve à constituer pour cette même année. Il s'agit donc de la provision relative à cette année.

La somme de chaque provision : $R = \sum_{i=0}^{n} R_i$ est le montant total de réserves dossier par dossier à provisionner à la date d'inventaire. Cette provision sert à faire face aux paiements futurs relatifs aux sinistres passés.

Ce traitement des données constitue donc la première mise en forme possible des données initiales.

1.1.2. Sinistres individuels

Un deuxième traitement des données consiste à conserver l'information « sinistre par sinistre ». Par conséquent et en terme de volume d'information, le fichier global à constituer ne se limite plus uniquement à des triangles agrégés mais à des sinistres estimés ligne à ligne.

Chaque montant de sinistre est classé en fonction de l'année de règlement. Pour un sinistre donné, les différents règlements effectués sont donc renseignés en fonction des dates de règlements. Les dates de règlements sont assimilées à des années comptables et ainsi regroupées. Ainsi, le tableau de données initiales nécessaire à la mise en place des modèles utilisant les sinistres ligne à ligne est de la forme :

Numéro de sinitre	Année comptable de règlement							
	1	2	…	…	k	…	K-1	K
1		$C_{1,2}$	…	…	$C_{1,k}$	…	$C_{1,K-1}$	$C_{1,K}$
…						…	$C_{.,K-1}$	$C_{.,K}$
i	$C_{i,1}$	$C_{i,2}$	…	…	$C_{i,k}$	…	$C_{i,K-1}$	$C_{i,K}$
…				…	$C_{.,k}$	…	$C_{.,K-1}$	$C_{i,K}$
…		$C_{.,2}$	…	…	$C_{.,k}$	…	$C_{.,K-1}$	$C_{i,K}$
…			…	…	$C_{.,k}$	…	$C_{.,K-1}$	$C_{i,K}$
n-1								$C_{n-1,K}$
n	$C_{n,1}$	$C_{n,2}$	…	…	$C_{n,k}$	…	$C_{n,K-1}$	$C_{n,K}$

Schéma 4 : Représentation des montants cumulés par sinistre

Les montants $C_{i,k}$ représentent les montants cumulés des coûts de sinistres pour chaque sinistre i en année comptable k. Certains sinistres sont entièrement réglés en année comptable $k \in [1;K]$. Néanmoins, les montants cumulés des sinistres sont ramenés jusqu'à l'année d'inventaire K qui correspond à l'année d'étude.

Par exemple, dans le tableau ci-dessus, le sinistre numéro 1 a commencé à être réglé en année 2. A compter de cette date, les différents règlements effectués sont cumulés jusqu'à obtenir en année K un montant total réglé de $C_{1,K}$ euros. Tel qu'est présenté le sinistre, l'information de clôture n'est pas connue.

Afin de pouvoir travailler sur ces données de cadence de sinistres, une information complémentaire doit être ajoutée : il s'agit de la date de clôture du sinistre. En réalité, dans la structure de l'étude, l'information sinistre « clos ou en cours » est suffisante. La date de clôture éventuelle est un élément supplémentaire qui permet de définir cette variable binaire.

Les données complètes nécessaires sont donc le tableau présenté ci-dessus auquel est ajouté l'information sinistre « clos ou en cours » qui traduit la notion de censure.

1.2. *Modèles de triangulation*

La première méthode employée dans le cadre du calcul des montants à provisionner est une méthode déterministe : la méthode de Chain- Ladder.

La méthode de Chain-Ladder utilisée pour estimer les réserves de sinistres sur la base de triangles de développement est d'autant plus performante que :

- ✓ le passé est régulier ;
- ✓ le présent et le futur sont structurellement peu différents du passé ;
- ✓ la branche d'assurance considérée est peu volatile ;
- ✓ les données sont nombreuses ;
- ✓ les données sont fiables.

Toutes ces exigences sur les données utilisées dans le calcul des provisions par la méthode de Chain-Ladder font de celle-ci une méthode intéressante car facilement utilisable et assez robuste mais dont toutes les hypothèses sous-jacentes ne sont pas validées.

En pratique, la méthode de Chain-Ladder estime les montants inconnus $Z_{i,k}$ pour $i+k>n$. Elle est basée sur le modèle suivant :

- ✓ Il y a des paramètres $\varphi_1,...,\varphi_n \in [1;+\infty[$ tels que $E[S_{i,k}]=E[S_{i,k-1}]*\varphi_k$ pour $i\in[0,...,n]$ et $k\in[0,...,n]$;
- ✓ Pour tout $k\in[0,...,n]$, on a $\sum_{j=0}^{n-k} S_{j,k-1}>0$.

Les paramètres $\varphi_1,...,\varphi_n$ sont inconnus.

La méthode de Chain-Ladder qui utilise les montants cumulés des sinistres est réalisée en deux étapes :

- ✓ Pour chaque année de développement $k\in[1,...,n]$, le facteur de Chain-Ladder $\hat{f}_k=\dfrac{\sum_{j=0}^{n-k} S_{j,k}}{\sum_{j=0}^{n-k-1} S_{j,k-1}}$ est défini comme estimateur de φ_k.
- ✓ Pour chaque année d'origine $i\in[0,...,n]$ et chaque année de développement $k\in[n-i,...,n]$, l'estimateur de Chain-Ladder $\hat{S}_{i,k}=S_{i,n-i}\prod_{l=n-i+1}^{k}\hat{f}_l$ est défini comme estimateur de l'espérance $E[S_{i,k}]$ de l'état $S_{i,k}$.

Donc, pour tout $k \in [n-i+1]$, $\hat{S}_{i,k} = \hat{S}_{i,k-1} * \hat{f}_k$. Ainsi le triangle de développement est complété par les estimations des montants de sinistres à venir. A partir de ces montants de sinistres cumulés estimés, les montants de sinistres annuels $\hat{Z}_{i,k}$ se déduisent par différence entre deux années.

Pour $i \in [0,...,n]$, la différence :

$$\hat{R}_i = \hat{S}_{i,n} - S_{i,n-i}$$

est appelée « réserve de Chain-Ladder » pour l'année d'origine i. Elle correspond à la différence entre le montant considéré comme total du sinistre et le montant réglé à la date d'inventaire.

La somme de ces réserves constitue la « réserve globale de Chain-Ladder » pour l'ensemble des années de développement :

$$\hat{R} = \sum_{i=0}^{n} \hat{R}_i \ .$$

Cette réserve constitue le montant total des provisions. Elle est estimée à partir du développement des montants de sinistres cumulés observés à la date d'inventaire et ce à l'aide des facteurs de développement.

1.3. *Modèles de données censurées – paramétriques*

1.3.1. Particularités des Données

Les données de montants de sinistres étudiées dans ce mémoire ont des caractéristiques communes avec les données étudiées dans le cadre des modèles de durée :

- ✓ les données sont générées par des variables aléatoires positives : les montants de sinistres réglés sont en effet positifs et aléatoires dans le sens où ils ne correspondent pas à des montants de remboursements forfaitaires ;
- ✓ pour certains individus, le résultat total n'est connu que partiellement : les sinistres dont la totalité du coût n'a pas été remboursée à la date d'étude constituent des données censurées. La présence de censure dans les données est caractéristique des modèles de durée qui s'interprètent ici en modèle de coût.

Ces particularités conduisent à définir les données à l'aide de fonctions de survie ou fonctions de hasard et non plus au travers de fonctions de répartition.

1.3.2. Notations et définitions

Soient :

- ✓ i le $ième$ sinistre ;
- ✓ X_i le coût total unitaire du sinistre i ;
- ✓ C_i la somme des règlements déjà effectués pour ce sinistre à la date d'étude;
- ✓ D_i l'indicatrice de l'état du sinistre à la date d'étude ;
- ✓ θ le paramètre de la loi de X .

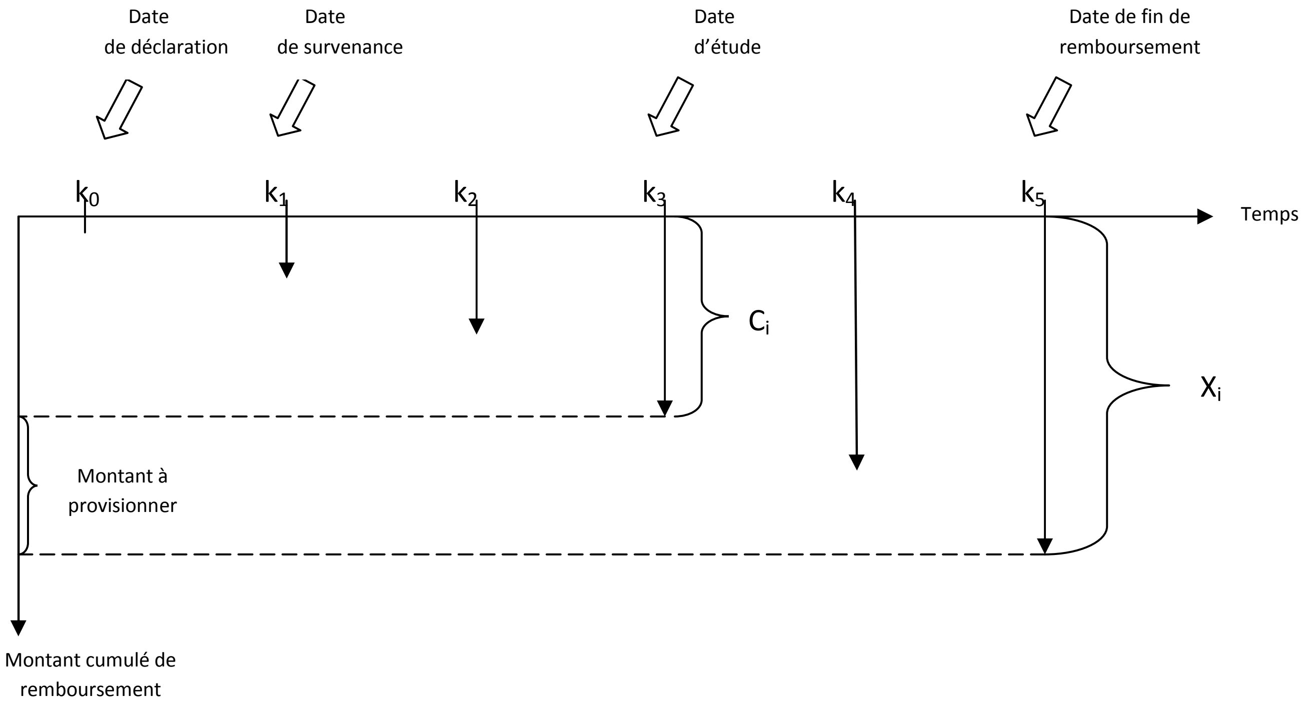


Schéma 4 : Schématisation des différentes données de l'étude

Ce schéma indique les différentes variables relatives à un sinistre i. Le sinistre ainsi représenté est intervenu en date k_0 . Les différents flux de remboursement ont lieu en date k_1, k_2, k_3, k_4 et k_5 . Or, la date d'inventaire est la date k_3 . Les flux postérieurs à cette date ne sont donc pas connus au moment de l'étude. La charge totale du sinistre X_i n'est pas encore connue et seul le montant C_i est déjà remboursé.

Deux cas peuvent alors être rencontrés : soit le montant total remboursé correspond au montant total du sinistre et donc les remboursements de sinistres s'arrêtent et aucune provision n'est alors à constituer ; soit la charge de sinistre n'est pas totalement remboursée et il faut donc constituer la provision qui correspond à la différence entre le montant déjà réglé et le montant total de sinistre.

Les variables du problème sont étudiées en tenant compte de la particularité des données et en faisant donc appel à des fonctions représentatives des données censurées [2]:

X[3] est une variable aléatoire à valeurs dans $[0;+\infty[$ de fonction de répartition $F_\theta(x) = P(X \leq x)$, de densité f_θ ; on désigne par S_θ la fonction de survie du modèle et h_θ la fonction de hasard.

Lorsque la densité de la variable X existe, elle se note :

$$f_\theta(x) = \frac{d}{dx} F_\theta(x) = \lim_{dx \to 0} \frac{P_\theta(x \leq X \leq x + dx)}{dx}$$

La fonction de survie est par définition le complément à 1 de la fonction de répartition. Elle représente la probabilité que le coût a de dépasser un certain seuil :

$$S_\theta(x) = 1 - F_\theta(x) = P_\theta(X > x).$$

La fonction de survie conditionnelle est par définition la survie d'un élément après un instant *x* sachant qu'il a déjà survécu jusqu'en *u*. Cela s'écrit donc :

$$S_{\theta,u}(x) = P_\theta(X > u + x \mid X > u) = \frac{P_\theta(X > u + x)}{P_\theta(X > u)} = \frac{S_\theta(u + x)}{S_\theta(u)}$$

La fonction de hasard aussi appelée taux de risque ou risque instantané est :

$$h_\theta(x) = \frac{f_\theta(x)}{S_\theta(x)} = -\frac{S_\theta'(x)}{S_\theta(x)} = -\frac{d}{dx} \ln S_\theta(x)$$

C représente la somme des règlements déjà effectués pendant la période couverte par les données. Cette période s'écoule depuis de la date de survenance des sinistres à la date d'étude qui est en pratique la date d'inventaire.

[2] « Modèles de Durée » de F. PLANCHET & P. THEROND – Economica 2006

[3] Les paramètres de la loi de X sont regroupés dans la notation θ

D renseigne sur l'état du sinistre à la date d'inventaire. Si celui-ci est en cours, la variable vaut 0 et donc $X > C$. Lorsque le sinistre est totalement réglé, l'indicatrice vaut 1 ; dans ce cas, il n'y a pas de montant à provisionner.

1.3.3. Données censurées

Soit X le coût unitaire d'un sinistre, $X \geq C$. Le modèle présente donc un cas de censure à droite par une variable aléatoire ; plus précisément, soit un échantillon de montants de sinistres $(X_1,..,X_n)$ et un second échantillon indépendant composé des règlements cumulés effectués à la date d'inventaire $(C_1,..,C_n)$. Cet échantillon présente les caractéristiques d'une censure de type III, car au lieu d'observer directement $(X_1,..,X_n)$ les observations portent sur $(T_1,D_1),..,(T_n,D_n)$ avec :

$$T_i = X_i \wedge C_i$$

$$\text{et } D_i = \begin{cases} 1 \; si \; X_i \leq C_i \\ 0 \; si \; X_i > C_i \end{cases}$$

La vraisemblance de l'échantillon $(T_1,D_1),..,(T_n,D_n)$ s'écrit, avec les notations classiques :

$$L(\theta) = \prod_{i=1}^{n} \left[f_X(T_i,\theta) S_C(T_i,\theta) \right]^{D_i} \left[f_C(T_i,\theta) S_X(T_i,\theta) \right]^{1-D_i}$$

Dans le cas des données ici étudiées, les différents règlements $(C_1,...,C_n)$ sont des données connues.

En retenant alors l'hypothèse que la censure est non informative, c'est à dire que la loi de censure est indépendante du paramètre θ, la vraisemblance se met sous la forme :

$$L(\theta) = const \prod_{i=1}^{n} f_\theta(T_i)^{D_i} \, S_\theta(C_i)^{1-D_i}$$

Cette hypothèse est valide lorsqu'il s'agit de censure planifiée par opposition aux censures dues à un autre évènement que celui recherché. Dans le cadre de cette étude, la date d'inventaire est planifiée et la censure provient seulement de l'arrêt de l'observation à cette date choisie. La censure est donc non informative.

Le terme *const* regroupe les informations en provenance de la loi de la censure, qui ne dépendent pas du paramètre. L'expression de la vraisemblance précédente traduit le fait que lorsque la sortie avant censure est observée, c'est le terme de densité qui intervient dans la vraisemblance et dans le cas contraire, c'est le terme discret avec comme valeur la fonction de survie au moment de la censure qui est pris en compte dans la vraisemblance.

Par définition, pour une information censurée, $T_i = C_i$. La vraisemblance peut alors s'écrire :

$$L(\theta) = const \prod_{i=1}^{n} f_\theta(T_i)^{D_i}\, S_\theta(T_i)^{1-D_i}$$

En écrivant ensuite la densité comme le produit de la fonction de hasard et de la fonction de survie $f_\theta(t) = h_\theta(t) S_\theta(t)$, la vraisemblance s'exprime alors ainsi :

$$L(\theta) = const \prod_{i=1}^{n} f_\theta(T_i)^{D_i}\, S_\theta(T_i)^{1-D_i}$$

Cette dernière expression conduit en passant au logarithme à :

$$\mathbf{ln}\, L(\theta) = \sum_{i=1}^{n} \mathbf{ln}\left(S_\theta(T_i)\right) + \sum_{i=1}^{n} D_i\, \mathbf{ln}\left(h_\theta(T_i)\right)$$

Il est ainsi possible d'estimer le paramètre θ par maximum de vraisemblance dès que les fonctions de survie et de hasard correspondant aux données censurées sont exprimées.

1.3.4. Les Modèles Statistiques

Les modèles d'étude mis en place dans le cadre de ce mémoire afin de connaitre la loi suivie par le coût total des sinistres sont les modèles non paramétriques et les modèles paramétriques.

