

Résumé

Mots clés : assurance non vie, modèle linéaire généralisé, temps opérationnel, risque de provisionnement, surinflation.

Le développement d'un modèle interne et la mise en œuvre de l'ORSA conduisent de plus en plus d'assureurs à effectuer un calcul stochastique de leurs provisions.

L'objet de ce mémoire est de présenter un modèle stochastique de provisionnement construit dans le cadre des modèles linéaires généralisés. On peut déduire du modèle l'évaluation de la meilleure estimation des provisions Best Estimate ou celle du besoin en capital au titre du risque de provisionnement.

Dans le calcul des provisions, nous montrons à partir de données en dommages corporels automobile, que le modèle étudié peut constituer une alternative aux méthodes par facteurs si les données ne respectent pas les hypothèses qui les sous tendent. En particulier, la structure souple du modèle permet de modéliser certains facteurs à l'origine des perturbations dans les cadences de paiements. Cela inclue l'instabilité des cadences de règlements ou la surinflation, deux phénomènes caractéristiques des branches à écoulement long.

Dans le calcul du risque de provisionnement, nous privilégions une approche simulateur dans un cadre Bootstrap. Nous présentons tout d'abord une version du Bootstrap adaptée au modèle, puis la modifions avec une vision sur un horizon d'un an. Le calcul du risque de provisionnement est également effectué par application de la formule standard du QIS4 pour comparaison.

Abstract

Key words : non-life insurance, generalized linear models, operational time, risk provisioning, superimposed inflation.

The development of an internal model and the implementation of the ORSA lead more and more insurers to do a stochastic calculus of their reserves.

The purpose of this paper is to present a stochastic model of provisioning built into the framework of generalized linear models. One can infer from the model assessing the best estimate of reserves or the required capital under the risk provisioning.

In the calculation of reserves, we show from data of Auto Bodily Injury Insurance, the model can be considered an alternative to the methods by factors if the data fail to meet the assumptions that underlie them. In particular, the flexible structure model is used to model some factors causing disturbances in the rates of payments. This includes the volatility of rates of settlements or superimposed inflation, two characteristic phenomena of long branches.

In computing the risk provisioning, we favour a simulation approach. We adapt the classical version of the Bootstrap to the model and modify it with a vision on a one-year horizon. The calculation of risk provisioning is done by applying the standard formula of QIS4 for comparison.

Remerciements

Je souhaitais ici remercier les personnes qui ont contribué directement ou non à la réalisation de ce mémoire :

Un grand merci à Stéphanie, ma femme, pour m'avoir supporté dans cette démarche.

Merci à Vincent Dupriez pour sa relecture et ses conseils expérimentés.

Merci enfin à tous mes professeurs du CNAM qui ont su par leur enseignement me persuader que l'actuariat est une matière riche et passionnante.

Sommaire

INTRODUCTION.....	7
PARTIE I. SOLVABILITE II : PRINCIPES DE CALCUL DU BEST ESTIMATE ET DU RISQUE DE PROVISION.....	8
Chapitre I. Le best estimate des provisions pour sinistres	9
I. Le périmètre de calcul.....	9
II. Les principes d'évaluation	9
II.1 Aspects méthodologiques.....	9
II.2 Aspects calculatoires	11
II.2.1 Les préconisations du Groupe Consultatif	11
II.2.2 Les préconisations de l'ACAM	11
III. L'expression de la provision pour sinistres	12
Chapitre II. le Risque de provisionnement	14
I. Les fondements du risque	14
I.1 L'incertitude des PSAP : une analyse par risques	14
I.2 De l'incertitude des PSAP au risque de provisionnement	15
I.2.1 L'exposition des branches à déroulement long.....	15
I.2.2 Le cas des dommages corporels en automobile.....	16
II. Le risque de provisionnement de Solvabilité II	17
II.1 La définition du risque.....	17
II.2 L'identification de la variable d'intérêt	18
II.3 Risque à un an contre risque ultime.....	19
III. Mesures de risques	20
III.1 L'indicateur du MSE : une mesure du risque à l'ultime	20
III.2 La mesure du risque à un an	21
III.2.1 Solvabilité II : La mesure par la VaR.....	21
III.2.2 Méthodes approchées.....	23
PARTIE II. LA SELECTION DU MODELE	25
Chapitre I. etude des données	26
I. L'exposé du problème.....	26
I.1 Limites de l'usage des méthodes par facteurs	26
I.2 Quelques méthodes alternatives	27
I.3 Le cas des données étudiées	28
II. La discussion sur les données	29
II.1 La description des données.....	29
II.1.1 Les données utilisées	29
II.1.2 L'estimation du nombre de sinistres survenus restant à déclarer.....	29
II.2 L'application de la méthode Chain Ladder	31
II.2.1 L'étude préalable des rapports historiques de développement.....	31
II.2.2 La sensibilité de la provision pour sinistres	32
II.3 De l'inadéquation des méthodes par facteurs	33
II.3.1 L'instabilité de la cadence des règlements.....	33
II.3.2 La liaison entre volume des prestations et nombre de règlements.....	34
II.3.3 Les coefficients de passage et le temps opérationnel.....	35
III. L'approche par les coûts moyens	36
III.1 Les coûts moyens et le temps opérationnel	36
III.2 La stationnarité des coûts moyens.....	39
III.3 La partition du temps opérationnel	40
III.4 Le lien moyenne - variance	42
IV. Les conséquences sur le choix du modèle	43
Chapitre II. la modelisation des couts moyens	44
I. La modélisation par les Moindres Carrés Ordinaires.....	44
I.1 Le modèle de régression.....	44
I.2 Les tendances élémentaires	45
I.3 Le terme stochastique	46

II.	La modélisation par les Modèles Linéaires Généralisés	47
II.1	GLM contre MCO	47
II.2	Introduction aux GLM	48
II.2.1	La famille des distributions exponentielles	48
II.2.2	L'estimation des paramètres	50
II.2.3	La déviance	52
II.2.4	Les résidus	53
II.3	Les hypothèses du modèle GLM	54
II.3.1	Le choix de la distribution des résidus	54
II.3.2	La fonction de lien	57
Chapitre III. la spécification du modele.....		58
I.	La capture des tendances	58
I.1	Le rappel du principe de parcimonie	58
I.2	Le modèle de séparation	58
I.3	La recherche des tendances du triangle	59
II.	La modélisation des tendances	62
II.1	Le temps opérationnel	62
II.2	La surinflation	64
II.3	L'interaction entre le temps opérationnel et la surinflation	66
III.	La modélisation des paiements partiels	69
III.1	L'exposé du problème	69
III.2	L'approche de Wright	70
III.3	L'application numérique	71
IV.	La sélection du modèle final	74
Chapitre IV. validation du modele par des tests statistiques		76
I.	Les points de validation	76
II.	Les tests sur la spécification du modèle	76
III.	Le test de robustesse	79
IV.	Le test sur la validité de la déviance	81
V.	Conclusion : la validation du modèle	82
PARTIE III. LA MISE EN ŒUVRE DES CALCULS		84
Chapitre I. le calcul du best estimate des pSAP		85
I.	Hypothèses et méthode	85
I.1	Hypothèses	85
I.2	Les étapes de calcul	86
II.	La projection en monnaie constante des prestations futures	88
II.1	La modélisation des prestations	88
II.2	L'application numérique	90
III.	L'inflation future	92
III.1	Le modèle d'inflation	92
III.2	Le calibrage des paramètres	94
III.3	La validation du modèle par back testing	96
III.4	La modélisation de l'inflation dans les prestations futures	99
IV.	L'évaluation du Best Estimate des PSAP	100
Chapitre II. le calcul du risque de provisionnement.....		103
I.	L'exposé du problème	103
II.	Le calcul du risque à l'ultime	104
II.1	La méthode du Bootstrap	104
II.1.1	Notations	104
II.1.2	La sélection des résidus	104
II.1.3	L'implémentation du Bootstrap	105
II.2	L'application numérique	107
III.	Le calcul du risque à un an	109
III.1	La modification de la méthode du Bootstrap	109
III.1.1	Principes	109
III.1.2	Algorithme	110
III.1.3	L'application numérique	112

III.2	Le calcul par la formule standard du QIS4.....	119
III.2.1	Les principes de calcul.....	119
III.2.2	L'application numérique	120
III.3	Synthèse des résultats	121
<i>BIBLIOGRAPHIE.....</i>		<i>123</i>
<i>ANNEXES.....</i>		<i>126</i>
A.	Triangle des incréments	126
B.	Triangle des nombres de sinistres réglés.....	127
C.	Triangle des nombres de sinistres déclarés et restant à déclarer	128
D.	Triangle des nombres de sinistres en cours.....	130
E.	Triangle des cadences de règlements	131
F.	Triangle des temps opérationnels.....	132
G.	Triangle des coûts moyens par sinistre réglé	133
H.	L'algorithme de Newton Raphson	134
I.	L'algorithme du score	135
J.	Expression de la déviance.....	137
K.	Evolution des coûts moyens en monnaie constante	138
L.	Rappel élémentaire de théorie du risque	139
M.	Modélisation des dates de règlements futurs.....	140
N.	Projection des règlements de sinistres.....	141
O.	Caractéristiques du processus d'inflation.....	142
P.	Equidistribution des résidus gamma de Pearson	144
Q.	Calcul du risque ultime : programmation de la procédure de rééchantillonnage	145
R.	Calcul du risque de provision : simulations des dates de règlements futurs	147
S.	Calcul du risque de provision : programmation de la régression à un an.....	148

INTRODUCTION

Le 22 Avril 2009, le Parlement européen votait la directive Solvabilité II. Cette réforme, qui place la gestion des risques au cœur du système prudentiel des assurances, se caractérise par de nouvelles exigences quantitatives visant à mieux identifier et mesurer les risques portés par les assureurs. Ces exigences incluent le calcul des capitaux propres et celui des provisions techniques, s'appuyant sur la notion de meilleure estimation.

Dans ces calculs, les nouveaux principes prudentiels distinguent en assurance non vie le cas des sinistres à survenir de celui des sinistres survenus en date de calcul. Dans ce mémoire, nous nous focalisons sur ce dernier cas. Dans le calcul des provisions, il s'agit alors d'évaluer la meilleure estimation des Provisions pour Sinistres. Et dans le calcul modulaire du capital de solvabilité, il est requis d'évaluer le risque de provisionnement, introduit par Solvabilité II dans le risque de souscription, qui constitue un des principaux risques de passif en assurance non-vie.

Dans le calcul des exigences quantitatives, la Commission Européenne laisse la possibilité aux assureurs d'utiliser une formule standard ou de recourir à un modèle interne. Ce mémoire présente un modèle de provisionnement en assurance non-vie, développé dans le cadre des Modèles Linéaires Généralisés et intégrable au sein d'un modèle interne. On peut en déduire la meilleure estimation des provisions pour sinistres ainsi que l'évaluation du risque de provisionnement.

L'approche adoptée consiste à modéliser le coût moyen des sinistres à partir de leur ordre de règlement, en utilisant les Modèles Linéaires Généralisés (GLM). La connaissance du nombre de sinistres survenus et restant à régler suffit alors à estimer le montant total des prestations futures, et par extension celui des provisions Best Estimate pour sinistres.

Les triangles des prestations et des nombre de sinistres réglés, nécessaires à la mise en œuvre du modèle, sont empruntés à Taylor (2000). Dans les calculs, nous utilisons et supposons certain le nombre estimé par Taylor des sinistres survenus et restant à régler. Dans le calcul du risque de provisionnement, cette simplification conduit à ne considérer que l'aléa sur les coûts unitaires des sinistres futurs. Cette restriction constitue une des limites de l'approche présentée, notamment dans le cas des dernières survenances où le nombre des sinistres futurs reste le plus incertain.