1.3.4.1. Les modèles non paramétriques

L'utilisation de modèles non paramétriques permet de ne pas faire d'hypothèses *a priori* sur la forme de la loi de survie. Cette fonction est estimée directement en vue des données. Dans le contexte de censure et de modèle de coût, l'estimateur principal de la fonction de survie est l'estimateur de Kaplan-Meier.

L'estimateur de Kaplan-Meier s'appuie sur la remarque suivante : la probabilité que le montant de sinistre dépasse un montant $t > s$ peut s'écrire :

$$S(t) = P(T > t \mid T > s) P(T > s) = P(T > t \mid T > s) S(s)$$

où T est la variable aléatoire représentant le coût total du sinistre. En renouvelant l'opération, la fonction de survie s'exprime comme un produit de termes en $P(T > t \mid T > s)$. En choisissant comme montants de conditionnement les montants correspondants à un montant total ou censuré selon que le sinistre est clos ou en cours au moment de l'inventaire ; estimer la fonction de survie revient à estimer la probabilité que le montant dépasse le montant i valant $T_{(i)}$.

Toujours en lien avec les modèles de durée, on exprime $p_i = 1 - q_i$ avec pour estimateur naturel de q_i :

$$\hat{q}_i = \frac{d_i}{r_i}$$

- ✓ r_i comptabilise le nombre de sinistres dont le montant est supérieur au montant T_i et dont le 1er règlement est inférieur au montant T_i ;
- ✓ d_i vaut 1 lorsque le sinistre est clos et 0 lorsque le sinistre est en cours et qu'il y a censure des données.

L'estimateur de Kaplan-Meier s'écrit donc finalement :

$$\hat{S}(t) = \prod_{T_{(i)} \le t} \left(1 - \frac{d_i}{r_i}\right)$$

La fonction de survie ainsi estimée caractérise entièrement les données du modèle. Cette fonction est graphiquement représentée comme une fonction en échelle. La fonction n'est donc pas continue et pour tout t, la valeur de la fonction de survie est une approximation.

C'est afin d'obtenir une fonction de survie continue[4] que les modèles paramétriques interviennent dans l'étude des modèles du durée et dans le cadre de notre étude dans l'estimation des montants de sinistres.

1.3.4.2. Les modèles paramétriques

Coût d'un sinistre

L'utilisation des modèles paramétriques afin d'évaluer la loi suivie par les données de coût de sinistre consiste dans un premier temps à supposer un type de loi. A partir de cette loi supposée sont déduites les fonctions de survie et de hasard.

En appliquant l'expression de la log-vraisemblance :

$$\ln L(\theta) = \sum_{i=1}^{n} \ln\left(S_\theta(T_i)\right) + \sum_{i=1}^{n} D_i \ln\left(h_\theta(T_i)\right),$$

les paramètres de la loi sont estimés. Ces paramètres fournissent donc une information complète sur la loi.

Ainsi, la loi suivie par le coût des sinistres est estimée de façon précise. La variable X est désormais connue grâce à l'estimation de l'échantillon. Il faut alors évaluer la loi suivie par la provision individuelle relative à chaque sinistre.

Provision individuelle

Le calcul de la provision nécessite l'introduction d'une nouvelle variable dont on cherche à connaitre la loi. La provision « best estimate » pour un sinistre est en effet par définition égale à l'espérance de la charge résiduelle du sinistre, charge définie par :

$$P = X - c \mid X \geq c.$$

---

[4] Une méthode à noyaux fournirait aussi un lissage de la fonction.

La loi de *X* étant connue, celle de $X - c$ l'est aussi. On cherche désormais à connaitre la loi conditionnelle de $X - c$ sachant que le coût total du sinistre égal ou dépasse le montant déjà réglé. L'espérance de cette variable *P* représente la provision à constituer pour un sinistre.

Cette partie de l'étude se fait de façon individuelle. Chaque sinistre engendre une provision qui lui est propre. Il faut étudier la loi conditionnelle au dépassement d'un certain seuil par la variable d'intérêt *X*. La fonction de survie conditionnelle est alors introduite. Soit *c*, le seuil dépassé par la variable *X*,

$$\begin{aligned} S_{\theta,c}(x) &= P(X > c + x \mid X > c) \\ &= \frac{P(X > c + x)}{P(X > c)} \\ &= \frac{S_\theta(c + x)}{S_\theta(c)} \end{aligned}$$

Or, la fonction de survie de la variable étudiée, c'est-à-dire la provision pour sinistre $P = X - c \mid X > c$ notée S_α est égale à :

$$\begin{aligned} S_\alpha(p) &= P\left((X - c \mid X > c) > p\right) \\ &= P(X - c > p \mid X > c) \\ &= \frac{P(X > p + c)}{P(X > c)} \\ &= S_{\theta,c}(p) \end{aligned}$$

Donc la fonction de survie de la provision individuelle à constituer est égale à la fonction de survie conditionnelle du coût du sinistre avec pour seuil dépassé la somme des règlements déjà effectués pour ce sinistre.

Provision totale

La loi du coût futur d'un sinistre est désormais connue, il faut alors s'intéresser au comportement de la provision totale appelée provision pour sinistres en cours. Pour connaître le montant de la provision à calculer pour l'ensemble du portefeuille, il convient dans un premier temps évaluer la loi suivie par cette somme et en prendre ensuite l'espérance.

La provision pour sinistres en cours est égale à l'espérance du montant résiduel total à payer, qui est une somme de variables indépendantes, mais non identiquement distribuées :

$$\Lambda = \sum_{i \in I_e} P_i ,$$

avec I_e l'ensemble des sinistres en cours (pour lesquels $D_i = 0$).

En fonction de la loi S_θ différentes méthodes peuvent être envisagées pour déterminer la distribution de Λ. Soit :

- ✓ le calcul explicite est possible (ce peut être le cas si les provisionnements individuels suivent des lois normales) ;
- ✓ on se trouve dans le domaine d'application du TCL, la condition de Lyapounov ou de Lindberg étant supposée satisfaite.

Calcul explicite

Lorsque ces quantités existent, le calcul de l'espérance et de la variance de *P* est possible à partir de la fonction de survie associée :

$$E_{\theta,c}(P) = \int_0^\infty S_{\theta,c}(p)\,dp$$

$$V_{\theta,c}(P) = 2\int_0^\infty pS_{\theta,c}(p)\,dp - E_{\theta,c}(P)^2$$

de sorte que, lorsqu'elle est possible, l'approximation normale de la loi de Λ est aisée à déterminer. La provision elle-même est alors estimée directement :

$$\mathbf{E}(\Lambda) = \sum_{i \in I_e} \mathbf{E}(P_i)$$

sa variance, lorsqu'elle existe, étant également de la forme

$$\mathbf{V}(\Lambda) = \sum_{i \in I_e} \mathbf{V}(P_i)$$

En particulier, lorsque l'approximation normale est valide, les paramètres de la loi limite se déduisent directement, au moins numériquement.

Domaine d'application du TCL

Le théorème central limite affirme que toute somme de variables aléatoires indépendantes et identiquement distribuées tend vers une variable aléatoire gaussienne. Les variables ici cumulées correspondant aux provisions individuelles sont bien indépendantes mais non identiquement distribuées. L'hypothèse d'identité des distributions peut ne pas être nécessaire si certaines conditions vérifiant qu'aucune variable n'exerce une influence significativement plus importante que les autres sont justifiées.

Les conditions de Lindeberg et Lyapounov permettent alors la généralisation du théorème qui ne nécessite plus que les variables soient de même lois. Dans le cadre de cette étude, les provisions constituées individuellement ne sont pas identiquement distribuées. Il est donc nécessaire que ces conditions soient remplies afin d'appliquer le théorème central limite.

1.3.4.3. Condition de Lyapounov

Le but de cette partie est de connaitre la loi de la somme des provisions individuelles. Soit $\left(P_1,..,P_n\right)$ la séquence de variables correspondant aux provisions individuelles. Soit μ_i, l'espérance finie de P_i et σ_i son écart-type fini, il est possible de définir alors :

$$s_n^{\ 2} = \sum_{i=1}^{n} \sigma_i^{\ 2}$$

De même, en supposant que les moments d'ordre 3 soient finis pour tout n, il est noté :

$$r_n^3 = \sum_{i=0}^{n} E\left(| X_i - \mu_i |^3\right)$$

La condition de Lyapounov devant être vérifiée est :

$$\lim_{n\to\infty} \frac{r_n}{s_n} = 0$$

Lorsque cette condition est vérifiée, l'application du théorème central limite est alors possible. Soit la somme des variables aléatoires notée : $S_n = X_1 + X_2 + ... + X_n$ de variance s_n et d'espérance mathématique : $m_n = \sum_{i=1}^{n} \mu_i$. Alors,

$$\frac{S_n - m_n}{s_n} \underset{n\to\infty}{\to} N(0,1)$$

1.3.4.4. Condition de Lindeberg

Avec les mêmes définitions et les mêmes notations que précédemment, il est possible de remplacer la condition de Lyapounov par la condition de Lindeberg qui est plus faible.

Cette condition s'écrit de la manière suivante : pour tout $\varepsilon > 0$,

$$\lim_{n\to\infty} \sum_{i=1}^{n} E\left(\frac{(X_i - \mu_i)^2}{s_n^{\ 2}} \middle| \left| X_i - \mu_i \right| > \varepsilon\, s_n \right)$$

Lorsque cette condition est vérifiée, le théorème central limite peut être appliqué et la convergence suivante est établie :

$$\frac{S_n - m_n}{s_n} \underset{n\to\infty}{\to} N(0,1)$$

1.3.4.5. Conséquences

Lorsque l'une des conditions ainsi définies est vérifiée, le théorème central limite est appliqué et ainsi, la loi de la somme des provisions est connue.

2. Chapitre 2 - Modèles de censure

Ce chapitre vise à présenter les lois qui seront testées en tant que lois d'adéquation aux variables montants des sinistres. La sélection des différents modèles étudiés se fait parmi les distributions continues les plus pertinentes pour la modélisation de montants de sinistres. Il s'agit classiquement des lois de Pareto et de Weibull.

Lorsque la loi du coût total des sinistres est connue de façon empirique grâce à l'utilisation du modèle non paramétrique de Kaplan – Meier, un test d'ajustement est réalisé afin de valider ou non la correspondance avec les lois citées.

2.1. *Présentation des lois utilisées*

2.1.1. Pourquoi ces lois ?

La présente étude consiste à trouver la loi suivie par la somme des provisions individuelles en supposant que la distribution des coûts individuels des sinistres est connue et qu'elle suit une loi de Pareto ou bien de Weibull.

Lorsque le type de loi est connu, le premier objectif de l'étude consiste donc à estimer les différents paramètres. Ainsi, la loi suivie par le montant total de sinistres est déterminée pour un groupe de risques homogènes. Il faut ensuite évaluer la loi suivie par la provision et ensuite la loi de la somme de ces provisions qui constitue alors le provisionnement total du portefeuille de risque.

2.1.2. Détail de la loi de Pareto

Dans un premier temps, il est supposé que la loi sous-jacente de la variable représentant le coût unitaire d'un sinistre est une Pareto de paramètres (m,α).

Ce modèle est intéressant du fait de l'invariance de la loi de Pareto par conditionnement au dépassement d'un seuil. Par ailleurs, cette loi apparaît naturellement si l'étude porte sur les « grands sinistres », dont le montant a par exemple dépassé un montant prédéfini.

L'invariance de la loi de Pareto conditionnellement au dépassement d'un seuil permettra de connaître la loi de la provision individuelle dès lors que la loi de X est estimée ainsi que ses paramètres.

2.1.2.1. Présentation générale

Soit X, une variable aléatoire supposée suivre une loi de Pareto de paramètres (m,α) et admettant pour tout $x \geq m$ la fonction de répartition suivante:

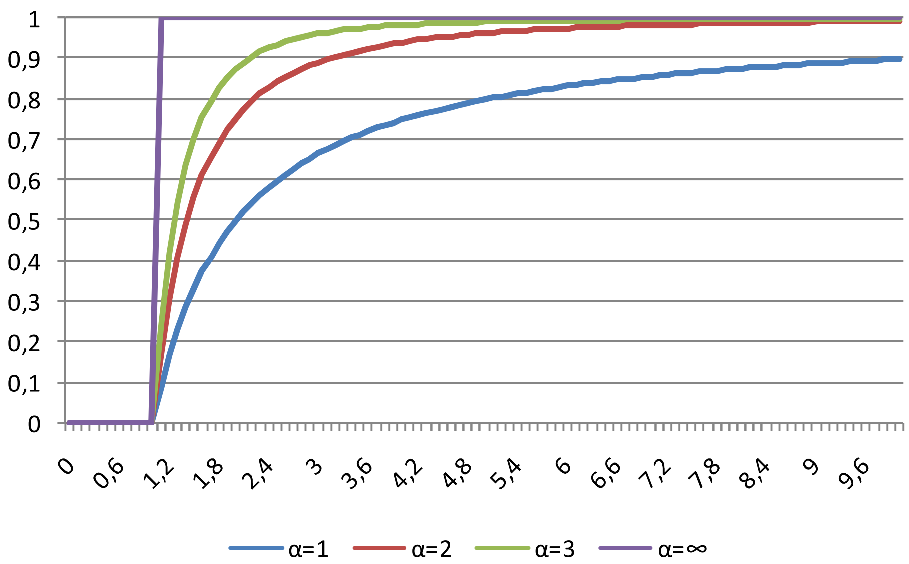
$$F_{m,\alpha}(x) = 1 - \left(\frac{x}{m}\right)^{-\alpha} \text{ avec}$$

- ✓ m est la valeur positive minimale que peut prendre X, c'est un paramètre de location qui détermine la position de l'origine ;
- ✓ α est un réel positif appelé indice de Pareto, c'est un paramètre qui régit la forme de la distribution.

La densité de probabilité de X est de la forme :

$$f_{m,\alpha}(x) = \alpha \frac{m^{\alpha}}{x^{\alpha+1}}.$$

La représentation graphique de la fonction de répartition d'une loi de Pareto de paramètres $m=1$ et α variable valant $1,2,3$ ou ∞ est la suivante :



Graphique 1 Différentes fonctions de répartition de Pareto

La fonction de répartition varie selon le paramètre α : plus celui-ci est grand, plus la fonction de répartition croit rapidement vers sa limite qui vaut 1.

L'espérance d'une variable aléatoire suivant une loi de Pareto est $E(X)=\frac{\alpha m}{\alpha-1}$ et sa variance vaut : $Var(X)=\left(\frac{m}{\alpha-1}\right)^2\frac{\alpha}{\alpha-2}$. Il est à noter que lorsque que $\alpha \le 1$, l'espérance est infinie et de même lorsque $\alpha \le 2$, la variance est infinie.

2.1.2.2. Application à la loi du coût du sinistre

Soit *X*, la variable aléatoire représentant le coût total unitaire d'un sinistre. Dans le cadre des données étudiées, les fonctions représentatives de la variable aléatoire sont les fonctions de survie et de hasard. La variable *X* admet pour fonction de survie :

$$S_{m,\alpha}(x)=\left(\frac{x}{m}\right)^{-\alpha} \text{ pour } x \ge m.$$

Sa fonction de hasard s'écrit pour $x \ge m$:

$$h_{m,\alpha}(x)=-\frac{d}{dx}\mathbf{ln}\, S_{m,\alpha}(x)=\frac{\alpha}{x}.$$

Ces deux fonctions permettent à l'aide de l'estimation du maximum de vraisemblance d'estimer les paramètres de la fonction dans le cadre de données censurées et ainsi de définir la loi de façon précise.

2.1.2.3. Estimation des paramètres

Pour estimer les paramètres, l'expression générale de la fonction de log-vraisemblance utilisée est la suivante :

$$\mathbf{ln}\, L(\theta)=\sum_{i=1}^{n}\mathbf{ln}\left(S_\theta(t_i)\right)+\sum_{i=1}^{n}d_i\,\mathbf{ln}\left(h_\theta(t_i)\right),$$

il en résulte avec application à la loi de Pareto (et en notant $d=\sum_{i=1}^{n}d_i$ le nombre de sinistres clos dont l'information n'est pas censurée) :

$$\ln L(m,\alpha) = -\alpha\sum_{i=1}^{n}\ln(t_i) + \alpha\times n\times\ln(m) + d\times\ln(\alpha) - \sum_{i=1}^{n} d_i \ln(t_i).$$

Cette expression prend bien en compte la notion de censure. Les données servant aux calculs sont en effet les couples de variables (t_i, d_i) qui correspondent aux montants de sinistres à la date d'inventaire et à la présence ou non de censure pour le sinistre i.