Le mémoire est structuré en trois parties. La première partie fournit un rappel des principes de calcul des provisions techniques et de l'exigence de capital au titre des sinistres survenus et restant à payer.

La seconde partie présente les caractéristiques du modèle de provisionnement, ses avantages par rapport aux méthodes par facteurs, et détaille sa mise en œuvre (choix du modèle, spécification, calibrage, ...).

La dernière partie constitue l'application numérique du mémoire. On y met en œuvre le calcul du Best Estimate des provisions pour sinistres ainsi que celle du risque de provisionnement.

PARTIE I. SOLVABILITE II : PRINCIPES DE CALCUL DU BEST ESTIMATE ET DU RISQUE DE PROVISION

Cette partie introduit les concepts et principes de calcul liés aux deux variables d'intérêt : le Best Estimate des provisions pour sinistres et le risque de provisionnement.

Le premier chapitre recense les points que nous jugeons importants de valider pour être conformes aux exigences de Solvabilité II dans le calcul du Best Estimate de la provision pour sinistres. Afin de conforter la validité de notre approche, nous présentons en parallèle les caractéristiques du modèle de provisionnement détaillés en partie II et III.

Le deuxième chapitre introduit les concepts liés au risque de provisionnement traité en partie III, et en présente deux méthodes approchées.

CHAPITRE I. LE BEST ESTIMATE DES PROVISIONS POUR SINISTRES

I. Le périmètre de calcul

L'article 76 de la directive européenne (Parlement Européen, 2009) indique que « la valeur des provisions techniques est égale à la somme de la meilleure estimation et de la marge de risque ». Les modalités d'évaluation de la marge de risque ne sont pas traitées dans ce mémoire qui se concentre sur le principe de la meilleure estimation (en anglais, Best Estimate).

Le calcul du Best Estimate en non-vie est segmenté selon deux niveaux :

- L'article 79 de la directive européenne précise que les provisions techniques doivent être calculées par groupes de risques homogènes. A cet effet, dix huit branches d'activité sont identifiées à l'annexe I de la directive, dont la branche « R.C. véhicules terrestres automoteurs » à laquelle se rattache le portefeuille étudié.
- Dans chaque branche, les spécifications techniques du QIS4 distinguent les sinistres survenus en date de calcul des sinistres à survenir. En ce sens, le Best Estimate est la somme des Best Estimate:
 - de la provision pour primes qui se substitue aux provisions pour primes non acquises et aux provisions pour risques en cours.
 - de la provision pour sinistres constituée au titre des sinistres survenus.

Ce mémoire se focalise donc sur cette dernière partie des provisions techniques. A ce sujet, notons que les spécifications du QIS4 (TS.II.E.8) préconisent de dissocier le calcul des provisions pour sinistres de celui des provisions pour primes.

II. Les principes d'évaluation

II.1 Aspects méthodologiques

Au centre des règles de calcul à respecter, le concept du Best Estimate organise le passage d'une évaluation prudente des provisions techniques assise sur des hypothèses réglementaires, à une évaluation économique plus réaliste. Plus précisément, l'article 76 définit le Best Estimate comme « la moyenne pondérée par leur probabilité des flux de trésorerie futurs, compte tenu de la valeur temporelle de l'argent [...], déterminée à partir de la courbe des taux sans risque pertinente ».

Comme il apparaît difficile de préconiser de façon systématique des méthodes de calcul en fonction de cas identifiés (duration, type de risque, ...), le Best Estimate est davantage un ensemble de principes qu'un guide méthodologique. Au lieu d'imposer des méthodes de calcul dans son évaluation, les principes réglementaires incitent plutôt l'assureur à s'inscrire dans une démarche méthodologique, dont certains aspects importants sont rappelés ci-après.

- **L'usage de méthodes reconnues**

Pour les branches de forte incertitude telle la branche étudiée, les spécifications techniques du QIS4 préconisent l'utilisation de méthodes avérées dites *best practice* en citant les méthodes construites à partir du triangle de liquidation (TS.II.E.10).

- ✓ A ce sujet le modèle étudié extrapole les prestations futures à partir du triangle de liquidation des nombres de règlements et du triangle des prestations payées.

- **La volonté de croiser différentes méthodes**

Les spécifications techniques du QIS4 conseillent d'estimer les PSAP à partir d'au moins « deux méthodes actuarielles considérées comme fiables conduisant à un faisceau convergent d'estimations. De ces méthodes, la plus appropriée, celle qui capture le mieux, de façon prudente, fiable et objective la nature du risque, est ensuite retenue. ». L'objectif d'une telle démarche consiste à contrôler l'absence d'erreur de modèle et de paramètres.

- ✓ Dans ce mémoire, nous nous consacrons à présenter une méthode compatible avec les exigences de Solvabilité II. Dans le cas de la branche corporelle automobile, le rapport de l'ACAM sur le Best Estimate en non-vie (ACAM, 2007) conçoit trois types de méthodes :
 - Les méthodes de cadences appliquées aux paiements, recours, nombres et charges de sinistres dossier par dossier.
 - Les méthodes d'exposition appliquées aux paiements, recours, nombres et charges de sinistres dossier par dossier.
 - L'approche basée sur des loss ratios ultimes fonction de la sinistralité déjà survenue et de la politique tarifaire.

- **La justification des méthodes utilisées**

La directive prévoit que les assureurs soient capables de justifier des méthodes de provisionnement utilisées. En ce sens, l'article 119 relatif aux normes de qualité statistique précise que « les entreprises d'assurance et de réassurance doivent être en mesure de justifier auprès des autorités de contrôle, les hypothèses qui sous-tendent leur modèle interne ».

- ✓ Sur ce point, les conclusions de l'étude des données présentée au début de la deuxième partie doivent permettre de justifier le recours au modèle présenté.

- **Une qualité minimale d'ajustement**

Les spécifications techniques du QIS4 conseillent la mise en œuvre de tests d'ajustements, de sorte à rejeter les méthodes conduisant à une qualité d'ajustement insuffisante (TS.II.E.13).

- ✓ A cet égard, le cadre des Modèles Linéaires Généralisés, dans lequel s'inscrit le modèle de provisionnement présenté, permet de procéder tant à des tests d'ajustements globaux qu'à des tests de significativité des variables explicatives.

II.2 Aspects calculatoires

Le croisement des principes réglementaires et des préconisations émises par l'ACAM ou le Groupe Consultatif Actuariel Européen¹ permet de préciser de façon suffisante les principes de calcul à respecter dans le calcul du Best Estimate en non-vie.

II.2.1 Les préconisations du Groupe Consultatif

Dans leur rapport intérimaire sur le Best Estimate en assurance non-vie, le Groupe Consultatif Actuariel Européen (2008) retranscrit l'article 76 de la directive en cinq points :

- La meilleure estimation correspond à la moyenne actualisée (et non à la médiane).
- L'estimation doit être sans biais et ne pas inclure de marge de prudence.
- Les paramètres utilisés combinent l'expérience de l'entreprise et du marché.
- Les flux futurs doivent refléter l'expérience future de l'entreprise.
- Les flux futurs doivent être actualisés à partir de la courbe des taux sans risque.

II.2.2 Les préconisations de l'ACAM

A l'initiative du Groupe Consultatif, des réflexions ont été menées au niveau national autour des méthodes approchées du calcul du Best Estimate en non-vie. En France, un groupe de travail est constitué sous la présidence de l'ACAM. En dehors des méthodes approchées, son rapport (ACAM, 2007) rappelle les orientations à suivre dans le calcul du Best Estimate.

- **Méthodes de calcul**

Le rapport précise que les modèles utilisés peuvent être :

- Soit stochastiques, impliquant l'estimation de la distribution des flux futurs, dont l'espérance actualisée définit la meilleure estimation des PSAP.
- Soit déterministes, auquel cas l'évaluation de la charge ultime moyenne après application de cadences et de la courbe des taux, conduit à la meilleure estimation.

- **Le retraitement des sinistres exceptionnels**

Dans le cas de la branche corporelle automobile, le rapport préconise de segmenter l'analyse par tranches de coût des sinistres. En particulier, il s'agit de séparer les sinistres hors graves des sinistres graves qui peuvent distordre les observations moyennes de coûts.

- ✓ A ce sujet, l'approche suivie permet de modéliser les coûts selon une loi de longue queue de distribution et selon une procédure d'estimation des paramètres cohérente avec cette distribution. Cette technique peut éliminer la nécessité de retraiter les sinistres graves.

¹ Le Groupe Consultatif Actuariel Européen a été créé en 1978 pour représenter l'opinion des actuaires devant les institutions européennes.

- **Intégration des éléments inflationnistes**

Le rapport préconise de distinguer les effets de la réassurance, de l'inflation et des recours dans le calcul de la meilleure estimation. Quatre étapes de calcul sont définies :

- La projection des triangles de charges et de règlements bruts de recours et de réassurance, avec prise en compte de l'inflation.
- La projection des triangles de recours.
- L'estimation de l'impact de la réassurance.
- L'actualisation des flux futurs au taux sans risque.

Au sujet de l'inflation, rappelons que l'article 77 de la directive prévoit la prise en compte de ses effets dans le calcul des provisions techniques. Parallèlement, le rapport de l'ACAM rappelle que la hausse des coûts dans la branche corporelle automobile est très supérieure à l'inflation constatée économiquement, du fait des dérives jurisprudentielles et conjoncturelles. En ce qui concerne la procédure de calcul, le rapport conseille d'estimer les liquidations de sinistres à partir des historiques retraités de l'inflation passée, puis d'appliquer les conditions d'inflation future aux projections de prestations.

- ✓ Le modèle présenté estime l'inflation passée de la branche à partir des historiques retraités de l'inflation. Cette première approche permet de projeter les prestations futures en monnaie constante. L'inflation générale contenue dans les prestations futures est ensuite estimée à partir d'un modèle ad hoc et appliquée aux projections précédentes.

- **Escompte des flux**

En ce qui concerne l'escompte des flux, le groupe de travail estime raisonnable de considérer les approximations suivantes :

- Dans la mesure où les prestations sont peu corrélées aux variations de la courbe des taux d'intérêt, l'utilisation d'une courbe des taux déterministe est suffisante. En ce sens, on se dispense de probabiliser l'évolution des taux dans le temps.
 - On peut également considérer des échéances déterministes, en se dispensant de probabiliser les dates de règlement des sinistres.
- ✓ Le calcul de l'espérance des prestations futures est conduit séparément de celui de leur actualisation, effectuée à partir d'une courbe des taux déterministe et selon des dates de règlements estimées en conformité avec les hypothèses du modèle.

III. L'expression de la provision pour sinistres

L'information en assurance dommages n'est dans certains cas que provisoire. On se retrouve ainsi en date d'évaluation des provisions pour sinistres dans une situation où l'observation des données ne correspond pas à la situation ultime. Dans ce contexte, l'objectif du provisionnement consiste à anticiper le développement des charges de sinistres pour les intégrer immédiatement dans le résultat.

Dans cet objectif, le provisionnement en assurance non-vie s'intéresse au déroulement temporel des paiements d'une branche d'assurance. Les sinistres sont rattachés à des années de survenance $i = 1, \dots, n$ ou de souscription et rapportés à des années d'inventaire.

Sans perte de généralités, on suppose pour simplifier la notation que les sinistres se déroulent sur n années : la charge ultime est donc connue de façon certaine passée ce délai. On désigne par x_{ij} ou X_{ij} le paiement effectué la $j^{\text{ème}}$ année de développement au titre des sinistres survenus l'année i . Ces variables supposées indépendantes sont appelées incréments. Lors de la clôture de fin d'année n , les montants x_{ij} antérieurs à cette date ($i+j \leq n+1$) sont disposés sous la forme d'un triangle de liquidation dont la forme est présentée au tableau 1.

Tableau 1 – le triangle de liquidation

i	j	1	J	l	n	Charge ultime
1		$x_{1,1}$			$x_{1,n}$	$C_{1,n}$
.						.
i			$x_{i,j}$			$C_{i,n}$
.						.
k				$x_{k,l}$		.
.						.
n		$x_{n,1}$			$x_{n,n}$	$C_{n,n}$

Les diagonales du tableau correspondent à des règlements calendaires : l'année n a donné lieu au paiement $\sum_{i=1}^n x_{i,n-i+1}$, toutes années de survenance confondues.

Parmi les n^2 variables du tableau 1, les variables $(X_{i,j})_{i+j-1 \leq n}$, du triangle supérieur ont été observées, les $(x_{i,j})_{i+j-1 \leq n}$ étant leurs réalisations. L'objectif consiste à compléter le triangle inférieur, constitué des règlements restant à effectuer au titre des années 2 à n . Ces quantités sont considérées comme des variables aléatoires $X_{i,j}$ ($i, j = 2, \dots, n$; $i+j-1 > n$).

Au 31/12/ n , la charge ultime au titre de l'année i est constituée de la somme des prestations passées (connues) et des prestations futures (aléatoires) :

$$C_{i,n} = \sum_{j=1}^{n-i+1} x_{i,j} + \sum_{j=n-i+2}^n X_{i,j}.$$

L'escompte des provisions techniques est un des points majeurs apportés par Solvabilité II qui prévoit que le Best Estimate doit être calculé « compte tenu de la valeur temporelle de l'argent ». En supposant par exemple que les prestations soient réglées en fin d'année, le Best Estimate de la provision pour sinistres constituée au 31/12/ n au titre de la survenance i est égal à la somme des prestations futures actualisées à la courbe des taux sans risque :

$$\hat{R}_i = \sum_{j=n-i+2}^n \hat{X}_{i,j} \cdot (1 + i(0, j - (n - i + 1)))^{n-i+1-j}$$

où $i(0, k)$ désigne le taux sans risque en 0 d'échéance k années et $\hat{X}_{i,j}$ l'estimation de la prestation payée en fin d'année de développement j et revalorisée de l'inflation.

Si comme nous le faisons ici, on suppose le risque totalement écoulé en n années, le Best Estimate de la provision pour sinistres s'écrit $\hat{R} = \sum_{i=2}^n \hat{R}_i$.

CHAPITRE II. LE RISQUE DE PROVISIONNEMENT

I. Les fondements du risque

I.1 L'incertitude des PSAP : une analyse par risques

Quel que soit l'horizon de solvabilité étudié, la source du risque de provisionnement est la même : l'incertitude liée à l'estimation des provisions. On peut associer les facteurs d'incertitude à trois risques : le risque de volatilité, le risque d'incertitude et les événements extrêmes.

■ Risque de volatilité

Le risque de volatilité est associé aux fluctuations aléatoires en termes de fréquence ou de valeur d'un événement. Il est considéré diversifiable dès lors que la loi des grands nombres assure la baisse de la volatilité du montant moyen de sinistres lorsque le nombre de risques assurés identiques et indépendants augmente.

■ Risque d'incertitude

Le risque d'incertitude correspond au risque d'une mauvaise spécification de la modélisation de la sinistralité ou d'une estimation erronée des paramètres. Il n'est donc pas un risque diversifiable, car il ne peut être réduit par l'augmentation de la taille du portefeuille.

• Risque de paramètres

L'estimation des provisions passe par l'extrapolation des prestations futures à partir des observations passées. Classiquement, les modèles de projection sont ajustés à partir des prestations observées. Ces modèles sont exposés à un risque d'échantillon, au sens où les historiques de prestations disponibles souvent courts contraignent à estimer leurs paramètres à partir d'un faible nombre d'observations. De façon générale, les branches sont d'autant plus exposées au risque qu'elles sont volatiles en nombre et en montant de prestations. Dans le cas des méthodes par facteurs, le risque est important dans le calcul des coefficients de passage des dernières années, basé sur peu de données.

• Risque de modèle

La capacité du modèle à estimer les prestations futures repose en fait sur le respect de deux hypothèses implicites. Tout d'abord, on suppose que le passé est correctement modélisé : les variables explicatives de la formation des prestations sont pertinentes et exhaustives et la structure du modèle bien spécifiée. Ensuite, en supposant que les caractéristiques des prestations futures de sinistres resteront conformes au modèle, on postule également l'absence de changements fondamentaux ou changements de régime dans la formation des prestations.

Ces changements de régime sont imputables à de nombreux facteurs. Par exemple, des décisions réglementaires ou judiciaires influant sur le niveau moyen des prestations, des changements dans la procédure interne de gestion des sinistres ou dans la nomenclature des produits, la révision de contrats de réassurance sont autant de causes susceptibles de rendre caduque la cadence des paiements utilisée dans la projection des prestations futures.

Ces facteurs d'incertitude peuvent conduire à des changements gradués ou brutaux des paramètres réels du modèle de provisionnement. Barnett et Zehnwirth² fournissent un exemple intéressant d'analyse des résidus définis par la différence entre les prestations payées et celles estimées par le modèle. Si les résidus présentent une hausse soudaine ou suivent une tendance régulière, il y a alors lieu de suspecter l'existence d'un changement de régime.

■ Evénements extrêmes

Ils sont caractérisés par une faible fréquence et un impact important. Du fait de leur rareté, leur occurrence ne peut être prédite à partir de l'extrapolation d'événements plus communs. Il est difficile de leur attribuer un montant de perte probable et par extension de capital requis.

I.2 De l'incertitude des PSAP au risque de provisionnement

I.2.1 L'exposition des branches à déroulement long

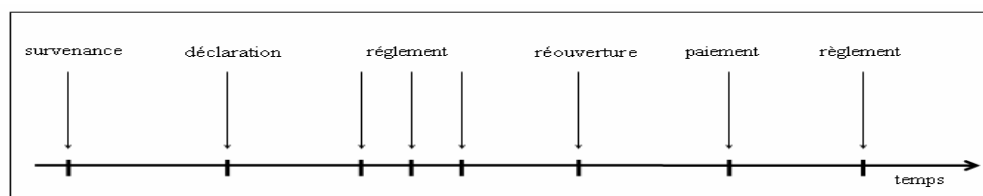
Les ratios de provisions techniques ont augmenté au fil du temps sur l'ensemble des risques dommages, en particulier dans les branches responsabilité civile et accidents. En assurance de responsabilité civile, les provisions représentent ainsi entre trois à cinq fois les primes acquises, contre 60% à 100% en assurance de biens et entre seulement 10% à 30% en assurance tous risque automobile.

Cette tendance est liée en partie à la part croissante des sinistres à développement long, dont les provisions sont davantage sujettes à dérive du fait de multiples facteurs.

Tout d'abord, l'identification et le règlement des sinistres s'inscrivent sur une période longue. Ainsi, la date du paiement peut varier sensiblement selon la gravité et la complexité du dommage du fait :

- du délai de déclaration à l'assureur, soit la durée entre la date de survenance et la date de déclaration du sinistre à l'assureur.
- du délai de règlement, entre la date de déclaration et la date de règlement final.
- de la réouverture de nouveaux sinistres inattendus.

Figure 1 – Processus de règlement des sinistres



Dans le cadre du risque de provisionnement, les sinistres particulièrement critiques sont ceux qui ne sont pas encore connus, ou ceux qui sont déjà survenus mais pas encore liquidés.

² Barnett G, Zehnwirth B. (2000), « Best Estimate for reserves », Proceedings of the Casualty Actuarial Society 87, 245-321.

Dans le provisionnement des risques longs, il est également primordial d'anticiper les tendances économiques, sociétales et juridiques exerçant une influence sur les engagements de l'assureur.

Dans les économies développées, ce sont avant tout l'explosion des coûts de santé publique et les exigences accrues en matière de soins à domicile et en établissement hospitalier. Selon les prévisions de l'OCDE, le taux d'augmentation des frais de soins de santé pourrait aller jusqu'à 7,3%³ par an jusqu'en 2050, de par l'effet conjugué de la hausse des salaires jusqu'ici bas dans le secteur de la santé et d'une plus grande longévité de la population.

La croissance des frais médicaux est quant à elle due au progrès constant de la médecine et aux développements de nouvelles thérapies.

De façon générale, on conçoit donc bien qu'il est délicat d'anticiper ces facteurs latents socio-économiques qui modélisent le niveau des indemnisations et contraignent l'assureur à contrôler périodiquement la suffisance des provisions constituées. Par exemple aux Etats-Unis, les provisions en assurance de responsabilité civile médicale constituées au titre de l'année de survenance 2000 ont été augmentées de 25% en 2006⁴.

Cette observation rappelle deux caractéristiques majeures de l'activité d'assurance :

- La rentabilité d'une activité doit être considérée après l'écoulement total du risque. En ce sens, la littérature s'est jusqu'ici concentré sur le risque dit ultime d'insuffisance des provisions à couvrir les engagements de l'assureur sur la totalité du développement des sinistres.
- Si cette vision de long terme est centrale dans le calcul des provisions, elle est inopérante à prémunir l'assureur du risque d'insolvabilité sur un horizon court.

Ainsi, d'une année sur l'autre, l'information supplémentaire obtenue (tendance, caractéristiques statistiques, déclarations tardives, règlement définitif,...) peut conduire l'assureur à réviser le montant de ses provisions à la hausse et occasionner des pertes importantes. Ce risque de court terme constitue le risque de provisionnement tel que le définit Solvabilité II. Il doit être estimé et conduire à la constitution d'une charge en capital.

I.2.2 Le cas des dommages corporels en automobile

L'assurance dommages corporels en automobile couvre par définition les dommages corporels en cas d'accident automobile. L'assureur payeur est l'assureur de l'automobiliste fautif, ou du conducteur, en cas de dommages occasionnés à des passagers.

Compte tenu du rapide inventaire des facteurs de variations amorcé dans la section précédente, on conçoit bien que la branche dommages corporels en automobile soit particulièrement exposée au risque de provisionnement.

Ainsi, l'assureur est tout d'abord exposé au risque de provision du fait des délais de déclaration et de règlement des sinistres. En particulier dans le cas des préjudices corporels graves, les réclamations découlant d'un événement dommageable ne sont parfois connues que très longtemps après la survenance du sinistre.

³ Münchener Rück, « La gestion des cycles, les risques de variations », Enterprise Risk Management, 2/2008.

⁴ Swiss Re, « Provisionnement des sinistres en assurance de dommages : progresser sur un défi stratégique », Sigma n°2/2008.

Et même dans le cas des sinistres connus, leur écoulement souvent long expose l'assureur à l'évolution de multiples facteurs environnementaux :

Tout d'abord, comme les niveaux de prestations restent étroitement liés au droit national de la responsabilité, l'assureur n'est pas à l'abri d'une décision de justice le contraignant à réviser les provisions constituées au titre des sinistres en cours.

Dans le cas des dommages corporels graves, les rentes versées ainsi que les frais de soins et de traitement médical doivent être payées pendant la durée de vie de la victime. Lorsqu'une indemnisation en rente est décidée, elle concerne en général les postes de préjudices suivants : frais médicaux, autres frais entraînés par l'infirmité (notamment les frais de tierce personne dans le cas des grands invalides) et pertes de revenus. Et dans un monde où les changements vont à un rythme croissant, les dimensions des tendances de long terme de l'ensemble de ces postes sont facilement sous-estimées.

Enfin, on observe dans la plupart des pays que les indemnisations revendiquées sont de plus en plus souvent versées sous formes de rentes. Ce mode d'indemnisation expose l'assureur à la hausse de ces coûts à plus d'un titre : tout d'abord si le montant à verser est fixé selon une base nominale, sans possibilité d'adaptation ultérieure, l'assureur prend en charge essentiellement le risque de longévité. De plus, et à l'inverse d'un paiement en capital immédiat et libératoire, le paiement sous forme d'arrérages l'expose également à un risque de baisse des taux et au renchérissement des provisions constituées qui en découle.