L'équation aux dérivées partielles par rapport au paramètre α s'écrit donc :

$$\frac{\partial}{\partial\alpha}\ln L(\alpha,m) = -\sum_{i=1}^{n}\ln t_i + n\ln m + \frac{d}{\alpha}$$

Il en résulte donc directement l'expression suivante des estimateurs :

$$\hat{m} = \min\{t_i; 1 \le i \le n\}$$

$$\hat{\alpha} = \frac{d}{\sum_{i=1}^{n}\ln\left(\frac{t_i}{\hat{m}}\right)}$$

En effet, comme $\hat{m} \le \min\{t_i; 1 \le i \le n\}$ et que la log-vraisemblance $\ln L(m,\alpha)$ est croissante en *m*, nécessairement $\hat{m} = \min\{t_i; 1 \le i \le n\}$. L'expression de $\hat{\alpha}$ découle directement de $\frac{\partial}{\partial\alpha}\ln L(\hat{m},\hat{\alpha}) = 0$.

Ainsi, la loi suivie par X est dans le cadre de ce modèle une distribution de Pareto de paramètres $(\hat{m},\hat{\alpha})$ avec des paramètres explicites.

2.1.2.4. Calcul de la provision individuelle

Le terme de provision individuelle correspond à la provision relative à un seul sinistre. La loi suivie par la variable X représentant le coût de sinistre individuel est commune à tous les sinistres. En revanche, la loi suivie par la provision à constituer est propre à chaque sinistre dans le sens où le montant déjà réglé pour un sinistre donné varie d'un sinistre à l'autre.

La loi de X définit à l'aide du modèle de Pareto étant complètement identifiée et estimée, il faut désormais évaluer la loi de la provision à mettre en place.

La loi de Pareto $P(m,\alpha)$ est caractérisée par deux propriétés importantes qui sont utiles à l'étude du provisionnement individuel du sinistre :

Propriété 1 : Si la variable X suit une loi de Pareto $P(m,\alpha)$; alors la loi conditionnelle au fait de dépasser un seuil $u \geq m$ notée $X \mid X > u$ est encore de Pareto, de paramètres (u,α).

Propriété 2 : Si la variable X suit une loi de Pareto $P(m,\alpha)$; alors $X - m$ suit une loi de Pareto de 2ème espèce $P_2(m,\alpha)$.

La loi de X est une Pareto dont les paramètres sont estimés par rapport aux données est telle que :

$$X \sim P(\hat{m},\hat{\alpha}) \text{ où } \hat{m} = \mathbf{min}\{t_i ; 1 \leq i \leq n\}.$$

Or, les provisions correspondent aux sinistres qui sont encore en cours (dans le domaine définit par I_e) et dont le coût total est donc supérieur au montant déjà réglé. Donc la variable D_i définit par

$$D_i = \begin{cases} 1 \; si \; X_i \leq C_i \\ 0 \; si \; X_i > C_i \end{cases}$$ vaut 0. Ainsi pour chaque i, $X_i > C_i$.

Par conséquent, pour tous les sinistres ici considérés, la variable T_i correspondant au minimum entre X_i et C_i est égale à la valeur de C_i. Le paramètre estimé de m est égal à :

$$\hat{m} = \mathbf{min}\{t_i ; 1 \leq i \leq n\}$$

D'après la propriété 1 d'une loi de Pareto, comme $t_i \leq c_i$, $X_i \mid X_i > c_i$ suit une loi de Pareto $P(c_i,\hat{\alpha})$ où c_i est propre à chaque sinistre.

D'après la propriété 2 d'une loi de Pareto, $P_i = X_i - c_i \mid X_i > c_i$ suit une loi de Pareto de 2ème espèce de paramètres identiques à ceux de la loi suivie par $X_i \mid X_i > c_i$. Donc :

$$P_i = X_i - c_i \mid X_i > c_i \sim P_2(c_i,\alpha)$$

La loi de Pareto de 2ème espèce a pour fonction de répartition $F_{c_i,\alpha}(x) = 1 - \left(\frac{c_i}{c_i + x}\right)^{\alpha}$.

Ses moments sont de la forme :

$$m_k = \frac{k! c_i^k}{(\alpha-1)\ldots(\alpha-k)}$$ pour $\alpha > k$ ou ($m_k = +\infty$ sinon).

Ainsi, pour un sinistre de distribution de Pareto dont le montant payé jusqu'à présent est c_i, le montant de provision à constituer est connu. Il s'agit en effet de l'espérance de P_i soit pour $\alpha > 1$,

$$E(P_i) = \frac{c_i}{\alpha - 1}$$

La variance de P_i pour $\alpha > 2$ est égale à

$$\mathbf{V}(P_i) = \frac{\alpha}{(\alpha-1)^2(\alpha-2)} \times c_i^2.$$

2.1.2.5. Calcul de la provision totale

Le provisionnement à effectuer pour un seul sinistre est alors connu. Il s'agit du cumul des espérances. L'espérance de la somme est égale à la somme des espérances. Sachant $\Lambda = \sum_{i \in I_e} P_i$, le montant à provisionner est égal à la somme des provisions calculées de façon individuelle ; soit pour $\alpha > 1$:

$$\mathbf{E}(\Lambda) = \frac{1}{\alpha - 1} \sum_{i \in I_e} c_i \, .$$

De même, par indépendance entre les provisions de sinistres, la variance de la somme des provisions est égale à la somme des variances ; soit pour $\alpha > 2$:

$$\mathbf{V}(\Lambda) = \frac{\alpha}{(\alpha-1)^2(\alpha-2)} \sum_{i \in I_e} c_i^2 \, .$$

Ainsi, l'utilisation du modèle de Pareto permet d'estimer le montant de la provision à constituer pour l'ensemble du portefeuille qui est constitué à la date d'évaluation d'une partie des montants de sinistres déjà réglés pour les sinistres encore en cours. Cette proportion est déterminée uniquement à l'aide du paramètre α de la loi de Pareto ; ce paramètre régissant la forme de la distribution et dépendant lui-même du nombre de données censurées et des montants de sinistres réglés à la date d'étude.

2.1.3. Détail de la loi de Weibull

Dans cette partie, il est désormais supposé que la loi suivie par la variable X représentant le coût unitaire d'un sinistre est une loi de Weibull de paramètres (α, l).

La loi de Weibull est une alternative souvent intéressante aux lois de coûts classiques de type log-normale ou gamma. Elle est souvent utilisée dans le cadre des modèles de durée.

2.1.3.1. Présentation générale

Soit X une variable aléatoire supposée suivre une loi de Weibull de paramètres (α, l). La variable admet la fonction de répartition suivante :

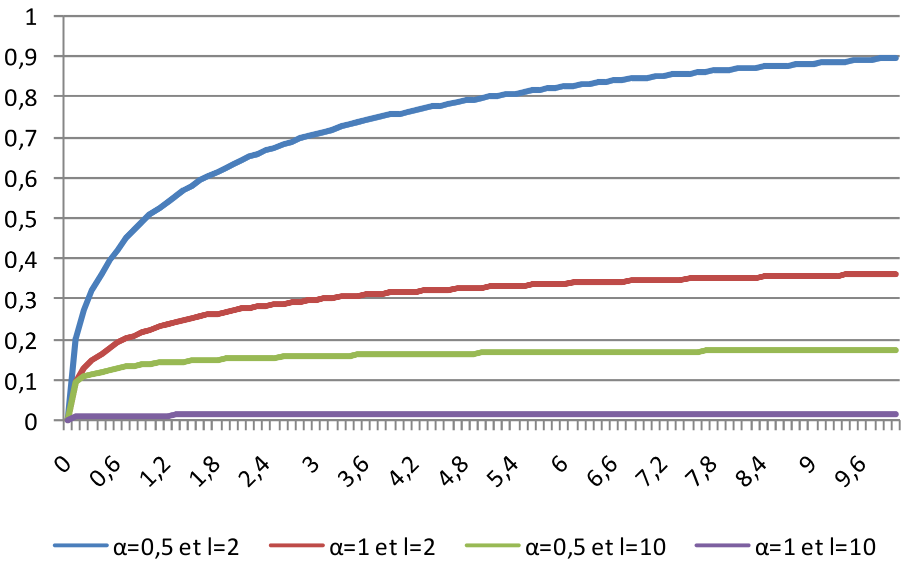
$$F_{\alpha,l}(x) = 1 - \mathbf{exp}\left\{-\left(\frac{x}{l}\right)^{\alpha}\right\} \text{ avec}$$

- ✓ α un paramètre de forme positif, dans le cas particulier où $\alpha = 1$, la loi de Weibull est en fait une loi exponentielle ;
- ✓ l un paramètre d'échelle positif.

La densité de probabilité de X est de la forme :

$$f_{\alpha,l}(x) = \frac{\alpha}{l^{\alpha}} x^{\alpha-1} \mathbf{exp}\left\{-\left(\frac{x}{l}\right)^{\alpha}\right\}.$$

La représentation graphique de la fonction de répartition d'une loi de Weibull de différents paramètres se présente comme suit :



Graphique 2 : Différentes fonctions de répartition de Weibull

La fonction de répartition croit plus rapidement avec des paramètres de faibles valeurs. Il est à remarquer graphiquement que le paramètre α régit la courbure de la courbe en son origine alors que pour α donné et deux paramètres l distincts un paramètre l plus petit engendre une croissance vers 1 plus forte.

L'espérance d'une variable aléatoire suivant une loi de Weibull est

$$E(X) = l\Gamma\left(\frac{1}{\alpha}+1\right).$$

2.1.3.2. Application à la loi du coût du sinistre

Soit *X*, la variable aléatoire représentant le coût total unitaire d'un sinistre. Dans le cadre des données étudiées, les fonctions représentatives de la variable aléatoire sont les fonctions de survie et de hasard. La variable *X* admet pour fonction de survie :

$$S_{\alpha,l}(x) = \mathbf{exp}\left\{-\left(\frac{x}{l}\right)^{\alpha}\right\}$$

Sa fonction de hasard s'écrit pour :

$$h_{m,\alpha}(x) = -\frac{d}{dx}\mathbf{ln}\,S_{m,\alpha}(x) = \alpha\frac{x^{\alpha-1}}{l^{\alpha}}$$

Ces deux fonctions permettent à l'aide de l'estimation du maximum de vraisemblance d'estimer les paramètres de la fonction dans le cadre de données censurées et de définir la loi de façon précise.

2.1.3.3. Estimation des paramètres

La vraisemblance dans le cadre de données censurées appliquée aux variables « montants de sinistres » est de la forme :

$$L(\alpha, l) \propto \prod_{i=1}^{n} f(t_i)^{d_i} S(t_i)^{1-d_i}$$

L'expression de cette vraisemblance pour une variable suivant un modèle de Weibull est (en notant $d = \sum_{i=1}^{n} d_i$ le nombre de sorties observées non censurées) :

$$L(\alpha, l) \propto \left(\frac{\alpha}{l^{\alpha}}\right)^{d} \prod_{i=1}^{n} t_i^{(\alpha-1)d_i} \, \mathbf{exp}\left\{-d_i \left(\frac{t_i}{l}\right)^{\alpha}\right\} \mathbf{exp}\left\{-(1-d_i)\left(\frac{t_i}{l}\right)^{\alpha}\right\}$$

$$\Leftrightarrow L(\alpha, l) \propto \left(\frac{\alpha}{l^{\alpha}}\right)^{d} \mathbf{exp}\left\{-l^{-\alpha} \sum_{i=1}^{n} t_i^{\alpha}\right\} \mathbf{exp}\left\{(\alpha-1) \sum_{i=1}^{n} d_i \, \mathbf{ln}\, t_i\right\}$$

De cette dernière expression, la fonction de log-vraisemblance appliquée au modèle de Weibull avec présence de données censurées s'écrit :

$$\mathbf{ln}\, L(\alpha, l) = \mathbf{ln}\, k + d(\mathbf{ln}\, \alpha - \alpha \, \mathbf{ln}\, l) - l^{-\alpha} \sum_{i=1}^{n} t_i^{\alpha} + (\alpha - 1) \sum_{i=1}^{n} d_i \, \mathbf{ln}\, t_i$$

Les équations aux dérivés partielles par rapport aux deux paramètres l et α s'écrivent donc :

$$\begin{cases} \dfrac{\partial}{\partial l} \mathbf{ln}\, L(\alpha, l) = -\dfrac{d}{l} + \alpha l^{-\alpha-1} \displaystyle\sum_{i=1}^{n} t_i^{\alpha} \\ \dfrac{\partial}{\partial \alpha} \mathbf{ln}\, L(\alpha, l) = d\left(\dfrac{1}{\alpha} - \mathbf{ln}\, l\right) + l^{-\tau}\left(\mathbf{ln}\, l \displaystyle\sum_{i=1}^{n} t_i^{\alpha} - \sum_{i=1}^{n} t_i^{\alpha} \, \mathbf{ln}\, t_i\right) + \sum_{i=1}^{n} d_i \, \mathbf{ln}\, t_i \end{cases}$$

Pour trouver les expressions des paramètres, les dérivées partielles sont supposées égales à 0 et la résolution des équations donnent les résultants suivants :

$$\begin{cases} l = \left(\dfrac{1}{d} \displaystyle\sum_{i=1}^{n} t_i^{\alpha} \right)^{1/\alpha} \\ \dfrac{1}{\alpha} = \dfrac{\displaystyle\sum_{i=1}^{n} t_i^{\alpha} \ln t_i}{\displaystyle\sum_{i=1}^{n} t_i^{\alpha}} - \dfrac{1}{d} \displaystyle\sum_{i=1}^{n} d_i \ln t_i \end{cases}$$

Les paramètres ne peuvent pas être exprimés de façon immédiate. En effet, la deuxième équation définit un algorithme qui converge vers $\hat{\alpha}$ pour autant que lui soit fournie une valeur initiale proche de sa valeur réelle. En pratique, cette valeur pourra être l'estimateur obtenu à l'aide d'un logiciel statistique. Une fois $\hat{\alpha}$ obtenu, $\hat{l}$ s'en déduit grâce à la première équation.

Il est donc supposé dans la suite de l'étude que les estimateurs des paramètres $\hat{\alpha}$ et $\hat{l}$ sont connus. La loi suivie par le coût des sinistres si elle correspond à une loi de Weibull est de fait caractérisée.

2.1.3.4. Calcul de la provision individuelle

Dans le chapitre précédent, il a été montré que la fonction de survie de la provision individuelle. peut s'exprimer en fonction de la fonction de survie de la variable « coût du sinistre » et du montant c_i déjà réglé à la date d'inventaire pour le sinistre i.

Soit p la réalisation de la variable provision individuelle. Cette provision valant pour chaque sinistre i : $P = X - c \mid X \geq c$. L'expression de la fonction de survie de la variable P est donc :

$$S_{\alpha,l,c}(p) = \frac{S_{l,\alpha}(c+p)}{S_{l,\alpha}(c)} = \mathbf{exp}\left\{ -\left(\frac{c+p}{l} \right)^{\alpha} + \left(\frac{c}{l} \right)^{\alpha} \right\}$$

A l'aide de cette expression de la fonction de survie, la variable est entièrement définie et connue. Le montant de la provision individuelle peut donc être déterminé de façon unique. Ce montant de provision à constituer représente l'espérance de la variable $P = X - c \mid X \geq c$. Cette espérance peut s'exprimer à l'aide de la fonction de survie :

$$E_{\alpha,l,c}(P) = \int_0^{\infty} S_{\alpha,l,c}(p)\, dp\,.$$

La résolution de cette intégrale se fait à l'aide de la fonction Gamma d'Euler notée $\Gamma(\theta)$ et de la fonction gamma incomplète qui s'expriment respectivement :

$$\Gamma(\theta)=\int_0^\infty t^{\theta-1}\,\mathbf{exp}(-t)\,dt\,.$$

$$\gamma(\theta;x)=\int_0^x t^{\theta-1}\,\mathbf{exp}(-t)\,dt$$

L'espérance de la provision individuelle vaut donc :

$$E_{\alpha,l,c}(P)=\int_0^\infty \mathbf{exp}\left\{-\left(\frac{c+p}{l}\right)^\alpha+\left(\frac{c}{l}\right)^\alpha\right\}dp$$

Le changement de variables suivant est mis en place :

$$u=\frac{c+p}{l}\qquad p=lu-c\qquad du=\frac{dp}{l}$$

$$E_{\alpha,l,c}(P)=l\,\mathbf{exp}\left(\left(\frac{c}{l}\right)^\alpha\right)\int_{c/l}^\infty \mathbf{exp}\left\{-u^\alpha\right\}du$$

Un nouveau changement de variable est effectué :

$$v=u^\alpha\qquad u=v^{1/\alpha}\qquad dv=\alpha u^{\alpha-1}du\qquad du=\frac{dv}{\alpha v^{\alpha-1/\alpha}}$$

$$E_{\alpha,l,c}(P)=\frac{l}{\alpha}\mathbf{exp}\left(\left(\frac{c}{l}\right)^\alpha\right)\int_{c/l^\alpha}^\infty \mathbf{exp}\{-v\}\,v^{\frac{1}{\alpha}-1}dv$$

$$E_{\alpha,l,c}(P)=\frac{l}{\alpha}\mathbf{exp}\left(\left(\frac{c}{l}\right)^\alpha\right)\left[\Gamma\left(\frac{1}{\alpha}\right)-\int_0^{c/l^\alpha}\mathbf{exp}\{-v\}\,v^{\frac{1}{\alpha}-1}dv\right]$$

$$E_{\alpha,l,c}(P)=\frac{l}{\alpha}\mathbf{exp}\left(\left(\frac{c}{l}\right)^\alpha\right)\left[\Gamma\left(\frac{1}{\alpha}\right)-\gamma\left(\frac{1}{\alpha};\left(\frac{c}{l}\right)^\alpha\right)\right]$$

La résolution de cette équation appliquée à chaque valeur c propre à chaque sinistre se fait en pratique à l'aide du logiciel R une fois que les paramètres de la fonction sont estimés.