Et dans tous les cas, s'il s'agit d'indemnisations sous forme de rentes, il est possible suite à une détérioration médicalement constatée que la réouverture d'un sinistre clôturé occasionne une augmentation subite des provisions.

II. Le risque de provisionnement de Solvabilité II

II.1 La définition du risque

Reprenons la définition du Capital de Solvabilité défini à l'article 100 de la directive :

« Le Capital de Solvabilité Requis correspond au capital économique dont à besoin une entreprise d'assurance ou de réassurance pour limiter la probabilité de ruine à 0.5%, c'est-à-dire à une seule occurrence tous les 200 ans. Son calcul utilise la méthode de la Value-at-risk, conformément à la formule standard ou dans le contexte d'un modèle interne : toutes les pertes potentielles sur les 12 mois à venir, y compris celles qui découleraient d'une réévaluation défavorable des actifs et passifs, doivent être évaluées [...]. »

En évaluant le risque en fin d'année n , l'application de cette définition au risque de provisionnement conduit à segmenter les risques qui lui sont associés selon deux périodes⁵ :

- **Une période de choc**, soit l'année $n+1$ durant laquelle un événement adverse survient, cet événement étant entaché d'une probabilité de survenance de 0.5%.
- **Une période affectée** débutant en fin d'année $n+1$ jusqu'à l'écoulement du risque, au titre de laquelle l'événement adverse aura conduit à réévaluer la charge de provisions.

⁵ International Association of Insurance Supervisors (Octobre 2008), « [Guidance paper No. 2.1.1 on the structure of regulatory capital es](#)

II.2 L'identification de la variable d'intérêt

La décomposition par Solvabilité II de la perte à un an en risques de prime⁶ et de provisionnement est assimilable à l'analyse traditionnelle du résultat par année de survenance. En d'autres termes, les risques de prime et de provisionnement correspondent respectivement et à peu de choses près⁷ à un risque de mali sur liquidation et à un risque de marge courante négative.

L'analyse par survenance distingue la survenance courante $n+1$ des survenances antérieures. Plaçons-nous au 31/12/ n et négligeons l'effet de l'actualisation des prestations. Le résultat technique de l'année $n+1$ est la somme de la marge courante $\tilde{T}$ et du résultat T de run-off :

$$\begin{cases} T = \sum_{i=1}^n \hat{R}_i^{(n)} - \left(\sum_{i=1}^n X_{i,n-i+2} + \sum_{i=1}^n \hat{R}_i^{(n+1)} \right) = R^{(n)} - X_{n+1} - R^{(n+1)} \\ \tilde{T} = P - \left(X_{n+1,1} + \hat{R}_{n+1} \right) = P - \tilde{X}_{n+1} - \tilde{R}^{(n+1)} \end{cases}$$

La marge courante $\tilde{T}$ est constituée par la différence entre les primes P payées l'année $n+1$ et la somme des provisions et prestations payées $\tilde{X}_{n+1}$ en $n+1$ qui leur sont associées.

Quant au résultat T , il est également appelé boni mali sur liquidation. Il correspond à la marge technique comptabilisée au titre des survenances antérieures à l'année $n+1$. Le résultat T est égal à la différence entre la provision d'ouverture $R^{(n)}$ et la charge de l'année constituée des prestations X_{n+1} payées durant l'année $n+1$ et de la provision de fermeture $R^{(n+1)}$.

Plaçons nous désormais au 31/12/ $n+1$, en considérant une année de survenance $i \leq n$. Le résultat qui lui est associé sur l'année $n+1$ s'écrit

$$T_i = \hat{R}_i^{(n)} + I - \left[X_{i,n-i+2} + \hat{R}_i^{(n+1)} \right].$$

Comme $\hat{R}_i^{(n)} = \sum_{j=n-i+2}^n \hat{X}_{i,j}^{(n)}$, on peut écrire

$$\begin{aligned} T_i &= \hat{R}_i^{(n)} + I - \left[X_{i,n-i+2} + \hat{R}_i^{(n+1)} \right] \\ &= \left(\hat{X}_{i,n-i+2}^{(n)} - X_{i,n-i+2} \right) + \sum_{j=n-i+3}^n \hat{X}_{i,j}^{(n)} - \sum_{j=n-i+3}^n \hat{X}_{i,j}^{(n+1)}. \end{aligned} \tag{1}$$

⁶ Le risque de primes concerne les sinistres non survenus, dont certains sont déjà provisionnés au travers de la provision pour primes, et dont les autres correspondant à des polices souscrites durant l'année $n+1$, sont couverts par les primes associées.

⁷ Il s'agit de réintégrer dans le risque de provisionnement les sinistres survenus en $n+1$ dont les primes associées ont été payées en n .

Dans l'équation (1) apparaissent les deux éléments constitutifs du boni mali de liquidation enregistrés l'année $n+1$ au titre de la survenance i :

- Le premier terme dit boni – mali sur prestations comptabilise la part de résultat liée à la différence entre le montant $x_{i,n-i+2}$ de prestations payées sur l'année $n+1$ et celui provisionné au 31/12/ n au titre de l'année $n+1$.
- Le deuxième terme dit boni – mali sur provisions représente la part de résultat liée à la réévaluation de la provision en fin d'année $n+1$, compte tenu de l'information obtenue durant l'année.

On distingue donc facilement le lien entre le niveau de la variable T et chacune des périodes en risque. L'effet d'un choc survenant l'année $n+1$ agit directement sur le niveau des prestations $X_{i,n+2-i}$ payées l'année même, tandis que la période affectée intervient indirectement à travers la réévaluation en fin d'année $n+1$ de la provision $R_i^{(n+1)}$, pour les survenances $i \leq n$. Par exemple, une décision de justice intervenant l'année $n+1$ peut conduire à une aggravation des sinistres sur l'année courante et à une révision à la fin de l'année des montants de provisions en couverture du risque affecté.

II.3 Risque à un an contre risque ultime

En ajoutant le terme $\sum_{j=1}^{n-i+1} x_{i,j}$ dans chacun des deux membres de la partie droite de l'équation (1), on a également

$$\begin{aligned} T_i &= \sum_{j=1}^{n-i+1} x_{i,j} + \sum_{j=n-i+2}^n \hat{X}_{i,j}^{(n)} - \left[\sum_{j=1}^{n-i+1} x_{i,j} + x_{i,n-i+2} + \sum_{j=n-i+3}^n \hat{X}_{i,j}^{(n+1)} \right] \\ &= \sum_{j=1}^{n-i+1} x_{i,j} + \hat{R}_i^{(n)} - \left[\sum_{j=1}^{n-i+2} x_{i,j} + \hat{R}_i^{(n+1)} \right] \\ &= \hat{C}_{i,n}^{(n)} - \hat{C}_{i,n}^{(n+1)} \end{aligned}$$

Le risque de provisionnement est capturé par la distribution de probabilité de la variable T , conditionnellement à l'information disponible à la fin de l'année n . Autrement dit, l'étude du risque de provisionnement à un an conduit à s'intéresser à la variabilité dans un an de l'estimateur de la charge ultime par rapport à sa valeur présente.

On note le contraste entre l'approche à un an requise par Solvabilité II et l'approche dite « à l'ultime » associée à la problématique du provisionnement. Ainsi, l'approche à l'ultime s'inscrit dans une perspective de long terme, en se focalisant sur l'incertitude liée au provisionnement jusqu'au développement total des sinistres. Elle se concentre sur l'étude de la variable $R_i^{(n)} - \sum_{j=n-i+2}^n x_{ij}$, soit la somme des pertes et profits constatés depuis la date d'évaluation jusqu'à l'écoulement total du risque, et dont seulement une partie viendra affecter le résultat de l'année $n+1$.

Dans une vision à l'ultime, on est donc intéressé par la variabilité de l'estimateur de la charge ultime par rapport à sa valeur à l'ultime. Dans une vision à un an, il s'agit de mesurer la variabilité de cet estimateur par rapport à sa valeur dans un an.

En guise d'illustration, prenons la méthode Chain Ladder. La technique consiste depuis la diagonale du triangle de liquidation à estimer à la fin de l'année n la charge ultime $C_{i,n}$ par le montant

$$\hat{C}_{i,n}^{(n)} = C_{i,n-i+1} \times \hat{f}_{n+1-i}^{(n)} \times \cdots \times \hat{f}_{n-1}^{(n)}, \text{ pour } 2 \leq i \leq n,$$

où les coefficients $\hat{f}_j^{(n)}$ représentent l'accroissement des paiements cumulés observés à la fin de l'année n , entre les années j et $j+1$.

Leur application successive aux montants connus $C_{i,n+1-i}$ permet pour chaque année de survénance i d'estimer la charge ultime $C_{i,n}$, $2 \leq i \leq n$ et ainsi la provision R_i .

En fin d'année n et $n+1$, les estimateurs des coefficients de développements sont

$$\hat{f}_j^{(n)} = \frac{\sum_{i=1}^{n-j} C_{i,j+1}}{\sum_{i=1}^{n-j} C_{i,j}} \quad \text{et} \quad \hat{f}_j^{(n+1)} = \frac{\sum_{i=1}^{n-j+1} C_{i,j+1}}{\sum_{i=1}^{n-j+1} C_{i,j}}.$$

Les estimations associées des charges ultimes s'écrivent

$$\hat{C}_{i,n}^{(n)} = C_{i,n-i+1} \prod_{j=n-i+1}^{n-1} \hat{f}_j^{(n)} \quad \text{et} \quad \hat{C}_{i,n}^{(n+1)} = C_{i,n-i+2} \prod_{j=n-i+2}^{n-1} \hat{f}_j^{(n+1)}, \quad \text{pour } i \geq 2.$$

Dans le cas de la méthode Chain Ladder, l'information obtenue sur l'année $n+1$ influe donc sur l'estimation de la charge ultime à deux niveaux :

- par l'intégration des prestations de l'année $n+1$ dans le calcul de la charge ultime.
- par l'application des nouveaux coefficients de passage à la nouvelle diagonale des règlements cumulés observés.

III. Mesures de risques

III.1 L'indicateur du MSE : une mesure du risque à l'ultime

L'erreur quadratique moyenne (« Mean Square Error ») constitue une mesure courante de l'incertitude liée à l'estimation des PSAP jusqu'au développement ultime des sinistres.

En désignant par $D = \{x_{ij} \mid i + j \leq n + 1\}$ l'information disponible en date de calcul, sa valeur correspond à la distance moyenne au carré entre la réalisation $C_{i,n}$ et sa valeur estimée :

$$\text{MSE}\left(\hat{C}_{i,n}\right) = E\left[\left(C_{i,n} - \hat{C}_{i,n}\right)^2 \mid D\right].$$

Comme⁸ $E[(X - c)^2] = \text{Var}(X) + [E(X) - c]^2$ pour toute constante c , on a par analogie

$$\text{MSE}\left(\hat{C}_{i,n}\right) = \text{Var}(C_{i,n} \mid D) + \left(E[(C_{i,n}) \mid D] - \hat{C}_{i,n}\right)^2,$$

car $\hat{C}_{i,n}$ est une constante connue dès lors que D est connu et $C_{i,n}$ est D - mesurable.

L'incertitude portant sur l'estimation de $C_{i,n}$ résulte donc de deux facteurs :

- Le premier terme est une mesure de l'erreur liée à la variabilité du processus ou du risque de volatilité (en anglais, process variance).
- Le second terme capture l'erreur liée à l'estimation des facteurs par l'écart au carré entre l'estimateur $\hat{C}_{i,n}$ et sa cible $E(C_{i,n} \mid D)$ (en anglais, parameter estimation error).

En remarquant que $C_{i,n} - \hat{C}_{i,n} = R_i - \hat{R}_i$, on a

$$\text{MSE}\left(\hat{R}_i\right) = E\left[\left(R_i - \hat{R}_i\right)^2 \mid D\right] = E\left[\left(C_{i,n} - \hat{C}_{i,n}\right)^2 \mid D\right] = \text{MSE}\left(\hat{C}_{i,n}\right)$$

Comme $E(C_{i,n}) = \hat{C}_{i,n}$, $\text{SE}(\hat{C}_{i,n}) = \sqrt{\text{MSE}\left(\hat{C}_{i,n}\right)}$ est l'estimateur de l'écart type de R .

III.2 La mesure du risque à un an

III.2.1 Solvabilité II : La mesure par la VaR

Reprenons la notation du II. On désigne la perte technique sur l'antériorité comptabilisée l'année $n+1$ par $L = -T$.

Dans un but de protection des assurés, Solvabilité II impose l'immobilisation d'un montant de capital minimal $\rho(L)$ ⁹, assurant le financement de ces engagements dans les cas où les ressources primaires constituées des primes, des provisions et des revenus financiers ne suffisent pas à honorer les engagements contractés auprès des assurés.

En munissant L d'une distribution de probabilité (risque neutre, historique,...), $\rho(L)$ doit en principe être choisi de telle sorte que l'évènement « $L > \rho(L)$ » soit improbable. A cette fin, Solvabilité II a privilégié la Value-at-Risk (VaR) parmi les mesures de risques.

⁸ Par linéarité de $E(X)$ on a $V(X) = E\{[X - c - E(X - c)]^2\} = E\{[X - c]^2\} - 2 \cdot \{E[X - c]\}^2 + \{E[X - c]\}^2 = E\{[X - c]^2\} - \{E[X - c]\}^2$.

⁹ La fonction ρ est une fonctionnelle à définir faisant correspondre au risque S un nombre positif noté $\rho(S)$ et vérifiant les propriétés de cohérence définies dans Artzner, Delbaen, Eber et Heath (1999), « Coherent Measures of Risk », Mathematical Finance, 9, pp. 203-228.

La VaR définit le α -quantile de L .

$$\text{VaR}_\alpha(L) = \inf\{x \mid P[L \leq x] \geq \alpha\} = Q_\alpha(L) = F_\zeta^{-1}(\alpha), \quad 0 \leq \alpha \leq 1,$$

où F_ζ^{-1} est la fonction quantile de L , soit l'inverse de sa fonction de répartition.

Un des reproches majeurs fait à la VaR est qu'elle ne précise ni l'importance de la perte, ni sa distribution au-delà d'un seuil¹⁰. Cet inconvénient peut être considéré comme majeur pour des distributions de pertes caractérisées par des queues épaisses.

C'est la raison pour laquelle on préfère souvent à la VaR la Tail VaR (TVaR) définie par

$$\text{TVaR}_\alpha(L) = \frac{1}{1-\alpha} \int_\alpha^1 \text{VaR}_\xi(L) d\xi, \quad 0 \leq \alpha \leq 1.$$

La TVaR est la moyenne des VaR de niveau de confiance supérieur ou égal à α . Pour des distributions continues, on a

$$\text{TVaR}_\alpha(L) = E[L \mid L > \text{VaR}_\alpha(L)].$$

Autrement dit, il s'agit de la traduction mathématique du concept de « perte moyenne dans les pires $(1 - \alpha)\%$ des cas ». A l'inverse de la VaR, la TVaR permet de distinguer deux distributions de perte qui ont le même α -quantile mais sont différentes par ailleurs (en particulier n'ont pas les mêmes queues de distribution, ou la même asymétrie).

La TVaR peut également s'exprimer en fonction d'une autre mesure de risque, celle d'*Expected Shortfall* définie par

$$\text{ES}_\alpha(L) = E[S - \rho(L)]^+$$

Le Shortfall peut s'interpréter comme la part de perte non couverte par l'assureur. En ce sens, elle définit le risque résiduel d'insolvabilité. Si $\rho(L) = \text{VaR}(L)$, la TVaR s'exprime comme une combinaison linéaire de la VaR et du Shortfall.

Comme

$$\begin{aligned} \text{ES}_\alpha(L) &= E[L - \text{VaR}_\alpha(L)]^+ = \int_0^1 (\text{VaR}_\xi(L) - \text{VaR}_\alpha(L))^+ d\xi \\ &= \int_\alpha^1 \text{VaR}_\xi(L) d\xi - \text{VaR}_\alpha(L) (1 - \alpha) \end{aligned}$$

$$\text{on a} \quad \text{TVaR}_\alpha(L) = \text{VaR}_\alpha(L) + \frac{1}{1-\alpha} E(L - \text{VaR}_\alpha(L))^+.$$

¹⁰ sauf quand une distribution théorique comme la loi normale est supposée a priori.

Même si la VaR est décriée à cause de son absence de sous additivité¹¹, Dhaene et al. (2004) montrent que cette mesure de risque est optimale au sens où elle minimise la fonction¹²

$$C(L, d) = E[(L - d)^+] + d\varepsilon,$$

tout en conduisant au plus petit niveau de capital d .

III.2.2 Méthodes approchées

■ Une approche historique

Une méthode simplifiée de calcul du capital requis est proposée dans l'étude de l'AISAM-ACME (2007). Elle consiste à calculer la charge de capital par application d'un chargement de solvabilité à la meilleure estimation de la provision d'ouverture. En ce sens, la solution est analogue à la pratique actuelle d'inclusion d'une marge de prudence dans la provision.

La valeur de la marge de prudence est définie à partir des données historiques de la branche. Plus précisément, l'approche vise à capturer l'augmentation relative à un an des provisions définie par

$$m = \frac{R^{(n)} - R^{(n+1)} - X_{n+1}}{R^{(n)}}$$

Alors, la charge en capital estimée au 31/12/n est définie par

$$C = R^{(n)} \times Q_{99,5\%}(m)$$

Cette approche rétrospective fournit une indication sur la volatilité historique des provisions à un an. Toutefois, la méthode reste dépendante de la définition de $R^{(n)}$. Par exemple, une méthode alternative de provisionnement réduisant de 10% le niveau des provisions conduit au même chargement en capital

$$m = \frac{0.9R^{(n)} - 0.9R^{(n+1)} - 0.9X_{n+1}}{0.9R^{(n)}} = \frac{R^{(n)} - R^{(n+1)} - X_{n+1}}{R^{(n)}},$$

mais diminue également de 10% le niveau de ressources $R^{(n)} \times (1+m)$ devant assurer la solvabilité à un an dans 99.5% des cas.

En ce sens, cet exemple met en avant la nécessité de définir le Best Estimate dans la distribution du risque. A ce sujet, nous avons vu que la notion de meilleure estimation correspondait à l'espérance actualisée de la somme des prestations futures.

¹¹ Une mesure de risque est sous-additive si $\rho(S_1 + S_2) \leq \rho(S_1) + \rho(S_2)$. Intuitivement, la propriété est légitime si en cas d'évènement adverse, le capital résiduel du risque i « excédentaire » tel que $\rho(S_i) - S_i > 0$, permet de couvrir la perte du risque « déficitaire ».

¹² $d\varepsilon$ est assimilable au coût de rémunération du capital immobilisé. La démonstration des auteurs est présentée en annexe.

■ Une approche par les quantiles

En prolongement de l'approche de l'AISAM-ACME, Dhaene et al. (2008) proposent une solution plus adaptée :

En occultant l'effet de l'actualisation, la quantité inconnue $R^{(n+1)} + X_{n+1}$ est le montant de ressources nécessaire durant l'année $n+1$. Cette quantité peut s'écrire en fonction de la provision certaine $R^{(n)}$ d'ouverture, avec

$$\begin{aligned} R^{(n)} + R^{(n)} \times \frac{R^{(n+1)} + X_{n+1} - R^{(n)}}{R^{(n)}} \\ = R^{(n)} + R^{(n)} \times \beta = R^{(n)} \times (1 + \beta) \end{aligned}$$

Le paramètre β désigne la révision du risque effectuée durant l'année $n+1$. Ainsi, le risque de provision considère les cas où β est strictement positif, soit

$$R^{(n+1)} + X_{n+1} > R^{(n)}.$$

Pour décrire le risque de provisionnement, il s'agit de trouver un estimateur de la volatilité du facteur d'accroissement β . Dans le cadre de Solvabilité II, le capital requis doit permettre de couvrir la perte technique liée au risque de mali sur liquidation dans 99,5% des cas, soit

$$Q_{99,5\%}(\beta \cdot R^{(n)}) = R^{(n)} \cdot Q_{99,5\%}(\beta)$$

Si l'on suppose β gaussien par simplification, on obtient

$$C = Q_{99,5\%}(\beta \cdot R^{(n)}) = R^{(n)} [E(\beta) + \sigma(\beta) \cdot \phi^{-1}(0.995)]$$

où les paramètres $E(\beta)$ et $\sigma(\beta)$ peuvent être estimés à partir de données historiques.

PARTIE II. LA SELECTION DU MODELE

L'aboutissement de cette partie est l'identification du modèle utilisé dans le calcul du Best Estimate et du risque de provisionnement. Nous choisissons d'étudier la branche des dommages corporels automobile caractérisée par un développement long.

Les réponses à certaines questions posées lors de la définition d'un modèle s'avèrent cruciales dans le cas des branches longues : comment modéliser la dérive inflationniste ; comment projeter les prestations au-delà de l'horizon observé dans le cas de risques partiellement écoulés, comment tenir compte d'une éventuelle dérive temporelle de la cadence des prestations ?

Inspirée de Wright (1992) et Taylor (2004), la méthode présentée dans cette partie propose des axes de réponse à ces questions. Tout d'abord, le modèle utilisé intègre implicitement la modélisation des facteurs de queue. Ensuite, sa structure souple facilite l'inclusion de facteurs influents dans la dérive temporelle des prestations de sinistres, tels les phénomènes d'inflation et de surinflation. Enfin, la méthode propose une modélisation alternative à la méthode par facteurs, dans le cas courant pour les branches longues où la cadence des prestations est instable.

L'approche est de type coût moyen. Elle suppose de disposer par survenance et année de développement des prestations de sinistres, des nombres de sinistres réglés et des estimations des nombres de règlements à l'ultime¹³.

Le cheminement au modèle final procède en quatre temps. La première partie étudie les données et justifie du recours au modèle de provisionnement utilisé. Tout en introduisant les outils propres aux GLM, la seconde partie définit les caractéristiques fondamentales du modèle. La troisième partie spécifie la structure du modèle tandis que la quatrième procède aux tests usuels de validation.

¹³ Et éventuellement les nombres de sinistres en cours, dans le cadre de la modélisation des paiements partiels.

CHAPITRE I. ETUDE DES DONNEES

I. L'exposé du problème

I.1 Limites de l'usage des méthodes par facteurs

Pour calculer la provision pour sinistres, nous disposons de différentes techniques qui se basent sur l'analyse du triangle de liquidation. La plupart de ces techniques passent par la recherche de facteurs de développement mesurant les taux de progression de la charge entre deux âges successifs.

Citons par exemple le modèle GLM *Poisson surdispersé*¹⁴ supposant les incréments indépendants et distribués selon une loi poisson, de moyenne et de variance

$$E(X_{ij}) = \mu_{ij} = e^{\alpha_i + \beta_j} \text{ et } \text{Var}(X_{ij}) = \phi \mu_{ij}.$$

Le lien exponentiel permet une interprétation simple des coefficients α_i et β_j . En reprenant la notation de la section précédente, on a

$$E(C_{i,n}) = \sum_{j=1}^n \mu_{ij} = e^{\alpha_i} \cdot \sum_{j=1}^n e^{\beta_j}.$$

soit $\ln[E(C_{i,n})] = c + \alpha_i$, où $c = \ln \left[\sum_{j=1}^n e^{\beta_j} \right]$.