2.1.3.5. Calcul de la provision totale

La provision totale a pour espérance :

$$\mathbf{E}(\Lambda) = \sum_{i \in I_e} \mathbf{E}(P_i)$$

et se calcule donc en pratique comme la somme des provisions individuelles.

2.2. *Présentation des tests d'ajustements*

Les tests d'ajustements visent à vérifier si les données observées correspondent bien à un modèle théorique supposé.

2.2.1. Le test du χ^2

Soit X la variable aléatoire « coût de sinistres » et soit $\{x_1, x_2, ..., x_n\}$ un échantillon de n réalisations indépendantes de la v.a. X. Soit $L(x)$ la loi de la distribution inconnue de X. L'hypothèse de départ sera que la loi de distribution de X est $L^*(x)$. Les distributions L^* testées dans ce mémoire seront les distributions de Pareto et de Weibull dont les paramètres seront estimés comme décrit précédemment.

Le test d'adéquation se formule donc ainsi :

$$H_0 : L(x) = L^*(x)$$
$$H_1 : L(x) \neq L^*(x)$$

A partir de l'échantillon $\{x_1, x_2, ..., x_n\}$ un histogramme en fréquence de k classes notées C_i avec $i \in [1;k]$ est construit. Le nombre d'observations de X appartenant à la classe C_i est noté O_i. Il est à noter que $\sum_{i=1}^{k} O_i = n$. La construction des classes n'est pas faite selon une règle précise néanmoins, il est souhaitable en pratique que l'effectif de chaque classe soit supérieur à 5 éléments.

La probabilité que la variable aléatoire X suivant la loi théorique $L^*(x)$ prenne une valeur sur le domaine définissant la classe C_i est notée p_i^*. Alors l'effectif théorique de la classe C_i noté E_i vaut $E_i = np_i$.

L'écart entre la réalité issue de l'échantillon et la théorie issue de l'hypothèse H_0 est mesuré par un indicateur I :

$$I = \sum_{i=1}^{k} \frac{\left(np_i^* - O_i\right)^2}{np_i^*} = \sum_{i=1}^{k} \frac{\left(E_i - O_i\right)^2}{E_i}$$

Sous l'hypothèse H_0, l'écart entre E_i et O_i qui représente l'écart entre la distribution théorique et la distribution empirique est distribué normalement. Dans ces conditions, l'indicateur I tend vers une loi du χ^2 à v degrés de liberté où $v =$ nombre de classes - 1 - nombre de paramètres nécessaires à la spécification complète de p_i^*.

La région d'acceptation du test est l'intervalle $\left(0, \chi^2_{v,1-\alpha}\right)$ tel que la probabilité d'une variable du χ^2 à v degrés de liberté prenne une valeur dans cet intervalle soit égale à $1-\alpha$ et avec α qui est l'erreur de première espèce relative au test.

Si la valeur de l'indicateur est supérieure à $\chi^2_{v,1-\alpha}$, alors l'hypothèse H_0 n'est pas retenue et la distribution théorique n'est pas considérée comme en adéquation avec l'échantillon de données.

2.2.2. L'adéquation aux modèles censurés

Une fois ces tests mis en place, la validation ou non de l'hypothèse H_0 permet de conclure ou non à l'adéquation de la distribution empirique des données à la distribution théorique.

Pour la validité des tests mis en place, il faut que les paramètres estimés le soit à l'aide de la méthode du maximum de vraisemblance. Dans les modèles paramétriques développés, la méthode du maximum de vraisemblance n'a pas toujours pu être mise en place. Ainsi, même si la comparaison avec la valeur $\chi^2_{v,1-\alpha}$ n'est pas possible, l'indicateur fournit une distance. C'est le plus petit indicateur qui révèle la loi la plus proche de la distribution des données.

En fonction de la réponse fournie par le test, le calcul de la provision se poursuit en gardant le modèle sélectionné.

3. Chapitre 3 - Mise en pratique et Résultats

3.1. *Les données*

3.1.1. Le type de données

L'étude réalisée dans ce mémoire nécessite des données correctement renseignées. En effet, les sinistres doivent être connus de façon individuelle donc pour la compagnie d'assurance, cela signifie qu'il faut enregistrer pour chaque sinistre, le montant de chacun des règlements effectués ainsi que les dates correspondantes. De plus, si un sinistre est clos, il faut préciser cette information au plus tôt afin d'avoir les données les plus correctes et complètes possibles.

Les données fournies par la compagnie d'assurance afin de réaliser ce mémoire sont des données parfaitement renseignées et qui conservent donc les informations sinistre par sinistre. Seuls les contrats sinistrés sont enregistrés dans le fichier de données. Il n'est donc pas possible de connaitre le ratio de sinistralité du portefeuille. Les données proviennent d'un inventaire du portefeuille réalisé au 31/12/2008. Cette date représente donc dans la suite, la date d'étude.

Les données initiales sont présentées sous forme d'un tableau comportant les renseignements suivants :

- ✓ exercice comptable ;
- ✓ numéro de sinistre ;
- ✓ code de contrat ;
- ✓ nom de l'assuré ;
- ✓ exercice de survenance ;
- ✓ règlement cumulé ;

et d'un autre fichier de données présentant en correspondance de chaque numéro de sinistre l'information :

- ✓ date de clôture du sinistre.

Dans un premier temps, ces deux fichiers d'information sont agrégés pour ne former qu'une seule base de données qui constituera la base de travail initiale à l'étude menée dans ce mémoire.

Les données sont classées par rapport aux années comptables qui correspondent aux années de règlement des sinistres. Chaque sinistre est présent sur plusieurs années comptables puisque les règlements ne sont pas effectués en une seule fois et c'est donc pour cela qu'il faut prévoir le coût total de sinistre et la provision à constituer pour faire face aux futurs règlements. Un sinistre correspond à un unique code de contrat mais un contrat peut comporter plusieurs sinistres qui seront traités séparément. C'est donc sur chaque sinistre que porte l'attention et non sur chaque contrat.

L'exercice de survenance correspond à l'année durant laquelle le sinistre a eu lieu. Cette information permet de connaitre par la suite la durée de remboursement de chaque sinistre.

Les règlements cumulés sont des montants cumulés de remboursement depuis la survenance du sinistre. Il se peut qu'il y ait des sauts de règlements dans le sens où un sinistre peut-être remboursé plusieurs années de suite puis le paiement cesse et reprend enfin une ou plusieurs années après. C'est l'information « montant cumulé » qui est importante et étudiée dans ce mémoire.

La date de clôture du sinistre va constituer l'information de censure. Si pour un sinistre, une date de clôture est connue, ce dernier est considéré comme terminé et l'information est donc complète[5]. En revanche, s'il n'y a pas de date de clôture, le sinistre est considéré comme encore en cours[6], il convient dans ce cas de constituer une provision alors que le coût total du sinistre n'est pas encore connu. C'est cette information qui est censurée et recherchée.

---

[5] Lorsque le sinistre est clos, il n'y a pas de censure, l'information est complète et l'indicatrice de censure sera égale à 1.

[6] Lorsque le sinistre est en cours, il y a censure de l'information « coût total du sinistre » et l'indicatrice de censure sera égale à 0.

3.1.2. Le portefeuille

Le portefeuille de contrats se décompose en trois types de contrats qui correspondent chacun à un risque couvert par l'assureur et contre lequel veut se prémunir l'assuré :

- ✓ l'assurance construction garantie décennale ;
- ✓ l'assurance responsabilité civile en cas d'accident;
- ✓ l'assurance dommages aux biens.

Ces trois types de contrats d'assurance s'adressent à des professionnels qui doivent se préserver dans le cadre de leurs activités contre différents sinistres.

Les descriptions faites des différents traitements réalisés sur les données ainsi que les méthodes de calcul mises en pratique seront dans la suite de ce mémoire détaillées pour les contrats correspondants à l'assurance construction garantie décennale.

Afin d'affiner autant que possible l'étude, un découpage et regroupement des sinistres semblables est nécessaire. Cependant, une segmentation trop grande engendrerait une information plus faible avec un nombre de données peut-être insuffisant. Ainsi, la segmentation du portefeuille consistant à ne garder que les contrats d'assurance construction garantie décennale apparait satisfaisante.

3.1.2.1. Le nettoyage des données

Le fichier de données initial correspondant aux contrats d'assurance construction garantie décennale est composé de 12 782 lignes. De ces données brutes, sont conservées les informations :

- ✓ exercice comptable ;
- ✓ numéro de sinistre ;
- ✓ exercice de survenance ;
- ✓ règlement cumulé ;
- ✓ date de clôture du sinistre.

Un croisement des différentes données permet de créer une nouvelle base de travail dont la lecture se fait sinistre par sinistre. Chaque sinistre ne correspond alors qu'à une seule ligne. Pour chaque sinistre (présenté en ligne) et chaque exercice comptable (présenté en colonne), le montant réglé cumulé est indiqué. Pour un numéro de sinistre, sont indiquées les informations d'année de survenance et de clôture du sinistre.

Les données de l'assurance construction garantie décennale sont composées de 902 sinistres distincts qui sont survenus entre 1983 et 2008, année d'inventaire dans le cadre du portefeuille étudié.

Sur ces sinistres étudiés, il est possible de rencontrer des informations fausses ou erronées. Les sinistres ainsi concernés doivent être écartés de l'analyse afin de ne pas en fausser les résultats. Les motifs d'exclusion d'un sinistre sur la base uniquement d'une analyse des données peuvent être de nature diverses :

- ✓ date de clôture incohérente ;
- ✓ montant total de sinistre réglé négatif ;
- ✓ montant total de sinistre réglé égal à 0[7].

L'analyse des données a permis d'identifier plusieurs valeurs aberrantes ce qui a conduit à réduire la base de données de l'étude à 825 sinistres.

3.1.2.2. La revalorisation des montants de sinistre

Les données fournies par la compagnie d'assurance sont composées de montants de sinistre non revalorisés ; c'est-à-dire que les montants de sinistres réglés sont en valeur de l'année comptable de chaque règlement. Pour qu'il y ait homogénéité entre tous les coûts de sinistres, il a été retenu de ramener tous les montants payés à une même date de calcul. Les montants payés sont donc tous ramenés en valeur 2009.

Les données initiales sont présentées en montant total cumulé par année comptable. Il a donc été nécessaire dans un premier temps d'isoler les montants annuels.

Afin de revaloriser les montants réglés, il a par ailleurs été nécessaire d'appliquer les facteurs d'incidence. Les facteurs d'incidence servent à transposer les montants réglés une certaine année passée en un montant qui aurait été payé si le sinistre considéré intervenait en 2009.

Les facteurs d'incidence

La construction des facteurs d'incidence est dépendante de :

---

[7] Ces montants égaux à 0 € correspondent à des sinistres clos ou dont la survenance est intervenue avant l'année 2000 et pour lesquels, il n'y a pourtant eu aucun dédommagement de réglé. C'est pour cela qu'il a été choisi d'exclure de tels sinistres de l'étude.

- ✓ l'évolution de l'indice de coût ;
- ✓ l'année de survenance du sinistre ;
- ✓ le délai de paiement.

L'indice des coûts des sinistres et plus particulièrement son évolution est fournie par le client. Pour les années 1983 et 1984, ce taux n'étant pas fournit, il a été fixé à 5 % au vu des indices des années suivantes. A compter de 2009, ce taux est pris fixe à 3 %. Il s'agit là d'hypothèses de calcul. Ce taux sera noté $tx_{évol}$.

Pour un sinistre survenu en année n, et réglé t années plus tard avec $t \in [0;25]$, le facteur d'incidence par lequel doit être multiplié le montant réglé en année $n+t$ est égal à :

$$f(n,t) = \frac{\text{indice de sinistre}_{\text{ année 2009 + t}}}{\text{indice de sinistre}_{\text{ année n}}}.$$

Ainsi, le montant réglé en année $n+t$ multiplié par $f(n,t)$ correspond au montant qui serait réglé avec un délai de t années pour un sinistre qui interviendrait en 2009. L'ensemble des facteurs d'incidence est donc représenté par un triangle dépendant de chaque année de survenance et de chaque année de paiement (c'est-à-dire du délai après lequel un certain montant a été réglé).

L'indice de sinistre

L'indice de sinistre est défini pour chaque année de façon récurrente. Ainsi, en 1982, cet indice est pris égal à 100 %. Ensuite, l'indice est défini par récurrence pour chaque année $n \in [1983;2034]$[8].

$$indice_n = indice_{n-1} * (1 + tx_{évol}).$$

Les chiffres des indices calculés chaque année sont indiqués en annexes.

Une fois les indices de sinistres calculés, les facteurs d'incidences sont déduits à l'aide de la formule indiquée précédemment. Le tableau des facteurs d'incidence en fonction de l'année de survenance et des délais de paiement est présenté en annexes de ce mémoire.

[8] 2034 correspond à un sinistre intervenu en année de 2009 et qui aurait un paiement après 25 ans. Les 25 ans correspondent à l'historique étudié (2008-1983)

Les montants revalorisés

Ainsi, à l'aide des facteurs d'incidence, pour chaque sinistre survenu en année n et chaque paiement réalisé en année $n+t$, le montant réglé est transposé au montant qui serait réglé si le sinistre intervenait en 2009. Le travail est effectué sur les montants annuels réglés. Les montants sont ensuite cumulés.

Les montants de sinistres réglés fin 2008 pour chaque sinistre sont ainsi obtenus, et ces montants correspondent à ce qui serait payé si les sinistres intervenaient en 2009. C'est à partir de ces coûts de sinistre clos ou en cours que le travail est effectué par la suite. La base de travail des coûts de sinistres est créée.

3.1.3. L'analyse des données

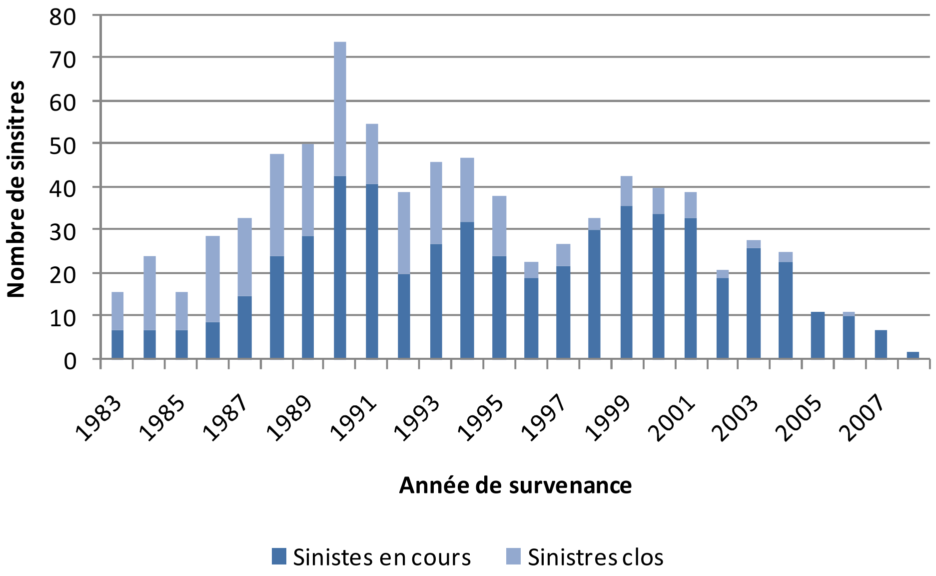
Une analyse de donnée peut-être menée sur les sinistres clos. Ceux-ci constituent la base de référence de la variable coût du sinistre. Les sinistres clos sont au nombre de 268 à la date d'étude contre 557 sinistres encore en cours.

3.1.3.1.La répartition des sinistres

La répartition des sinistres clos ou en cours est la suivante :

Répartition	Sinistes en cours	Sinistres clos
Effectif	557	268
Proportion	68%	32%

La répartition des 825 sinistres étudiés par année de survenance et en fonction du fait qu'ils soient clos ou en cours, est la suivante :



Graphique 3 : Répartition de la survenance des sinistres intervenus sur les contrats de construction garantie décennale depuis 1983 jusqu'à la date d'étude en fonction de l'information de censure.

Cette répartition montre que des sinistres survenus dès 1983 peuvent encore être en cours plus de 20 après. La garantie proposée par l'assureur a une durée de vie de 10 ans et pourtant les délais de remboursements complets des sinistres peuvent être beaucoup plus importants.

Il est à noter par ailleurs une évolution en fonction de l'année de survenance ; globalement, plus l'année de survenance est proche de la date d'étude, plus le nombre de sinistres clos est faible.

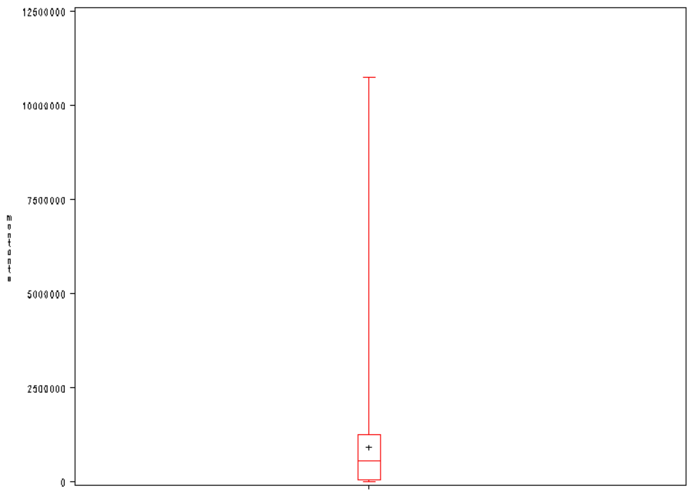
3.1.3.2. L'analyse des montants de sinistres clos

Les sinistres clos sont au nombre de 268. C'est sur ces coûts de sinistres que porte la recherche à une loi d'adéquation. Ils représentent l'historique sur lequel se base l'étude. Une étude statistique est faite sur ces données « coûts de sinistres » en fonction des montants payés mais aussi des dates de survenance du sinistre et de clôture du contrat.

Les statistiques majeures des montants de sinistres clos sont (les chiffres sont en euros):

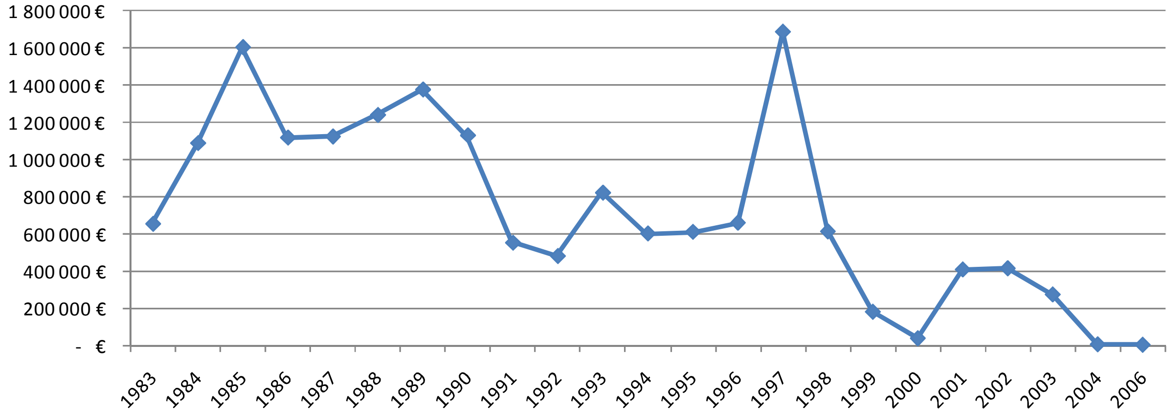
```
Min.    :      24
1st Qu.:   54664
Median :  555096
Mean   :  910840
3rd Qu.: 1240454
Max.    :10742608
```

La moyenne des coûts de sinistres est supérieure à la médiane. Ceci signifie qu'il existe des montants de sinistres très élevés mais qui ne représentent qu'une faible proportion, comme illustré par la représentation des boxplot :



Graphique 4 : Représentation à l'aide d'une boxplot de la dispersion des coûts de sinistres

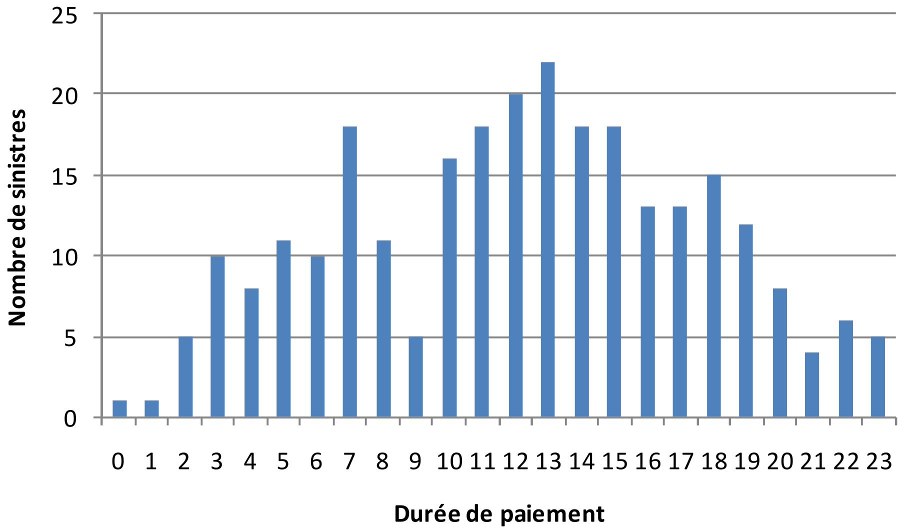
Cette disparité dans les montants de sinistres se retrouve aussi lors de l'étude des coûts par année de survenance. Sur les données de sinistres clos, les années de survenance s'étalent de 1983 à 2006. En 2007 et 2008, les sinistres qui sont survenus ne sont pas encore clos à la date d'étude.



Graphique 5 : Moyenne des montants de sinistres en fonction de l'année de survenance

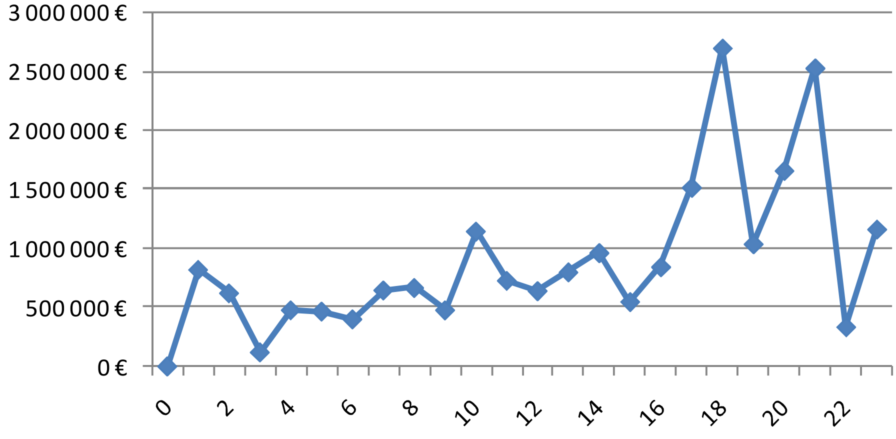
3.1.3.3. Analyse sur les durées de paiements

Les durées de paiements des sinistres peuvent varier entre 0 année (ce qui signifie que le sinistre a totalement été réglé durant l'année de survenance) et 23 ans.



Graphique 6 : Répartition des sinistres en fonction de leur durée moyenne de paiements

Les durées de paiements sont variables selon les sinistres. La durée moyenne de règlement est d'un peu plus de 12 ans. Il est intéressant de connaitre le coût moyen d'un sinistre en fonction de sa durée de règlement.



Graphique 7 : Moyenne des montants de sinistres en fonction de la durée de paiement

L'évolution des coûts moyens des sinistres en fonction de la durée de règlement met en avant le fait que les sinistres dont la durée de paiement est importante ont en moyenne des coûts plus élevés.

3.1.4. La mise en place des triangles de liquidation

Pour l'exécution des calculs sur la base de Chain-Ladder, les données doivent être présentées sous forme de triangles de liquidation comme cela a été décrit dans le 1er chapitre de ce mémoire. Pour cela, les sinistres ne sont plus considérés individuellement mais agrégés par année de survenance et par année de développement. Aucune distinction n'est faite entre les sinistres clos et les sinistres encore en cours au moment de l'étude. Ce qui compte, c'est le passage entre deux années de développement.

Ce travail de mise en place des données est donc réalisé sur la base des montants de sinistres revalorisés.

Il faut dans un premier temps sommer les sinistres par année de survenance et par année de paiements. Cela forme un triangle supérieur droit. Ensuite, au lieu d'exprimer en fonction des années de règlement, les montants réglés sont considérés par rapport aux délais de paiements c'est-à-dire par rapport aux années de développement. Cela constitue donc un triangle supérieur gauche au format traditionnellement utilisé dans la mise en place de Chain-Ladder.

Il est à noter en observant le triangle ainsi formé que pour les années de survenance de 1983 à 1988 ; le montant réglé la première année, c'est-à-dire avec un délai de paiement nul, est égal à 0. Ceci provient peut-être d'une information insuffisante. Il faudra prendre en compte ce fait lors du calcul des facteurs de développement.

Les montants présentés dans les triangles sont des paiements cumulés des sinistres.

3.1.5. La mise en place des modèles statistiques

Les données qui servent de base à la mise en place des modèles non paramétriques et paramétriques sont des données ligne à ligne où chaque sinistre correspond à une seule ligne et inversement. Pour chaque ligne de sinistre, sont indiqués :

- ✓ le numéro de sinistre ;
- ✓ le montant cumulé réglé chaque année comptable ;
- ✓ la date de clôture du sinistre le cas échéant ;

Pour parfaire la mise en forme des données par rapport au travail nécessaire à effectuer, une indicatrice de sinistre est créée et notée :

$$D_i = \begin{cases} 1 \text{ si le sinistre est clos} \\ 0 \text{ si le sinistre est en cours} \end{cases}$$

Lorsque l'indicatrice vaut 0, l'information n'est pas complète et le sinistre est en cours. La notion de censure intervient à ce niveau. L'information « coût total de sinistre » est censurée.

En pratique, seuls les montants cumulés pour l'année d'étude qui est l'année 2008 sont nécessaires à l'étude, ils constituent en effet les observations des variables $T_i = X_i \wedge C_i$.

Ainsi, toute l'information nécessaire à la mise en place de modèle de censure est connue. Ces informations peuvent être synthétisées en deux colonnes :

- ✓ T_i : le montant total réglé en 2008 ;
- ✓ D_i : information de censure.

Les données sont composées de 830 sinistres répartis en 269 sinistres clos et 561 sinistres dont le dossier n'est pas clos au 31/12/2008.

3.2. *Application à l'aide des méthodes de triangulations*

La méthode déterministe de Chain-Ladder permet de connaitre les provisions à mettre en place en correspondance des années de survenance des sinistres et ainsi d'en déduire la provision globale à constituer à la date d'inventaire.

Pour cela, la première étape consiste à évaluer les facteurs de Chain-Ladder aussi appelés facteurs de développement. Ce facteur reflète le passage d'une année de développement à l'autre. Il se calcule comme explicité dans le premier chapitre de ce mémoire.

Il a été noté lors de l'observation du triangle que certaines informations avaient pu être omises notamment certains montants concernant l'année de développement 0. Dans le calcul du facteur de Chain-Ladder f_1, les années prises en compte seront donc les années 1989 à 2007 uniquement (au lieu des années 1983 à 2007).

Les coefficients de développement sont les suivants :

f_1	3,651	f_6	1,257	f_{11}	1,190	f_{16}	1,072	f_{21}	1,035
f_2	2,375	f_7	1,236	f_{12}	1,113	f_{17}	1,067	f_{22}	1,024
f_3	1,744	f_8	1,187	f_{13}	1,107	f_{18}	1,059	f_{23}	1,016
f_4	1,334	f_9	1,193	f_{14}	1,088	f_{19}	1,033	f_{24}	1,006
f_5	1,232	f_{10}	1,151	f_{15}	1,081	f_{20}	1,041	f_{25}	1,004

Ces coefficients sont ensuite appliqués au triangle des montants cumulés afin d'obtenir les totaux de montants de sinistres réglés après un délai de développement de 25 ans. Le triangle de paiements connus est ainsi complété pour former un rectangle de paiements totaux. Les différences sont ensuite faites entre les coûts totaux estimés des sinistres par année de survenance et les montants payés à fin 2008.

Le montant total de provision s'élève à **442 Millions d'Euros**.

Ce montant de provision estimé sert de comparaison avec le montant estimé à l'aide des modèles de durée en tenant compte de la censure.

3.3. *Application avec les modèles de données censurées*

3.3.1. Etude non paramétrique- Estimateur de Kaplan-Meier

Les données sources sur lesquelles sera appliquée la méthode de Kaplan-Meier sont constituées des coûts des sinistres notées $T_i = \inf\left(X_i, C_i\right)$ et de l'information de censure sur ces sinistres. Cette information est notée D_i qui vaut 1 si le sinistre est clos et donc si l'information sur celui-ci est complète, et 0 lorsque le sinistre est en cours au moment de l'étude et donc il y a censure. Pour chaque sinistre i les couple d'information $\left(T_i, D_i\right)$ est disponible.

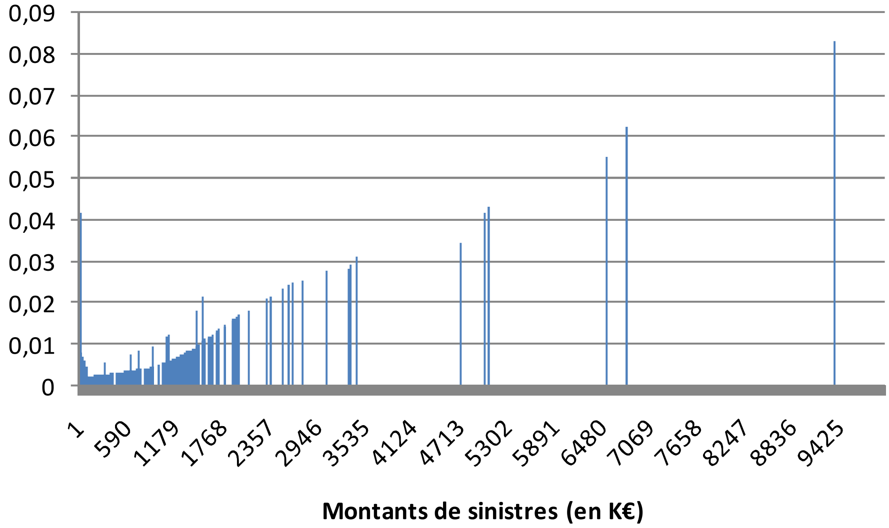
L'estimateur de Kaplan-Meier est un estimateur non paramétrique de la fonction de survie notée $S(t)$. Il permet d'approcher la valeur empirique prise par le risque de sortie de l'état qui correspond dans le cadre de cette étude à la clôture du sinistre. Cette approximation ne nécessite pas d'adopter une quelconque spécification de loi. Elle est principalement utilisée dans le cadre de données incomplètes telles que les données censurées. Les méthodes de calculs habituellement utilisées pour calculer des durées de vie sont appliquées ici aux montants des sinistres.

L'estimation par la méthode de Kaplan-Meier se fait en pratique à l'aide du logiciel R[9].

A partir des résultats fournis par le logiciel, un travail est effectué sur les taux conditionnels de sortie à un moment donné. Ces taux sont notés q_x et correspondent à la probabilité que le sinistre ait atteint son montant cumulé total (c'est-à-dire qu'il soit totalement réglé et donc devienne clos) entre le montant x € et $x+1000$ €.

La représentation de ces taux en fonction des montants de sinistres donne le graphique suivant :

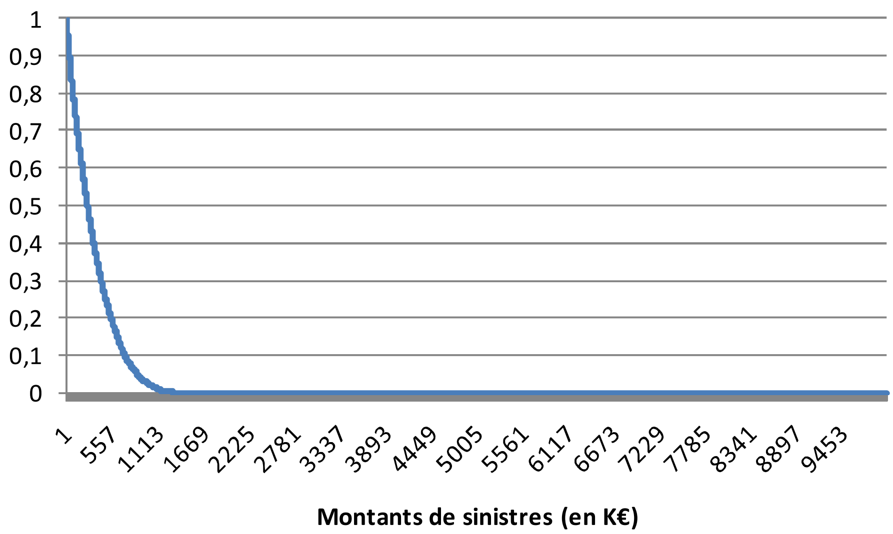
[9] Logiciel R version 2.4.0.