Cette égalité donne une interprétation du paramètre α_i en termes de charge ultime espérée de sinistres.

Par ailleurs, comme

$$e^{\beta_j} = \frac{e^{\alpha_i + \beta_j}}{e^{\alpha_i}} = \frac{E(X_{ij})}{E(C_{i,n}) / \sum_{j=1}^n e^{\beta_j}},$$

on a $\frac{E(X_{ij})}{E(C_{i,n})} = \frac{e^{\beta_j}}{\sum_{j=1}^n e^{\beta_j}} \quad \forall i.$

Cette dernière quantité représente la proportion moyenne des paiements attendus au cours de l'année de développement j . Remarquons que cette proportion est indépendante des années de survenance. En d'autres termes, la projection des incréments futurs repose sur l'application à chaque survenance d'une unique cadence de développement.

Hors, la pertinence de ce choix de modélisation suppose que les données observées suivent des évolutions relativement régulières. De ce fait, les méthodes par facteurs doivent donc être utilisées avec précaution, notamment dans le cas des branches longues dont la durée d'écoulement rend plus improbable l'hypothèse de stabilité des cadences de prestations.

¹⁴ On rappelle que ce modèle conduit aux mêmes résultats que ceux issus de la méthode déterministe Chain Ladder.

I.2 Quelques méthodes alternatives

Avec le temps, de nombreux facteurs peuvent en effet venir influencer sur le rythme de paiement des prestations. A ce sujet, le rapport de l'ACAM (2007) rappelle quelques méthodes alternatives à la méthode des facteurs, dans des cas où l'application d'une cadence unique n'est pas souhaitable.

- **Tendances des facteurs**

Les méthodes par facteurs calculent des coefficients de développement de façon globale. De ce fait, elles ne permettent pas de distinguer des évolutions dans le temps de la cadence des paiements. Une solution consiste à utiliser des techniques qui passent par le calcul de facteurs par année de survenance.

Selon les tendances observées, on peut retenir des facteurs calculés :

- A partir d'une moyenne pondérée, en attribuant un poids plus important aux facteurs des survenances les plus récentes.
- A partir d'une moyenne portant sur les derniers facteurs. Se pose alors la question de l'horizon à retenir.

Si on observe des tendances, il est impératif d'en tenir compte dans le calcul des facteurs de développement. La solution choisie doit alors être d'autant plus justifiée qu'elle garde dans tous les cas une part de subjectivité.

- **Dispersion des facteurs**

Si le développement des sinistres suit une loi générale, il est normal d'observer des fluctuations dans les trajectoires de prestations par survenance. Dans tous les cas, que les facteurs soient peu ou fortement volatiles, l'application d'une cadence unique à l'ensemble des survenances demeure une solution moyenne. Alors, les diverses solutions possibles basées sur les facteurs peuvent selon leur dispersion diverger sensiblement.

Différents modèles stochastiques permettent de définir un intervalle de confiance dans l'estimation de la provision (Munich Chain Ladder, Bootstrap). L'utilisation de ces modèles est souvent conditionnée par l'hypothèse sous jacente d'indépendance des facteurs de développement¹⁵. Ce postulat exclut en particulier les phénomènes de compensation des règlements de sinistres d'un délai au suivant pour une même année de survenance.

- **Volatilité des facteurs**

Dans son rapport, le groupe de travail dissocie la dispersion naturelle de la volatilité des facteurs de développement. Cette dernière faisant référence à une disparité plus heurtée, est fréquemment observée dans le cas des branches longues pour lesquelles une faible proportion de la charge ultime est connue lors des premières années de développement.

¹⁵ Dans son article « Measuring the variability of Chain Ladder Reserve Estimates » de 1993, T. Mack propose un test de contrôle de l'absence de corrélation entre les facteurs.

A ce sujet, une illustration de l'effet de l'application d'une cadence unique en présence de volatilité des facteurs est donnée à partir du portefeuille étudié (voir II.2.2).

Dans ce cas, le rapport de l'ACAM préconise deux types d'approches :

- La méthode Bornhuetter-Fergusson combine l'approche traditionnelle des facteurs avec une estimation « a priori » de la charge ultime. Au fil des années de développement, cette approche de type crédibilité accorde une pondération plus importante à l'observation des facteurs.
- Les méthodes basées sur des facteurs de développement en valeur absolue, construites par exemple sur des historiques de rapports de sinistres à primes.

Ces deux méthodes permettent de limiter l'instabilité des projections liée à la sensibilité de l'approche par facteurs des premières années. En revanche, leur utilisation reste délicate. Pour la première, la principale difficulté réside dans la justification de l'estimation « a priori » de la charge ultime. Quant à la seconde, elle reste conditionnée à la stabilité des loss ratio d'une survenance à l'autre.

- **Eléments de rupture**

Indépendamment de leur volatilité naturelle, les trajectoires de développement des sinistres peuvent également être perturbées par des éléments de rupture. Ces ruptures interviennent souvent lors d'un même exercice en touchant l'ensemble des années de survenance¹⁶. Elles peuvent être causées tant par des sources internes (changement des règles de gestion, des dates de cut-off ou de la procédure de sélection des risques, ...) qu'externes (évolution jurisprudentielle, économique ou sociale, ...) à l'entreprise.

Dans tous les cas, il est primordial de retraiter les données de ces éventuels phénomènes, de sorte à ne pas utiliser des facteurs erronés dans le calcul des provisions.

I.3 Le cas des données étudiées

Les données que nous étudions sont issues de Taylor (2000). Nous ne disposons d'aucune statistique relative au portefeuille étudié (benchmark, loss ratio, ...). La seule information connue concerne la branche du portefeuille, soit celle des dommages corporels en automobile.

En fait, les caractéristiques de cette branche nous situent au carrefour des différentes limites des méthodes par facteurs qui viennent d'être évoquées. Dans le cas des données étudiées, l'analyse à venir invalide l'usage des méthodes par facteurs pour différentes causes. On y montre notamment :

- que les facteurs de développement sont volatiles et sujets à tendances.
- Que le volume des paiements est dépendant des procédures de gestion, à travers les nombres de règlements dont les cadences sont par ailleurs instables.

¹⁶ Dans son article « Measuring the variability of Chain Ladder Reserve Estimates », T. Mack propose également un test de contrôle vérifiant qu'il n'existe pas de rupture d'ensemble des facteurs sur une année d'exercice.

II. La discussion sur les données

II.1 La description des données

II.1.1 Les données utilisées

Si les techniques de provisionnement passent invariablement par l'analyse du triangle de liquidation, il n'existe pas de consensus clair sur la nature des données à prendre en compte dans l'estimation de la charge ultime : étudie-t-on les sinistres payés ou survenus ?.

Dans le premier cas, les incréments X_{ij} représentent les sinistres survenus l'année i et payés avec j années de retard. Il s'agit du triangle des prestations réglées.

Dans le second cas, les incréments X_{ij} représentent les sinistres survenus l'année i et déclarés la $j^{\text{ème}}$ année de développement : dans ce dernier cas où les sinistres n'ont pas été entièrement réglés, le triangle inclut leurs estimations en lieu et place des règlements à venir.

Les deux issues présentent des avantages et des inconvénients :

- Si on inclut seulement les sinistres payés, les données du triangle reflètent la situation réelle, sans prise en compte de jugement humain. Cependant, la connaissance des IBNR n'est pas exploitée, et le triangle n'apporte aucune information sur les paiements futurs.
- A l'inverse, les estimations des IBNR peuvent être volatiles, si elles sont soumises à différentes influences économiques et politiques plus ou moins objectives.

Cette réalité complique à double titre la lecture de la cadence des règlements. Tout d'abord, du fait des réévaluations des estimations de sinistres, les montants réglés peuvent différer sensiblement des montants estimés. Ensuite, les sinistres sont considérés réglés en date d'estimation et non de règlement.

Le modèle utilisé définit le coût moyen d'un sinistre en fonction de l'ordre d'occurrence de leur règlement. Son utilisation requiert la connaissance des données suivantes :

- Les sinistres payés retraités de l'inflation sont présentés en annexe A. Les prestations sont exprimées en dollars du 31/12/1995.
- Les nombres N_{ij} de sinistres survenus au titre de la survenance i et réglés avec j années de retard sont présentés en annexe B.
- Les nombres N_i des sinistres survenus et à survenir au titre de la survenance i sont présentés en annexe C. Les estimations des nombres de sinistres survenus restant à déclarer dits tardifs correspondent aux estimations de Taylor (2000).

II.1.2 L'estimation du nombre de sinistres survenus restant à déclarer

Considérons la survenance i et notons D_{ij} le nombre de sinistres déclarés durant l'année de développement j . Classiquement, nous disposons du triangle des D_{ij} , pour $j = 1, \dots, n - i + 1$, et nous cherchons à extrapoler la série des D_{ij} pour $j = n - i + 2, \dots, \infty$.

On définit le nombre cumulé des sinistres déclarés au titre de i par

$$A_{ij} = \sum_{m=1}^n D_{i,m}$$

Taylor (2000) extrapole le nombre des sinistres tardifs à partir de la série historique des D_{ij} . Plus précisément, l'extrapolation des tardifs est établie en deux temps¹⁷ :

- Il estime les facteurs de passage v_j à partir de la méthode Chain Ladder pour $j \leq 7$. Autrement dit, on définit comme estimateur de v_j la quantité

$$\hat{v}_j = \frac{\sum_i A_{i,j+1}}{\sum_i A_{i,j}}, \quad j = 1, \dots, 7,$$

où la sommation en i se fait pour l'ensemble des valeurs disponibles¹⁸.

- Pour les coefficients suivants, il ajuste une courbe aux quantités $\hat{v}_j - 1, j=1, \dots, 7$, qu'il utilise pour les délais $j > 7$. Le modèle ajusté est le suivant :

$$\hat{v}_j - 1 = 0.0011 \times 0.59^{j-7}$$

La reconstitution des coefficients de passage est exposée en annexe T. Pour finir, remarquons que le modèle permet l'extrapolation des tardifs pour les délais importants. Par exemple pour $j \geq 10$, comme $\hat{v}_j - 1 \rightarrow 0$, on a

$$\prod_{j=10}^{\infty} \hat{v}_j \cong 1 + \sum_{j=10}^{\infty} [\hat{v}_j - 1]$$

Comme

$$\begin{aligned} \sum_{j=10}^{\infty} [\hat{v}_j - 1] &= 0.0011 \times \sum_{j=10}^{\infty} 0.59^{j-7} = 0.0011 \times 0.59^3 \sum_{j=0}^{\infty} 0.59^j \\ &= 0.0011 \times \frac{0.59^3}{0.41} = \frac{\hat{v}_{10} - 1}{0.41} = \frac{0.59}{0.41} \times (\hat{v}_9 - 1) \end{aligned}$$

D'où
$$\prod_{j=10}^{\infty} \hat{v}_j \cong 1 + 0.59/0.41 \times (\hat{v}_9 - 1).$$

Au final, on obtient
$$\hat{A}_{i,\infty} = A_{i,n-i+1} \prod_{j=n+1-i}^{\infty} \hat{v}_j.$$

¹⁷ Pour plus d'explications, se reporter à la page 17 de Taylor (2000). On y trouvera plusieurs techniques d'extrapolation des sinistres tardifs.

¹⁸ Sur cet exemple, Taylor estime les coefficients de passage selon des historiques différents détaillés en annexe C (tableau 29 bis).

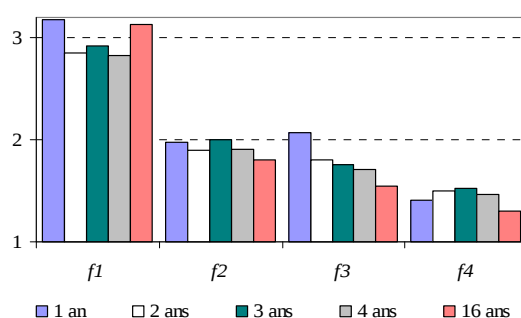
II.2 L'application de la méthode Chain Ladder

II.2.1 L'étude préalable des rapports historiques de développement

Les coefficients f_j calculés par Chain Ladder et les facteurs de développement cumulés associés f_j' sont présentés en bas de triangle de l'annexe A.

Les figures 2 et 3 présentent les facteurs f_1 à f_8 , calculés à partir d'un horizon allant du dernier exercice comptable 1995 (en bleu) à l'ensemble des seize années disponibles (en rose). Elles témoignent clairement de l'instabilité de la cadence des prestations dans le temps.

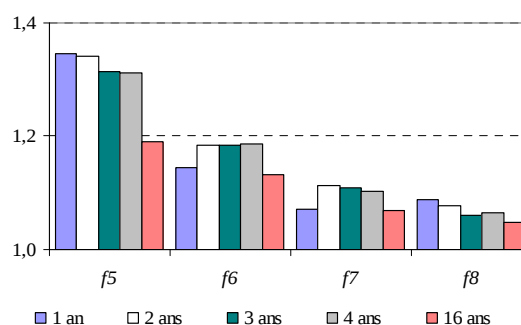
Figure 2 – Rapports moyens de développement f_j , $j = 1$ à 4



A priori, la modélisation d'une cadence unique de paiements basée sur les années de développement apparaît compromise à double titre du fait :

- De la tendance générale à la hausse des rapports historiques¹⁹.
- De la volatilité des coefficients, notamment dans le cas des délais 1 et 3.

Figure 3 – Rapports moyens de développement f_j , $j = 5$ à 8



Dans le cas des branches longues, l'hétérogénéité des cadences de paiements peut conduire à une instabilité forte dans le calcul des provisions. Autrement dit, l'application de facteurs multiplicatifs de niveau élevé peut entraîner des effets fortement amplifiés sur l'évaluation de la charge totale et en conséquence conduire à des niveaux de provisions très volatiles.

¹⁹ Les valeurs des rapports historiques sont présentées en annexe A pour les neuf derniers exercices.

Par exemple, l'application de la méthode Chain Ladder ajustée à partir des seize années d'historique suppose que la part de sinistres réglés en fin de 1^{ère} année est en moyenne de

$$\frac{1}{f_1} = \frac{1}{18,355} = 5,45\%.$$

Autrement dit, pour 0.0545 euro réglé en fin de 1^{ère} année de développement, la méthode estime le montant de la charge ultime à 1 euro. A ce sujet, rappelons que l'application d'une cadence moyenne de paiements à l'ensemble des survenances suppose implicitement l'homogénéité du développement de leurs prestations.

Par exemple dans le cas de la survenance 1979 où la charge ultime est a priori connue, l'étude à postériori de la cadence de ses prestations conclut que 9% des prestations étaient payées en fin de 1^{ère} année. L'application du coefficient moyen Chain Ladder à cette survenance aurait abouti en fin de 1^{ère} année à une surestimation de sa charge ultime de

$$(0,09-0,0545) \times 18,355 = 64,9\%.$$

II.2.2 La sensibilité de la provision pour sinistres

La section précédente donne une mesure de la variabilité de l'estimation des PSAP du fait de la volatilité des facteurs. Dans le cas des données étudiées, le recours à une méthode par facteurs n'est de toute façon pas pertinent du fait des tendances décelées dans les facteurs.

Le tableau 2 présente les montants de PSAP obtenus en fonction du nombre d'exercices retenus dans le calcul des coefficients Chain Ladder, et le rapport (colonne %) de chaque provision par rapport à la PSAP calculée à partir des seize exercices observés.

Tableau 2 – PSAP - fin d'année 1995

Exercices retenus	(%)	PSAP
1995	1,768	382 249 638
1990 à 1995	1,474	318 723 185
1985 à 1995	1,055	228 155 210
1979 à 1995	1,000	216 250 173

Par exemple, l'utilisation des facteurs observés durant l'exercice 1995 conduit à une provision supérieure de 76,8% à celle estimée en recourant à l'ensemble des exercices disponibles.

De façon plus générale, on observe sans surprise du fait de la hausse des premiers facteurs de développement avec le temps que l'intégration progressive des exercices passés dans le calcul des coefficients de passage conduit à une baisse tendancielle du niveau des provisions. Dans ces conditions, le recours à un unique profil de développement dans le calcul des PSAP apparaît difficilement justifiable, d'autant que s'il est possible de tenir compte de la tendance haussière des facteurs en raccourcissant l'historique intervenant dans le calcul des coefficients f_j , il n'existe aucune méthode adéquate permettant d'en définir la profondeur.

II.3 De l'inadéquation des méthodes par facteurs

II.3.1 L'instabilité de la cadence des règlements

Il est courant que l'instabilité des facteurs de développement soit liée à celle du rythme de règlement des sinistres. C'est pour cela que Berquist et Sherman (1977) suggèrent de fixer le montant des IBNR à partir de la durée de règlement des sinistres restant à courir.

Dans cette même démarche, Reid (1978) introduit la notion de temps opérationnel en associant un indicateur de temps moyen

$$\tau_{ij} = \frac{N_{i1} + \dots + \frac{1}{2} \cdot N_{i,j}}{N_i}, \text{ avec } \tau_{ij} \in [0, 1],$$

à chaque cellule (i,j) du triangle, définissant la proportion de tous les sinistres clôturés au titre de la survenance i au milieu de l'année de développement j.

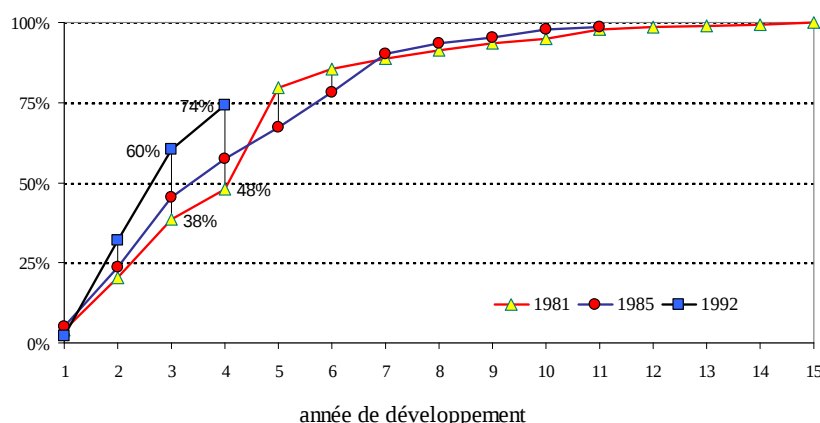
Les sinistres clôturés en (i,j) sont définis par l'égalité

$$N_{ij} = O_{i,j-1} + D_{ij} - O_{i,j}$$

où O_{ij} désigne le nombre de sinistres restant ouverts à la fin de la période associée à la cellule (i,j) et D_{ij} le nombre de sinistres déclarés durant cette même période.

La notion de temps opérationnel fait référence à la vitesse de règlement d'une cohorte de sinistres, correspondant dans notre cas aux sinistres survenus durant une même année de survenance. L'activité de la cohorte débute au temps opérationnel $\tau = 0$ et se termine lorsque le dernier sinistre de la cohorte est réglé en $\tau = 1$. La figure 4 reprend trois trajectoires de temps opérationnel du triangle présenté en annexe F.

Figure 4 – Trajectoires de temps opérationnel de trois années de survenances

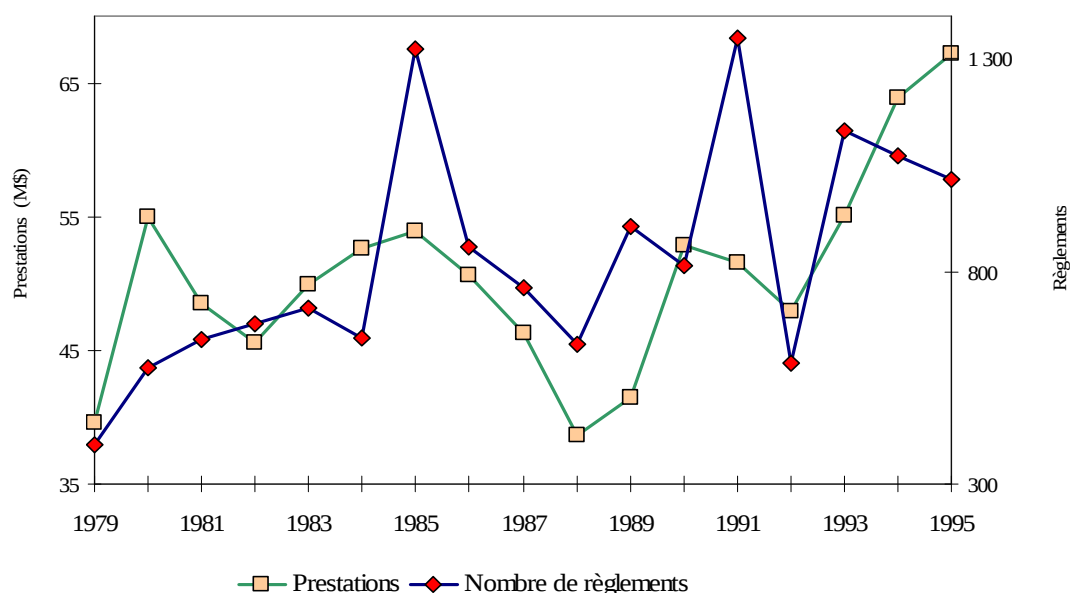


Les années de survenance présentées sont sélectionnées pour illustrer la forte volatilité des trajectoires de règlement, notamment dans le cas des premières années de développement. Par exemple, on observe que 74% des sinistres de la survenance 1992 ont été clôturés en fin de 4^{ème} année de développement, contre seulement 48% pour la survenance 1981.

II.3.2 La liaison entre volume des prestations et nombre de règlements

La figure 5 issue de Taylor (2000) superpose le nombre de sinistres réglés et le volume de prestations par année comptable. On observe le lien étroit entre ces deux variables.

Figure 5 – Volume des prestations (M\$) et règlements de sinistres



L'observation de la figure 5 nous rappelle cependant que la justesse de cette relation peut dans la pratique être contrariée, notamment du fait de deux phénomènes :

- Certains règlements de sinistres ne conduisent pas forcément à prestations :

Ainsi, l'activité de règlement des sinistres peut tout simplement être artificielle, au sens où elle répond davantage à une motivation administrative. Par exemple, on peut considérer une année d'inventaire durant laquelle l'assureur décide de clôturer le maximum de sinistres laissés ouverts sans utilité.

Sur la figure 7, on remarque que le pic de règlement de l'année comptable 1985 ne s'accompagne pas d'une hausse du volume des prestations. On observe également que cette année est adjacente à un minima local de la courbe des sinistres réglés en 1984. On peut imaginer qu'une partie des sinistres réglés en 1985 correspond à un rattrapage des sinistres sans suite laissés en suspens en 1984.

- Un certain nombre de prestations ne s'accompagnent pas d'un règlement :

En effet, un sinistre peut déclencher de nombreux versements. On parle de paiements partiels. Par exemple, ce phénomène peut être accentué dans le cas d'indemnisations sous la forme d'arrérages : le sinistre est considéré réglé avec le paiement du dernier arrérage.

II.3.3 Les coefficients de passage et le temps opérationnel

Si le volume des prestations dépend du nombre de sinistres réglés, on peut s'attendre à observer un lien entre l'évolution du temps opérationnel et celle des facteurs de développement.

La figure 6 superpose les évolutions dans le temps des rapports historiques de première année f_1 et du temps opérationnel τ_2 de fin de deuxième année de développement. La figure 8 effectue la même comparaison pour le facteur de troisième année et le temps opérationnel de fin de quatrième année de développement.