Graphique 8 : Taux conditionnels de sortie

Pour de nombreux montants, le taux de sortie q_x est nul. Cela provient du fait que les données ne présentent pas de sorties dans l'intervalle $[x; x+1000]$. Il n'y a pas de sorties car le nombre de données est insuffisant. Le choix est donc fait d'adapter un vecteur q_x modifié. Les q_x sont globalement croissants. Lorsque q_x est égal à 0, il est donc supposé égal à la valeur q_{x-1}.

La fonction de survie introduite à partir des q_x adaptés est de la forme suivante :



Graphique 9 : Fonction de survie estimée à partir de Kaplan-Meier

L'étude non paramétrique fournit une fonction de survie au plus proche des données. Dans le cadre de ce mémoire, une étude paramétrique est aussi réalisée afin de lisser la fonction de répartition et d'exprimer la loi de la provision en plus de la loi du coût d'un sinistre.

3.3.2. Etude paramétrique

Pour cette étape, l'étude paramétrique consiste à approcher et à estimer la loi suivie par les variables « montants de sinistres » à l'aide de distributions connues. Les modèles étudiés dans le cadre de ce mémoire sont dans un premier temps les distributions de Pareto et de Weibull qui sont des lois rencontrées couramment dans l'étude des montants de sinistres.

3.3.2.1. Modèle de Pareto

Le modèle de Pareto est le premier modèle paramétrique mis en place dans le cadre de cette étude. Lorsqu'un échantillon de données suit une loi de Pareto, sa fonction de survie et sa fonction de hasard s'expriment aisément. De là, peut se déduire l'expression de la log-vraisemblance dans le cadre de données censurées :

$$\ln L(m,\alpha) = -\alpha \sum_{i=1}^{n} \ln(t_i) + \alpha \times n \times \ln(m) + d \times \ln(\alpha) - \sum_{i=1}^{n} d_i \ln(t_i).$$

On cherche à vérifier à l'aide des modèles paramétriques si la loi suivie par les coûts des sinistres est une loi de Pareto. Pour cela, on étudie sa fonction de survie. Celle-ci s'exprime en fonction de deux paramètres qui sont estimés par rapport aux données.

$\hat{m}$ se définit en tant que montant minimum des coûts des sinistres.

$\hat{\alpha} = \dfrac{d}{\sum_{i=1}^{n} \ln\left(\frac{t_i}{\hat{m}}\right)}$ s'exprime donc en fonction du nombre de sinistres clos parmi les données et des montants connus reversés à la date d'inventaire (c'est-à-dire le coût du sinistre si celui-ci est clos ou le montant déjà réglé si le sinistre est encore en cours).

Estimation des paramètres

Le montant minimum des sinistres est connu à la lecture des données. Tous les sinistres sans distinguer ceux qui sont en cours de ceux qui sont clos sont pris en compte. La valeur de d est égale au nombre de sinistres clos c'est-à-dire les sinistres dont l'information est non censurée et pour lesquels l'indicatrice $D_i = 1$.

L'évaluation des paramètres estimés de la loi de Pareto à l'aide de la méthode du maximum de vraisemblance donne les résultats suivants :

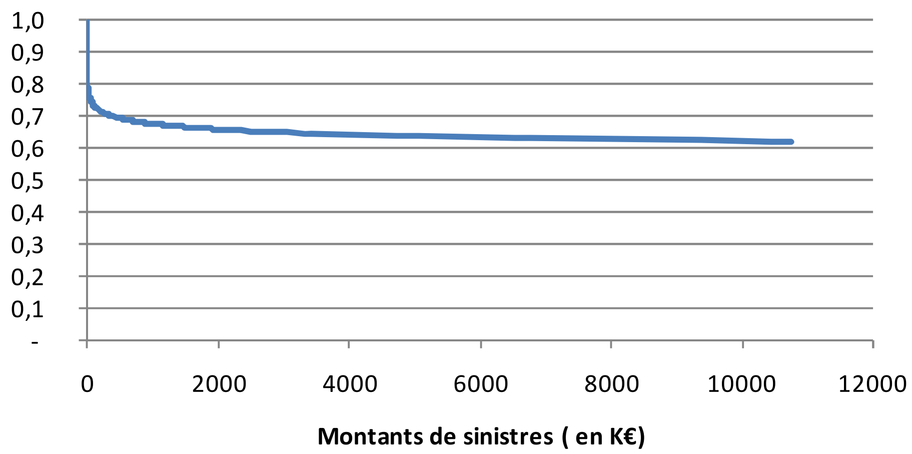
Paramètres estimés	
m	24
alpha	0,04

Représentations de la fonction de survie

Les paramètres ainsi estimés, il est alors possible de tracer la fonction de survie relative à un modèle de Pareto dans le cadre des modèles censurés. La fonction de survie s'exprime :

$$S_{\hat{m},\hat{\alpha}}(x) = \left(\frac{x}{\hat{m}}\right)^{-\hat{\alpha}}$$

La fonction est calculée pour tous les X_i connus c'est-à-dire pour tous les montants relatifs aux sinistres clos à la date d'inventaire. La fonction de survie du modèle paramétrique de Pareto se représente comme suit :



Graphique 10 : Fonction de survie de Pareto des montants de sinistres

La fonction de survie ainsi représentée à l'échelle illustre bien le problème de convergence de la fonction. En effet, la fonction de survie est décroissante de 1 vers 0. Or, la fonction de survie modélisée à l'aide d'une loi de Pareto de paramètres estimés en tenant compte de la censure ne décroit vers 0 que très lentement. L'adéquation des montants de sinistres à une loi de Pareto est donc rejetée.

Une loi de Pareto avec des paramètres différents non estimés par maximum de vraisemblance ne donne pas non plus des résultats satisfaisants en termes d'adéquation.

L'étude continue donc vers la recherche d'une adéquation entre la loi des montants de sinistres et la loi de Weibull.

3.3.2.2. Modèle de Weibull

L'application des données au modèle de Weibull suppose donc que les données suivent une loi de Weibull dont les paramètres doivent être estimés.

Les équations de la vraisemblance conduisent à un algorithme. L'estimation du paramètre α fournit ensuite la valeur du deuxième paramètre l.

Dans le cadre de ce mémoire, l'estimation des paramètres se fait à l'aide du logiciel statistique EayFit[10]. Ce logiciel permet, en insérant les données brutes, de tracer les fonctions de densité, de répartitions, de survie, de hasard et de hasard instantané. Il fournit également une analyse descriptive des données. Le logiciel met en parallèle les données et les lois statistiques et paramétriques connues. Pour chacune de ces lois, il fournit l'estimation des paramètres optimaux (c'est-à-dire ceux qui se rapprochent le plus des données).

L'utilisation de ce logiciel nécessite de ne travailler que sur des données non censurées. L'estimateur de Kaplan-Meier a fourni une fonction de répartition empirique. C'est à partir de cette fonction empirique que va être construit un vecteur de données pouvant être utilisé dans le logiciel EasyFit, prenant ainsi en compte la censure.

[10] EasyFit Professional Version 5.2.

Construction d'un unique vecteur de données prenant en compte la censure

Afin d'utiliser correctement le logiciel EasyFit, à partir des deux vecteurs (coût du sinistre et indicatrice de la censure), il convient de construire un unique vecteur regroupant ces deux informations. L'estimateur de Kaplan-Meier permet d'établir la fonction de répartition empirique tenant compte de la censure. C'est donc sur la base de la fonction de répartition empirique que va être construit le vecteur servant à l'utilisation d'EasyFit. En effet, l'expression de la fonction empirique permet de construire un nouveau vecteur (X_i).

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \le x\}}$$

Pour vérifier que le vecteur représente bien la synthèse des informations (T_i, D_i), la fonction de répartition de celui-ci est tracée. Elle est identique à la fonction de répartition construite à partir de Kaplan-Meier. Le vecteur construit peut donc être utilisé dans le logiciel EasyFit.

Utilisation du logiciel EasyFit

Le vecteur est entré dans le logiciel EasyFit afin que celui-ci estime les paramètres des lois statistiques se rapprochant des données brutes.

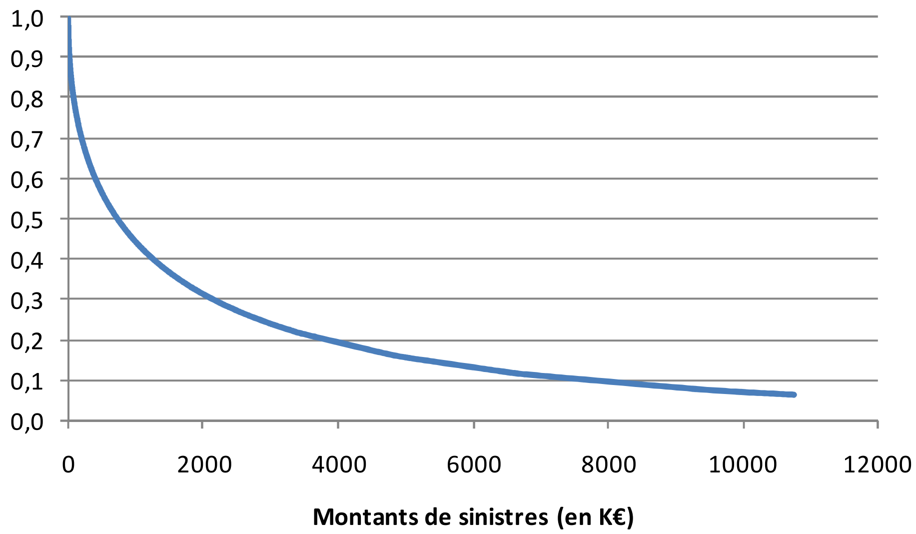
Les paramètres fournit par le logiciel sont donc :

Paramètres estimés	
l	508 418
alpha	0,49

Les paramètres ainsi estimés ne correspondent pas aux paramètres estimés à l'aide du maximum de vraisemblance. Ils caractérisent néanmoins la loi de Weibull la plus proche des données. Ce sont donc les paramètres retenus. Il est alors possible de tracer la fonction de survie relative à un modèle de Weibull dans le cadre des modèles censurés. La fonction de survie s'exprime :

$$S_{\alpha,l}(x) = \mathbf{exp}\left\{-\left(\frac{x}{l}\right)^{\alpha}\right\}$$

Sa représentation graphique est la suivante :



Graphique 11 : Fonction de survie de Weibull des montants de sinistres

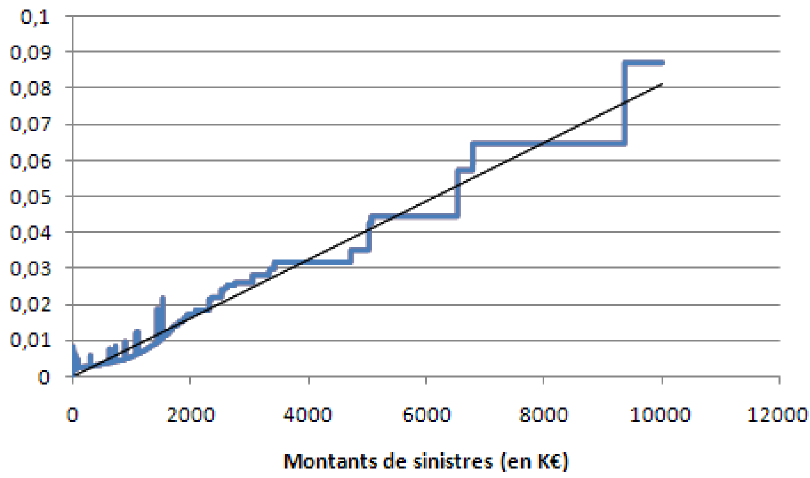
Au vu de la représentation graphique de la fonction de survie, il n'y a pas adéquation entre la loi de Weibull de paramètres estimés (α, l) et les données.

L'adéquation des données aux modèles paramétriques de Pareto et Weibull n'a pas été mise en avant. Une autre méthode est donc mise en place.

3.3.2.3. Un autre modèle paramétrique

Un nouveau modèle paramétrique est mis en place. Celui-ci est inspiré de la représentation des q_x adaptés en fonction des montants de sinistres. Afin de faciliter les calculs par la suite, le travail se fait sur les valeurs de : $-\ln(1-q_x)$.

La représentation graphique de la fonction $f : x \to -\ln(1-q_x)$ est la suivante :



Graphique 12 : Fonction f et droite de tendance

La fonction $f : x \to -\ln(1-q_x)$ peut être modélisée à l'aide d'une droite d'équation :

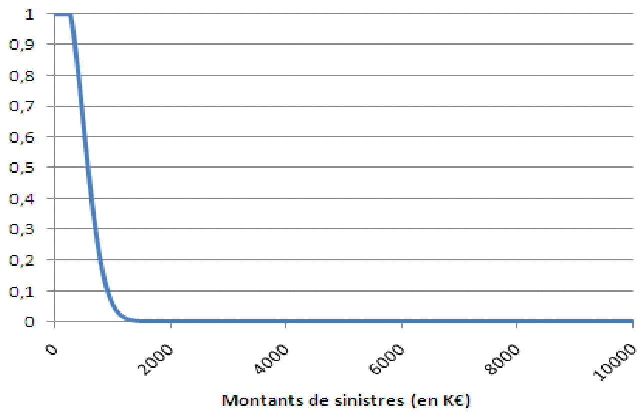
$$-\ln(1-q_x) = ax + b$$

Paramètres estimés	
a	7,88E-09
b	-1,07E-03

La fonction de survie s'exprime [11]:

$$S_{a,b}(x) = \exp\left(-\left(\frac{ax^2}{2} + \left(b - \frac{a}{2}\right)x\right)\right)$$

[11] Le calcul est présenté en annexes



Graphique 13 : Fonction de survie estimée

La fonction de survie ainsi estimée est proche de la fonction de survie estimée à l'aide des modèles non paramétriques.

3.3.2.4. Comparaison des modèles

Les différents modèles paramétriques mis en place dans le cadre de cette étude doivent désormais être comparés afin d'obtenir comme distribution de référence, le modèle le plus proche possible des données afin d'obtenir un montant de provision pertinent.

Pour réaliser une comparaison des différents modèles, l'indicateur du test du χ^2 est calculé pour le modèle de Pareto, de Weibull et pour le modèle relatif à la fonction $f : x \to -\ln(1-q_x)$.

Pour comparer les différents modèles paramétriques mis en place, l'indicateur du χ^2 est pris comme indicateur de la distance entre les données et les valeurs fournies par la distribution modélisée. La mise en place du calcul de l'indicateur est la même pour toutes les adéquations testées.

Les données initiales sont les coûts de sinistres clos. Ces données constituent la distribution d'une variable. Le test d'adéquation consiste à comparer cette distribution à la distribution qui serait obtenue si la variable suivait une des lois statistique choisie.

La procédure de réalisation du test se fait comme explicité dans le chapitre 2 de ce mémoire en plusieurs étapes. La description faite ici correspond aux tests d'adéquation des deux modèles.

1. Segmentation des données et décompte des effectifs

Les valeurs de la variable « coût de sinistres » qui sont comprises entre 24 et 10 742 608 € sont segmentées en k classes. Le choix des classes C_i est fait de façon aléatoire mais en cherchant à garder une certaine homogénéité. La segmentation des classes étant un élément important, deux méthodes ont été retenues pour le choix des classes.

La première segmentation consiste à choisir des classes dont le « pas » est quasiment identique pour chacune d'entre elles (sauf aux extrémités de l'intervalle de valeurs). Ceci fournit donc $k = 21$ classes dont les effectifs notés O_i sont compris entre 5 et 16 sauf pour la première classe composée de 66 éléments constitutifs.

La seconde segmentation mise en place est basée sur le nombre d'éléments par classe. Il est choisi que chaque classe hormis la dernière doit comporter 20 éléments. Les bornes des classes sont donc déduites des montants des coûts de sinistres. Ceci fournit donc 13 classes de 20 éléments et 1 classe de 8 éléments ce qui amène à $k = 14$ classes.

2. Probabilités des lois testées

Pour chaque classe C_i, la probabilité que la variable aléatoire suivant la distribution empirique appartienne à cette classe est calculée. Cette probabilité est notée p_i*. Pour la calculer, il faut se servir de la fonction de répartition du modèle. La fonction de répartition $F(X)$ est égale à $F(X)=1-S(X)$ où $S(X)$ est la fonction de survie calculée à l'aide des paramètres du modèle. Ainsi,

$$\begin{aligned} p_i* &= P(X \in C_i) \\ &= P(X \in [a_i ; b_i[) \\ &= P(a_i \leq X < b_i) \\ &= F(b_i) - F(a_i) \end{aligned}$$

La fonction de répartition est donc évaluée pour chaque borne de classes.

Ensuite, l'effectif théorique de chaque classe est évalué en multipliant cette probabilité par l'effectif total, c'est-à-dire par le nombre de sinistres clos à la date d'étude. L'effectif est ici de 268 coûts de sinistres. Pour chaque classe C_i, l'effectif théorique est noté $E_i = n \times p_i^*$.

3. Mesure de l'écart

Pour chaque classe, l'écart entre l'effectif réel et l'effectif théorique est mesuré. L'indicateur d'écart est calculé :

$$I=\sum_{i=1}^{k}\frac{\left(np_i^*-O_i\right)^2}{np_i^*}=\sum_{i=1}^{k}\frac{\left(E_i-O_i\right)^2}{E_i}$$

Pour l'adéquation à la loi de Weibull, la séparation des classes par effectif donne un indicateur égal à 115 alors que la séparation des classes par montants donne un indicateur égal à 91.

Pour l'adéquation à la loi de Pareto, la séparation des classes par effectif donne un indicateur égal à 1800 alors que la séparation des classes par montants donne un indicateur égal à 1847. Ce chiffre important vient du fait que la dernière classe contient en théorie un nombre important d'éléments par rapport aux données observées.

Pour l'adéquation à la loi définit à partir de la fonction $f(x)$, la séparation des classes par effectif donne un indicateur égal à 35 alors que la séparation des classes par montants donne un indicateur égal à 33.

Les indicateurs les plus faibles sont obtenus pour l'adéquation à la fonction $f(x)$.

3.3.2.5. Calcul de la provision

Les comparaisons des modèles ont permis de sélectionner le modèle paramétrique relatif à la fonction *f*. A partir de cette expression et du calcul exposé dans le chapitre 1, la loi de la provision s'exprime ainsi :

$$E(P_i)=\frac{1}{S_{a,b}(c_i)}\times\int_{c_i}^{+\infty}S_{a,b}(t)dt$$

Les valeurs des coûts de sinistres c_i relatifs aux sinistres en cours sont renseignées dans R[12]. L'équation fournissant la valeur de l'espérance de la provision individuelle pour chaque sinistre i est résolue. L'ensemble des provisions individuelles est alors connue et la somme de ces espérances donne par linéarité l'espérance du total des provisions à mettre en place pour faire face aux sinistres en cours.

Le montant de la provision totale est de **220 Millions d'Euros.**

[12] Le code est présenté en annexes

3.3.2.6. Conclusion

Le montant de la provision estimée à l'aide des modèles paramétriques est inférieur au montant estimé à l'aide de la méthode Chain-Ladder. Ceci s'explique en partie car la méthode Chain-Ladder prend en considération les sinistres dits tardifs.

Dans la branche assurance construction garantie décennale, les provisions relatives aux tardifs représentent environ 30 % du montant des provisions globales calculées. Pour comparer la provision trouvée à l'aide des modèles paramétriques au montant trouvé à l'aide de Chain-Ladder ; les provisions relatives aux tardifs sont déduites.

Les provisions relatives aux sinistres déclarés sont estimées par la méthode Chain-Ladder à **309 Millions d'Euros**.

Le modèle paramétrique fournit donc un montant de provision plus faible que la méthode Chain-Ladder.

Conclusion

Ce mémoire propose une approche nouvelle pour le calcul des provisions de sinistres à payer. En effet, le choix est fait d'exploiter la notion de censure qui intervient dans les données et d'utiliser la théorie des modèles censurés pour évaluer les provisions de sinistres.

Pour mettre en œuvre une telle approche, il a été nécessaire d'identifier la loi de distribution des sinistres à partir des sinistres clos et non clos et sur la base d'un modèle paramétrique. Ce mémoire qui a eu recours au modèle Kaplan-Meier a permis, à partir de distribution empirique, d'établir une loi paramétrique correspondant au coût de sinistre (c'est-à-dire à partir d'une observation empirique d'en déduire la distribution continue correspondante) : dans un premier temps au coût de sinistre puis à la provision de sinistre.

Les modèles paramétriques de Pareto ou de Weibull n'ont pas fourni de résultats satisfaisants sur notre périmètre. Seul le modèle polynomial a permis d'approximer au mieux la distribution paramétrique du modèle.

Les résultats obtenus par ce modèle paraissent cohérents au regard de l'estimation des provisions de sinistres obtenus ex-post ou à partir d'une méthode plus classique telle que la modèle Chain-Ladder.

En d'autres termes, si l'approche d'un modèle de censure sur un échantillon statistique IARD fortement volatil et de grande amplitude apporte une réponse a priori assez bien adaptée à ce cas de figure, elle met en évidence plusieurs difficultés à résoudre :

- elle requiert en effet un échantillon statistique suffisant a priori difficilement compatible avec l'échantillon statistique considéré (des sinistres volatils de grande amplitude);

- le modèle de censure n'introduit pas par ailleurs de notion de durée de liquidation ce qui a conduit à retraiter dans l'échantillon considéré l'inflation.

Même si il est possible de compléter l'étude par des modèles d'inflation, il est vraisemblable que la combinaison de ces modèles ne donne pas des résultats satisfaisants dans une optique purement comptable. En effet, dans le cas des branches de responsabilité civile et en particulier garantie décennale, la dérive des coûts liée à l'inflation est un facteur pesant sur le dispositif.

L'application d'un tel modèle pourrait constituer une continuité intéressante à l'approche qui a été développée.

Bibliographie

PARTRAT C., LECOEUR E., NESSI J.M., NISIPASU E., REIZ O. [2007] *Provisionnement technique en assurance non-vie , Perspectives actuarielles modernes*, Paris : Economica

PLANCHET F., THEROND P. [2006] *Modèles de durée : applications actuarielles*, Paris : Economica

DENUIT M., CHARPENTIER A. [2005] *Mathématiques d l'assurance non-vie*, Paris : Economica

REGAZZONI Y., SANDER J. [1997] « Les provisions techniques : une approche par simulation », *Bulletin Français d'Actuariat*, Vol. 1, n°2, pp. 81-95.

ROEHNER B., WINIWARTER P. [1985] « Aggregation of independent Paretian random variables », *Advances in applied probability*, vol. 17, no2, pp. 465-469.

WEIBULL W. [1951] « A Statistical Distribution Function of Wide applicability ». *Journal of Applied Mechanic*, 18, pp. 292-297.

Annexes

A. Code des Assurances[13] :

- ✓ Article R 331 – 6 :

« Les provisions techniques correspondant aux autres opérations d'assurance sont les suivantes : […]

4° Provision pour sinistres à payer : valeur estimative des dépenses en principal et en frais, tant internes qu'externes, nécessaires au règlement de tous les sinistres survenus et non payés, y compris les capitaux constitutifs des rentes non encore mises à la charge de l'entreprise ; […] »

- ✓ Article R 331 – 15 :

« La provision pour sinistres à payer est calculée exercice par exercice.

Sans préjudice de l'application des règles spécifiques à certaines branches prévues à la présente section, l'évaluation des sinistres connus est effectuée dossier par dossier, le coût d'un dossier comprenant toutes les charges externes individualisables ; elle est augmentée d'une estimation du coût des sinistres survenus mais non déclarés.

La provision pour sinistres à payer doit toujours être calculée pour son montant brut, sans tenir compte des recours à exercer ; les recours à recevoir font l'objet d'une évaluation distincte.

Par dérogation […], l'entreprise peut, avec l'accord de l'Autorité de contrôle des assurances et des mutuelles, utiliser des méthodes statistiques pour l'estimation des sinistres survenus au cours des deux derniers exercices. »

[13] Code cité en Introduction

B. Programmation sous R :

Calcul de la provision individuelle:

```
c=read.table("encourstotal.txt")
c=c/10^3
a= 0.000007884943
b=-0.001068260157

L=as.vector(1:nrow(c))
e=10^-10

survie=function(t) {pmax(0,pmin(1,exp(-t*(a*t/2+b-a/2))))}
g=function(x){integrate(survie,lower=x,upper=Inf)$value}
L=apply(as.matrix(c[,1]),1,g)

P=L/pmax(survie(c[,1]),e)
P
write.table(P,"provisionajustementln1q.txt")
```

C. Calcul :

$$S(x)=\mathbf{exp}\left(-\left(\frac{ax^2}{2}+\left(b-\frac{a}{2}\right)x\right)\right)$$

q_x s'exprime à partir de la fonction de survie :

$$q_x=1-\frac{S(x+1)}{S(x)}$$

d'où

$$\begin{aligned}\mathbf{ln}(1-q_x)&=\mathbf{ln}\left(\frac{S_{x+1}}{S_x}\right)\\&=\mathbf{ln}(S_{x+1})-\mathbf{ln}(S_x)\\&=-ax-b\end{aligned}$$

D. Les facteurs d'incidence :

Facteurs d'incidence

Année de survenance	Année de paiement 1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
1983	289,55%	284,04%	275,42%	274,69%	275,57%	275,19%	273,65%	275,14%	275,86%	276,10%	277,98%	280,61%	282,95%
1984		273,03%	264,75%	264,05%	264,90%	264,53%	263,05%	264,48%	265,18%	265,40%	267,21%	269,74%	271,99%
1985			254,50%	253,82%	254,63%	254,28%	252,86%	254,24%	254,90%	255,12%	256,86%	259,29%	261,45%
1986				243,99%	244,77%	244,43%	243,06%	244,39%	245,03%	245,24%	246,91%	249,24%	251,32%
1987					235,29%	234,96%	233,65%	234,92%	235,54%	235,74%	237,34%	239,59%	241,59%
1988						225,86%	224,59%	225,82%	226,41%	226,60%	228,15%	230,31%	232,23%
1989							215,89%	217,07%	217,64%	217,83%	219,31%	221,39%	223,23%
1990								208,66%	209,21%	209,39%	210,81%	212,81%	214,58%
1991									201,11%	201,28%	202,65%	204,56%	206,27%
1992										193,48%	194,80%	196,64%	198,28%
1993											187,25%	189,02%	190,60%
1994												181,70%	183,22%
1995													176,12%
1996													
1997													
1998													
1999													
2000													
2001													
2002													
2003													
2004													
2005													
2006													
2007													
2008													

Année de survenance	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
1983	288,61%	291,89%	295,61%	302,53%	304,17%	303,09%	303,33%	303,13%	298,60%	297,46%	292,50%	287,11%	279,33%
1984	277,43%	280,59%	284,16%	290,81%	292,39%	291,35%	291,58%	291,39%	287,03%	285,94%	281,17%	275,99%	268,51%
1985	266,68%	269,72%	273,15%	279,54%	281,06%	280,06%	280,28%	280,10%	275,91%	274,86%	270,28%	265,30%	258,11%
1986	256,35%	259,27%	262,57%	268,71%	270,17%	269,21%	269,43%	269,25%	265,23%	264,21%	259,81%	255,02%	248,11%
1987	246,42%	249,22%	252,40%	258,30%	259,71%	258,78%	258,99%	258,82%	254,95%	253,98%	249,74%	245,14%	238,50%
1988	236,88%	239,57%	242,62%	248,30%	249,65%	248,76%	248,96%	248,79%	245,07%	244,14%	240,07%	235,65%	229,26%
1989	227,70%	230,29%	233,22%	238,68%	239,97%	239,12%	239,31%	239,15%	235,58%	234,68%	230,77%	226,52%	220,38%
1990	218,88%	221,37%	224,18%	229,43%	230,68%	229,86%	230,04%	229,89%	226,45%	225,59%	221,83%	217,74%	211,84%
1991	210,40%	212,79%	215,50%	220,55%	221,74%	220,95%	221,13%	220,98%	217,68%	216,85%	213,24%	209,31%	203,64%
1992	202,25%	204,55%	207,15%	212,00%	213,15%	212,39%	212,56%	212,42%	209,25%	208,45%	204,98%	201,20%	195,75%
1993	194,41%	196,63%	199,13%	203,79%	204,90%	204,17%	204,33%	204,19%	201,14%	200,37%	197,03%	193,41%	188,16%
1994	186,88%	189,01%	191,41%	195,89%	196,96%	196,26%	196,41%	196,28%	193,35%	192,61%	189,40%	185,91%	180,88%
1995	179,64%	181,69%	184,00%	188,31%	189,33%	188,65%	188,81%	188,68%	185,86%	185,15%	182,06%	178,71%	173,87%
1996	172,68%	174,65%	176,87%	181,01%	181,99%	181,35%	181,49%	181,37%	178,66%	177,98%	175,01%	171,79%	167,13%
1997		167,88%	170,02%	174,00%	174,94%	174,32%	174,46%	174,35%	171,74%	171,08%	168,23%	165,13%	160,66%
1998			163,43%	167,26%	168,17%	167,57%	167,70%	167,59%	165,09%	164,46%	161,71%	158,74%	154,43%
1999				160,78%	161,65%	161,08%	161,21%	161,10%	158,69%	158,08%	155,45%	152,59%	148,45%
2000					155,39%	154,84%	154,96%	154,86%	152,54%	151,96%	149,43%	146,68%	142,70%
2001						148,84%	148,96%	148,86%	146,63%	146,07%	143,64%	140,99%	137,17%
2002							143,19%	143,09%	140,95%	140,41%	138,08%	135,53%	131,86%
2003								137,55%	135,49%	134,98%	132,73%	130,28%	126,75%
2004									130,25%	129,75%	127,58%	125,23%	121,84%
2005										124,72%	122,64%	120,38%	117,12%
2006											117,89%	115,72%	112,58%
2007												111,24%	108,22%
2008													104,03%

E. Les triangles de liquidation :

Extrait de la base de données qui initie les triangles de liquidation

Étiquettes de lignes	1984	1985	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	Exercice de survenance	Cloture du sinistre	Cloture - Survenance
76D0302867	-	-	-	-	-	-	-	-	-	1 859	2 988	8 222	8 222	20 052	1994		
24D0301571	-	-	-	-	-	-	-	-	1 388	4 123	6 245	12 271	15 096	20 111	1993		
90D9903203	-	-	-	-	-	5 028	8 586	11 354	17 733	18 804	18 804	18 804	20 616	20 616	1989		
15D0700034	-	-	-	-	-	-	-	-	-	-	-	-	5 661	20 642	2004		
90D9102974	-	-	771 761	353 056	20 684	20 684	20 684	20 684	20 684	20 684	20 684	20 684	20 684	20 684	1991	1999	8
24D0401615	-	-	-	-	-	-	-	-	-	1 100	7 757	9 104	10 509	20 728	1993		
64D0800362	-	-	-	-	-	-	-	-	-	-	-	-	-	21 349	2008		
91D0301047	-	-	-	-	-	-	-	-	1 799	2 834	3 653	6 302	9 939	21 448	1994		
17D9900274	-	-	-	-	1 709	1 709	3 695	6 639	11 386	19 195	23 926	21 481	21 481	21 481	1999	2006	7
23D0305265	-	-	-	-	-	-	-	-	-	2 245	3 784	5 297	21 622	21 622	1984	2007	23
22D0503222	-	-	-	-	-	-	-	-	-	-	-	6 160	16 767	21 633	2001		
90D9702923	-	-	3 236	5 929	10 644	11 537	15 538	20 070	21 759	21 759	21 759	21 759	21 759	21 759	1992		
91D9401310	-	-	21 796	21 796	21 796	21 796	21 796	21 796	21 796	21 796	21 796	21 796	21 796	21 796	1994	1997	3
23D0203228	-	-	-	-	-	-	-	990	9 756	18 097	19 387	20 229	20 229	22 023	2001		
43D0200689	-	-	-	-	-	-	-	-	-	-	-	22 193	22 193	22 193	1988	2007	19
91D0701807	-	-	-	-	-	-	-	-	-	-	-	-	555	22 337	2004		
76D0501778	-	-	-	-	-	-	-	-	-	-	-	16 921	22 469	22 469	1999		
90D0002159	-	-	-	-	-	-	1 252	5 120	5 120	5 120	10 882	19 872	21 246	22 544	1994		
95D0000044	-	-	-	-	-	2 401	7 413	10 256	11 686	11 686	12 411	12 411	22 775	22 775	2000		
29D9501177	-	-	19 145	156 661	175 080	578 798	578 798	578 990	578 990	589 396	31 082	31 082	31 082	23 297	1994		
90D9402085	-	-	-	-	1 762	1 762	1 762	1 762	7 538	10 451	16 774	19 589	21 981	23 372	1989		
64D0704709	-	-	-	-	-	-	-	-	-	-	-	-	-	23 550	2006		
24D0101789	-	-	-	-	-	-	1 442	608 026	23 569	23 569	23 569	23 569	23 569	23 569	2001	2003	2
25D0400655	-	-	-	-	-	-	-	-	-	-	6 716	15 624	21 172	23 625	1998		
39D0000220	-	-	-	-	-	160 890	894 481	23 782	23 782	23 782	23 782	23 782	23 782	23 782	2000	2002	2
77D0700316	-	-	-	-	-	-	-	-	-	-	-	-	9 643	24 004	2007		
90D0502446	-	-	-	-	-	-	-	-	-	-	5 503	18 109	24 124	24 124	1997		
90D9900578	-	-	-	-	2 663	9 469	10 329	12 061	12 725	23 136	23 136	24 203	24 203	24 203	1995		
92D9300790	-	-	18 330	19 213	19 213	19 213	19 213	19 213	24 484	24 484	24 484	24 484	24 484	24 484	1993	2005	12
23D9802594	-	-	-	1 207	2 983	6 748	12 170	12 958	13 846	18 426	19 403	19 403	22 407	24 531	1998		
24D9900525	-	-	-	-	3 553	3 553	5 336	6 668	6 668	13 771	13 771	20 025	20 025	24 751	1988		
21D9900688	-	-	-	-	2 524	7 046	7 046	8 684	9 586	11 385	13 252	23 654	23 654	24 772	1999		
76D0401837	-	-	-	-	-	-	-	-	-	-	11 897	22 882	22 944	24 786	1993		
90D9000912	-	-	16 377	16 377	19 436	22 930	25 030	25 030	25 030	25 030	25 030	25 030	25 030	25 030	1983	2005	22
23D8900913	-	-	981	1 985	1 985	1 985	1 985	3 680	18 386	18 386	25 300	25 300	25 300	25 300	1985		
21D9902875	-	-	-	-	-	7 184	10 244	14 612	17 135	23 802	25 477	25 477	25 477	25 477	1997		
90D9700701	-	-	936	936	7 977	9 919	13 882	13 882	19 089	19 089	19 089	25 921	25 921	25 921	1990	2007	17
16D9900826	-	-	-	-	-	5	878	878	878	878	878	878	878	26 517	1998		
90D9301201	-	-	-	-	-	-	-	-	18 044	26 859	26 859	26 859	26 859	26 859	1990	2006	16
90D0202478	-	-	-	-	-	-	-	1 603	7 871	13 459	19 328	22 782	27 242	26 870	2002		

Le triangle de liquidation de la provision

Année de survenance	Année de développement																									
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1983		3659,91879	4960,54211	17600,6592	23176,7632	601363,994	611958,64	619154,81	638311,202	666259,548	1064497,17	2998176,86	4186417,58	4919403,47	5049591,43	5167318,43	6202579,27	7355866,67	6209575,39	6467909,74	8144736,25	8147983,86	8665513,73	11149398,6	11067940,4	11108442
1984	0	0	6375,27769	26203,8339	136222,751	167571,92	2576645,88	3953445,78	5506233,56	5559222,85	5836536,13	5864200,04	6967564,42	6438466,21	10239599,6	10290186,7	15043199,8	18344930,7	18585164	18644920,8	19915434,4	22232568,1	23947136,1	24573995,7	24898032	
1985	0	1600674,03	1713229,31	1829453,56	1995512,37	2169038,03	3763199,98	4409956,36	4543567,81	8471628,19	7452915,02	9148132,43	10250364,9	9979787,86	12513053,1	14382771,4	15298712,6	17549035,7	17574135,7	17186684,9	17180438,4	17192851,5	16775108,8	16808286,3		
1986	0	14169,4943	1884503,29	2501098,83	3106622,52	4745372,9	4022987,68	5686956,13	8672867,56	9201372,01	12238590,6	15637675,9	15859082,4	15459008,7	16681611,4	18890282,5	21841361,5	21996641,5	28662664,6	30321105,8	31221824,6	31030240,7	31150252			
1987	0	31655,4022	275003,016	813113,171	1762175,98	2542128,3	3830461,36	6173847,95	9212193,52	9967936,78	11869484,2	14872230,7	18283855	17721740,7	17612737,3	21184806,8	25512065,5	23597402,7	32154098,2	32570825,6	33324281,7	35090337				
1988	0	46356,3022	568676,623	1851460,92	4117097,42	5885504,44	7825890,98	11484850,9	16518648,3	19692395,8	21360848,9	29905650,2	40484269,7	40946000,1	45630258,4	45045777,6	46545805,1	51010644,6	52548528,3	56575709	58761552,9					
1989	4965,98566	266797,25	984126,664	2056737,18	5126763,23	7762688,92	8966450,19	10562669,4	11604228,9	21027965	19542805,2	26783346,4	27403099,3	39644309	40531274,2	51374235,6	52089989,7	52466296,4	52880661,2	53764505,2						
1990	7094,60087	126008,992	2323487,96	3017931,77	5266920,66	8823918,98	12198004,6	13798510,5	23263675,6	25817505,2	29158994,8	40085752,1	50631260,6	57573249	61657530,1	64432884,6	68602606,3	74394096,8	86441862,9							
1991	26239,7237	639255,054	2443362,45	5441063,27	12905843,8	17189646,8	19670850,7	20253607,3	20936495,9	24029053,5	26827769,4	28044902	30490599,6	33704070,2	40527683,4	57540641,5	56710806,9	64782092,4								
1992	867294,199	2423865,72	4531840,32	4999471,32	6184092,1	8432858,17	11153164	11502154,6	14125264,3	16322449,8	17481779,8	18594449,7	19573278,8	21179629,9	21179082,3	21284379,7	23999693,7									
1993	36529,4975	769976,886	1619571,91	2995317,31	4376836,38	4785322,45	5547673,41	6843314,72	9175616,23	12173230,7	20205398	25723500,4	25137615,5	28557410	28759179,3	29744773,3										
1994	67890,1236	5014229,88	5816274,92	6242709,79	8376178,7	10317799,9	10982511,5	13737873,9	15415304,1	17441071	20737324,3	25757075,5	22589810,6	25919031,5	28304334,8											
1995	44073,4546	425712,695	2451333,03	3346444,11	5548175,71	6206512,83	7301143,88	11325535,4	12635952,3	18925355,5	21366305,3	22006657,2	26057251	27674856,8												
1996	41473,6725	115906,96	8326651,1	32518225,7	35751050,5	36182682	37260037,4	40310163,2	41951375,2	42883882,6	42471805,7	43222598,7	45585815,8													
1997	54550,9824	521666,574	1755508,62	2599573,63	4098670,99	5068491,02	12887631,1	10423573	14606140,8	14097511,6	15343267,8	16350302,1														
1998	34553,7924	219177,933	2095017,95	2263540,85	2648225,69	5041053,35	5508219,3	8519855,26	9338792,19	11216625,5	23186684,2															
1999	1365346,7	1714851,53	2451481,67	6957156,79	7195340,61	7472684,95	8672562,78	19821138,9	18394470,9	24792444,6																
2000	329635,265	3983610,76	5036434,18	5361496,6	4319871,22	3809210,84	4747630,58	4967763,85	6158027,83																	
2001	509704,543	2469290,91	3680530,56	3021635,86	3948428,39	4537115,54	7145707,07	13993091,9																		
2002	58585,7789	975766,689	2110371,45	2910987,83	5306545,46	7441641,72	12799328,8																			
2003	858642,071	1764610,82	4880689,11	4920527,52	5639676,26	8380743,12																				
2004	416962,846	279898,178	507089,384	1269739,18	1554221,01																					
2005	27036,3018	205089,495	524212,291	685280,441																						
2006	16197,6624	384098,832	992767,83																							
2007	100101,815	570679,05																								
2008	23437,6261																									

La liquidation des coûts totaux estimés et la provision

Année de survenance	Année de développement 0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	Provision
1983		3 660	4 961	17 601	23 177	601 364	611 959	619 155	638 311	666 260	1 064 497	2 998 177	4 186 418	4 919 403	5 049 591	5 167 318	6 202 579	7 355 867	6 209 575	6 467 910	8 144 736	8 147 984	8 665 514	11 149 399	11 067 940	11 108 442	
1984	-	-	6 375	26 204	136 223	167 572	2 576 646	3 953 446	5 506 234	5 559 223	5 836 536	5 864 200	6 967 564	6 438 466	10 239 600	10 290 187	15 043 200	18 344 931	18 585 164	18 644 921	19 915 434	22 232 568	23 947 136	24 573 996	24 898 032	24 989 143	91 111
1985	-	1 600 674	1 713 229	1 829 454	1 995 512	2 169 038	3 763 200	4 409 956	4 543 568	8 471 628	7 452 915	9 148 132	10 250 365	9 979 788	12 513 053	14 382 771	15 298 713	17 549 036	17 574 136	17 186 685	17 180 438	17 192 852	16 775 109	16 808 286	16 909 136	16 971 013	162 726
1986	-	14 169	1 884 503	2 501 099	3 106 623	4 745 373	4 022 988	5 686 956	8 672 868	9 201 372	12 238 591	15 637 676	15 859 082	15 459 009	16 681 611	18 890 283	21 841 361	21 996 642	28 662 665	30 321 106	31 221 825	31 030 241	31 150 252	31 648 656	31 838 548	31 955 057	804 805
1987	-	31 655	275 003	813 113	1 762 176	2 542 128	3 830 461	6 173 848	9 212 194	9 967 937	11 869 484	14 872 231	18 283 855	17 721 741	17 612 737	21 184 807	25 512 065	23 597 403	32 154 098	32 570 826	33 324 282	35 090 337	35 932 505	36 507 425	36 726 470	36 860 865	1 770 528
1988	-	46 356	568 677	1 851 461	4 117 097	5 885 504	7 825 891	11 484 851	16 518 648	19 692 396	21 360 849	29 905 650	40 484 270	40 946 000	45 630 258	45 045 778	46 545 805	51 010 645	52 548 528	56 575 709	58 761 553	60 818 207	62 277 844	63 274 290	63 653 935	63 886 868	5 125 315
1989	4 966	266 797	984 127	2 056 737	5 126 763	7 762 689	8 966 450	10 562 669	11 604 229	21 027 965	19 542 805	26 783 346	27 403 099	39 644 309	40 531 274	51 374 236	52 089 990	52 466 296	52 880 661	53 764 505	55 968 850	57 927 760	59 318 026	60 267 114	60 628 717	60 850 579	7 086 074
1990	7 095	126 009	2 323 488	3 017 932	5 266 921	8 823 919	12 198 005	13 798 511	23 263 676	25 817 505	29 158 995	40 085 752	50 631 261	57 573 249	61 657 530	64 432 885	68 602 606	74 394 097	86 441 863	89 294 444	92 955 517	96 208 960	98 517 975	100 094 262	100 694 828	101 063 306	14 621 443
1991	26 240	639 255	2 443 362	5 441 063	12 905 844	17 189 647	19 670 851	20 253 607	20 936 496	24 029 054	26 827 769	28 044 902	30 490 600	33 704 070	40 527 683	57 540 641	56 710 807	64 782 092	68 604 236	70 868 176	73 773 771	76 355 853	78 188 393	79 439 408	79 916 044	80 208 485	15 426 393
1992	867 294	2 423 866	4 531 840	4 999 471	6 184 092	8 432 858	11 153 164	11 502 155	14 125 264	16 322 450	17 481 780	18 594 450	19 573 279	21 179 630	21 179 082	21 284 380	23 999 694	25 607 673	27 118 526	28 013 437	29 161 988	30 182 658	30 907 041	31 401 554	31 589 963	31 705 562	7 705 869
1993	36 529	769 977	1 619 572	2 995 317	4 376 836	4 785 322	5 547 673	6 843 315	9 175 616	12 173 231	20 205 398	25 723 500	25 137 616	28 557 410	28 759 179	29 744 773	31 882 825	34 018 974	36 026 093	37 214 955	38 740 768	40 096 695	41 059 015	41 715 959	41 966 255	42 119 825	12 375 051
1994	67 890	5 014 230	5 816 275	6 242 710	8 376 179	10 317 800	10 982 512	13 737 874	15 415 304	17 441 071	20 737 324	25 757 076	22 589 811	25 919 032	28 304 335	30 596 986	32 796 294	34 993 646	37 058 271	38 281 194	39 850 723	41 245 498	42 235 390	42 911 157	43 168 623	43 326 593	15 022 258
1995	44 073	425 713	2 451 333	3 346 444	5 548 176	6 206 513	7 301 144	11 325 535	12 635 952	18 925 356	21 366 305	22 006 657	26 057 251	27 674 857	30 116 120	32 555 525	34 895 613	37 233 619	39 430 403	40 731 606	42 401 602	43 885 658	44 938 914	45 657 936	45 931 884	46 099 965	18 425 108
1996	41 474	115 907	8 326 651	32 518 226	35 751 050	36 182 682	37 260 037	40 310 163	41 951 375	42 883 883	42 471 806	43 222 599	45 585 816	50 452 121	54 902 618	59 349 730	63 615 783	67 878 041	71 882 845	74 254 979	77 299 433	80 004 913	81 925 031	83 235 832	83 735 247	84 041 664	38 455 848
1997	54 551	521 667	1 755 509	2 599 574	4 098 671	5 068 491	12 887 631	10 423 573	14 606 141	14 097 512	15 343 268	16 350 302	18 196 780	20 139 294	21 915 828	23 691 010	25 393 917	27 095 310	28 693 933	29 640 833	30 856 107	31 936 071	32 702 537	33 225 777	33 425 132	33 547 446	17 197 144
1998	34 554	219 178	2 095 018	2 263 541	2 648 226	5 041 053	5 508 219	8 519 855	9 338 792	11 216 625	23 186 684	27 592 154	30 708 202	33 986 315	36 984 326	39 980 056	42 853 819	45 725 025	48 422 801	50 020 754	52 071 604	53 894 111	55 187 569	56 070 570	56 406 994	56 613 407	33 426 723
1999	1 365 347	1 714 852	2 451 482	6 957 157	7 195 341	7 472 685	8 672 563	19 821 139	18 394 471	24 792 445	28 536 104	33 957 964	37 792 917	41 827 326	45 517 011	49 203 889	52 740 660	56 274 284	59 594 467	61 561 084	64 085 088	66 328 067	67 919 940	69 006 659	69 420 699	69 674 734	44 882 290
2000	329 635	3 983 611	5 036 434	5 361 497	4 319 871	3 809 211	4 747 631	4 967 764	6 158 028	7 348 993	8 458 691	10 065 842	11 202 602	12 398 484	13 492 183	14 585 050	15 633 422	16 680 861	17 665 032	18 247 978	18 996 145	19 661 010	20 132 874	20 455 000	20 577 730	20 653 031	14 495 004
2001	509 705	2 469 291	3 680 531	3 021 636	3 948 428	4 537 116	7 145 707	13 993 092	16 609 800	19 822 143	22 815 286	27 150 190	30 216 326	33 441 931	36 391 921	39 339 666	42 167 398	44 992 613	47 647 178	49 219 535	51 237 535	53 030 849	54 303 590	55 172 447	55 503 482	55 706 589	41 713 497
2002	58 586	975 767	2 110 371	2 910 988	5 306 545	7 441 642	12 799 329	15 819 970	18 778 305	22 410 037	25 793 953	30 694 804	34 161 241	37 807 967	41 143 095	44 475 686	47 672 594	50 866 658	53 867 790	55 645 428	57 926 890	59 954 331	61 393 235	62 375 527	62 749 780	62 979 404	50 180 075
2003	858 642	1 764 611	4 880 689	4 920 528	5 639 676	8 380 743	10 531 758	13 017 253	15 451 480	18 439 803	21 224 213	25 256 813	28 109 125	31 109 786	33 854 052	36 596 230	39 226 763	41 854 956	44 324 399	45 787 104	47 664 375	49 332 628	50 516 611	51 324 877	51 632 827	51 821 769	43 441 026
2004	416 963	279 898	507 089	1 269 739	1 554 221	1 914 800	2 406 256	2 974 133	3 530 295	4 213 056	4 849 227	5 770 581	6 422 266	7 107 845	7 734 845	8 361 367	8 962 382	9 562 861	10 127 070	10 461 263	10 890 175	11 271 331	11 541 843	11 726 513	11 796 872	11 840 041	10 285 820
2005	27 036	205 089	524 212	685 280	914 454	1 126 607	1 415 764	1 749 885	2 077 113	2 478 828	2 853 131	3 395 225	3 778 656	4 182 029	4 550 936	4 919 562	5 273 179	5 626 482	5 958 445	6 155 073	6 407 431	6 631 691	6 790 852	6 899 506	6 940 903	6 966 302	6 281 021
2006	16 198	384 099	992 768	1 731 368	2 310 378	2 846 385	3 576 943	4 421 102	5 247 848	6 262 784	7 208 464	8 578 072	9 546 814	10 565 941	11 497 987	12 429 324	13 322 742	14 215 366	15 054 073	15 550 857	16 188 442	16 755 038	17 157 159	17 431 673	17 536 263	17 600 435	16 607 667
2007	100 102	570 679	1 355 363	2 363 727	3 154 212	3 885 989	4 883 373	6 035 849	7 164 553	8 550 180	9 841 258	11 711 096	13 033 658	14 425 007	15 697 470	16 968 965	18 188 692	19 407 335	20 552 368	21 230 596	22 101 050	22 874 587	23 423 577	23 798 354	23 941 144	24 028 753	23 458 074
2008	23 438	85 571	203 231	354 430	472 960	582 687	732 240	905 049	1 074 293	1 282 061	1 475 653	1 756 027	1 954 339	2 162 965	2 353 765	2 544 420	2 727 313	2 910 043	3 081 736	3 183 433	3 313 954	3 429 942	3 512 261	3 568 457	3 589 868	3 603 004	3 579 567
Facteurs de développeme		3,651	2,375	1,744	1,334	1,232	1,257	1,236	1,187	1,193	1,151	1,190	1,113	1,107	1,088	1,081	1,072	1,067	1,059	1,033	1,041	1,035	1,024	1,016	1,006	1,004	442 620 436