Figure 6 – Temps opérationnel τ_2 et rapport de développement f_1

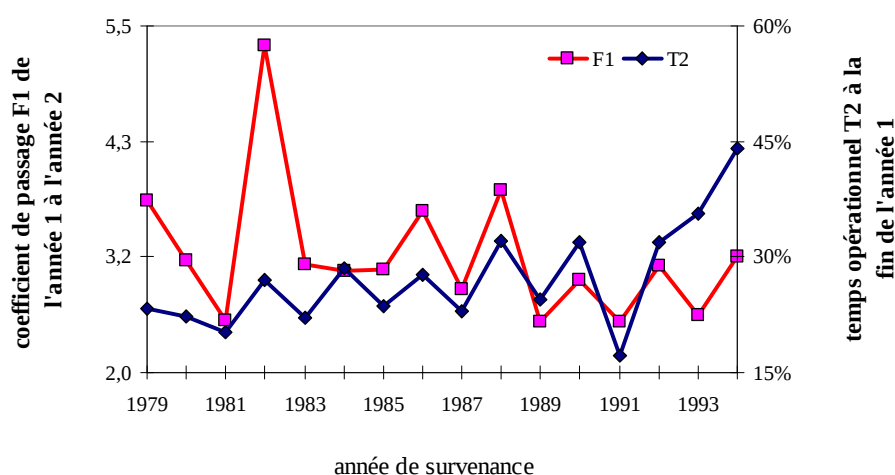
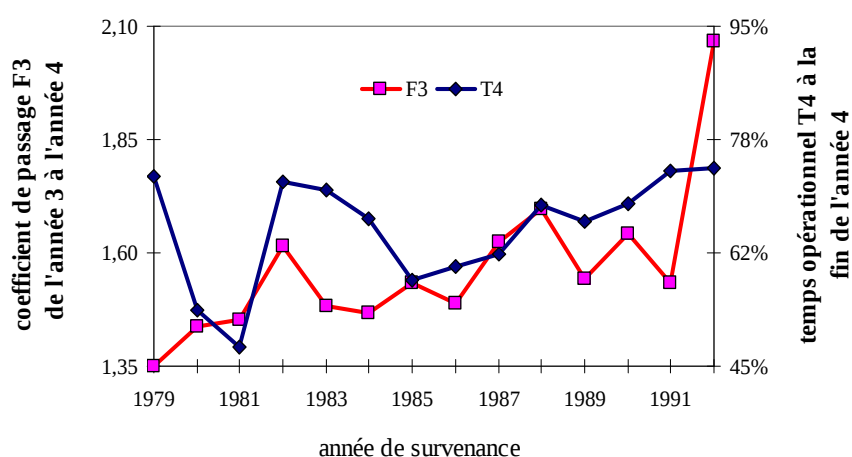


Figure 7 – Temps opérationnel τ_4 et rapport de développement f_3



Dans les deux cas, la croissance des facteurs de développement apparaît coïncider avec celle du temps opérationnel, même si les coefficients de corrélation entre les séries f_j et τ_{j+1} sont loin d'être parfaits. On observe par exemple que la hausse sensible du taux de règlement τ_2 pour les survenances postérieures à 1991 ne s'accompagne pas d'une augmentation du coefficient de développement associé.

III. L'approche par les coûts moyens

III.1 Les coûts moyens et le temps opérationnel

La section précédente met en avant l'existence d'un lien entre le volume des prestations payées et le nombre de règlements de sinistres. Le modèle le plus simple reconnaissant la dépendance entre ces deux variables suppose la relation

$$E(X_{ij}) = N_{ij} \cdot S_j, \quad (1)$$

où $S_j = \sum_{i \in I} X_{ij} / \sum_{i \in I} N_{ij}$ désigne l'estimateur du coût moyen par sinistre réglé²⁰ observé sur le délai j et I un ensemble approprié d'années de survénance.

L'utilisation de la relation (1) suppose implicitement pour un même délai que les coûts moyens par sinistre réglé restent stables d'une année de survénance à une autre. Dans notre cas, cette hypothèse n'est donc pas réaliste, du fait du lien établi entre le volume des prestations et le nombre des règlements de sinistres instables en cadence.

Dans le cas d'instabilité avérée de la cadence des règlements, supposer *l'hypothèse d'invariance de l'ordre* peut constituer un axe de modélisation plausible. Autrement dit, il s'agit de postuler que les variations du rythme de règlement n'ont pour seule conséquence que de compresser ou d'étendre la durée durant laquelle ces sinistres sont payés, sans affecter ni leur ordre de paiement, ni leur niveau de prestation.

Imaginons sur un exemple simple deux années de survénance engendrant chacune trois sinistres, de montants 1000, 2000 et 3000, réglés dans cet ordre pour chacune des survénances, mais selon une répartition dans le temps exposée au tableau 3.

Du fait de la répartition différente des règlements de sinistres dans le temps, les deux survénances présentent des niveaux de coûts moyens par délai très hétérogènes. La modélisation des incréments à partir de ces coûts conduirait à des problèmes similaires à ceux évoqués dans le cas des facteurs de développement. Si par contre, le développement des sinistres est mesuré en temps opérationnel, on observe pour les deux survénances les mêmes coûts par sinistre réglé de 1000 en $\tau = \frac{1}{3}$, 2000 en $\tau = \frac{2}{3}$ et 3000 en $\tau = 1$.

Tableau 3 – Prestation cumulée par sinistre réglé

Délai j	nombre de sinistres réglés		Prestations		f_j		Prestation par sinistre réglé		Temps Opérationnel	
	Surv. 1	Surv. 2	Surv. 1	Surv. 2	Surv. 1	Surv. 2	Surv. 1	Surv. 2	Surv. 1	Surv. 2
1	1	2	1000	3000	6	1	1000	1500	33,3%	66,7%
2	2	0	5000	0	1	2	2500	0	100,0%	66,7%
3	0	1	0	3000			0	3000	100,0%	100,0%

²⁰ Dans la littérature anglaise, la quantité S_j est usuellement appelée PPCF, pour « Payment Per Claim Finalised ».

Dans ces conditions, l'approche par les coûts moyens reste valide. Il s'agit en termes de modélisation de substituer l'indice τ à j , en passant à une structure paramétrique du type

$$E(X_{i\tau}) = N_{i\tau} \cdot S_{\tau} \quad (2)$$

où $S_{\tau} = \sum_{i,j \in [u_{\tau}, v_{\tau}]} X_{ij} / N_{ij}$ désigne le coût moyen des sinistres réglés dans le $\tau^{\text{ème}}$ intervalle de temps opérationnel.

Dans le cas des données étudiées, les figures 8 et 9 témoignent de l'effet du temps opérationnel sur l'homogénéité des coûts moyens. Le hamac²¹ de la figure 8 fournit une lecture rapide des propriétés de la distribution des coûts moyens par année de développement présentés en annexe G. La figure 9 est son homologue en temps opérationnel.

La construction de la figure 9 requiert de partitionner l'intervalle $]0,1]$ en sous intervalles $[u_{\tau}, v_{\tau}]$, $\tau = 1, \dots, 15$, avec $u_1 = 0$, $v_{15} = 1$, $u_{\tau+1} = v_{\tau}$. On associe le coût moyen $S_{ij} = \frac{X_{ij}}{N_{ij}}$ de chaque cellule (i, j) du triangle au τ^e sous intervalle $[u_{\tau}, v_{\tau}]$, tel que

$$u_{\tau} < \tau_{ij} \leq v_{\tau}, \text{ où } v_{\tau} = \frac{\tau}{15}.$$

L'observation des figures 8 et 9 conduit à deux enseignements. Tout d'abord, on observe que l'usage du temps opérationnel fait baisser la variabilité intra classe des coûts moyens, plus homogènes sur la figure 9 que sur la figure 8. Ensuite, on voit que leur regroupement selon leur temps opérationnel aboutit à une trajectoire plus régulière.

Par ailleurs, on remarque que la trajectoire des coûts moyens de la figure 9 n'est pas conforme à celle monotone croissante plus usuelle. En effet, il est courant de constater un lien de proportionnalité entre le temps opérationnel et le coût des sinistres. Autrement dit, les « petits » sinistres sont réglés plus rapidement que les « gros », davantage sujets à de longue procédure. En outre, ce postulat souvent vérifié empiriquement est la principale justification du recours au temps opérationnel. On trouve dans la littérature deux sources d'explication possibles de la baisse des coûts moyens dans les premières valeurs de temps opérationnel.

La première est avancée par Taylor (1981) qui rappelle que ce phénomène peut s'expliquer par l'effet de la répartition dans le temps des sinistres réglés sans suite²². Sans intervenir ni sur l'ordre, ni sur la distribution des prestations, ces sinistres sont en effet comptabilisés dans le nombre de règlements intervenant dans le calcul du coût moyen. En supposant que ces règlements interviennent majoritairement dans les premières années, le coût moyen peut donc baisser pour remonter par la suite.

La seconde est exposée par Wright (1992) qui montre que cette baisse peut être liée à l'existence de paiements partiels importants. Cette problématique est développée plus en avant. Nous verrons en fin de partie que la modélisation des paiements partiels développée par Wright fournit dans le cas des données étudiées une explication fortuite à ce phénomène.

²¹ Le point dans la boîte à moustache représente la moyenne, le haut et le bas de la boîte respectivement les 75^{ème} et 25^{ème} quantiles et les extrémités des barres supérieures et inférieures les incréments de valeurs extrêmes.

²² Sinistre dont le règlement n'entraîne pas une indemnisation de l'assureur.

Précisons que l'hypothèse d'invariance de l'ordre reste bien sûr une idéalisation. En l'occurrence, sa pertinence même si elle est avérée par le passé, peut se voir remise en cause d'une année sur l'autre par de nombreux facteurs. Supposons par exemple que sous l'impulsion d'une décision managériale, les règlements de sinistres sans suite interviennent plus tôt. La distribution des prestations par année de développement restera inchangée. A l'inverse, le nombre de règlements effectués dans les premiers intervalles de temps opérationnel augmentera, et diminuera en conséquence sur les derniers intervalles. Dans le cas où les données ne sont pas retraitées de ces sinistres, les coûts moyens baisseront dans les 1^{ers} intervalles par rapport aux coûts moyens historiques.

Figure 8 – Coûts moyens par année de développement

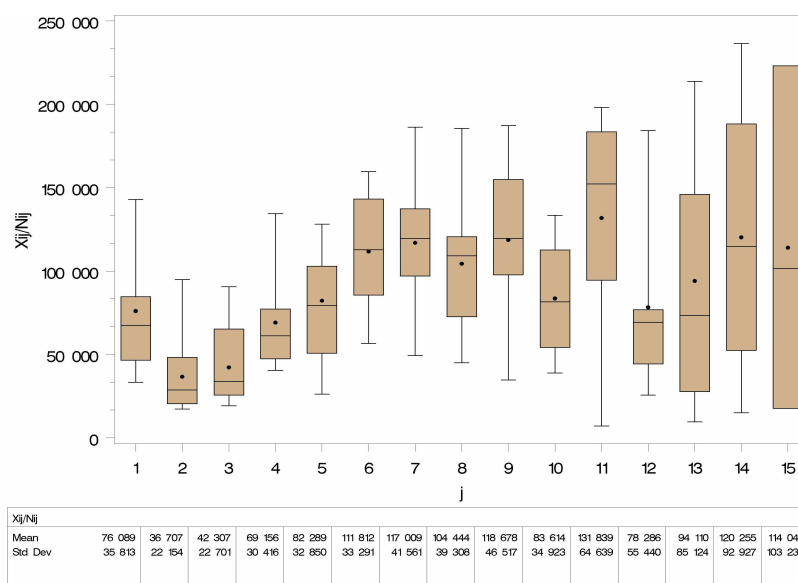
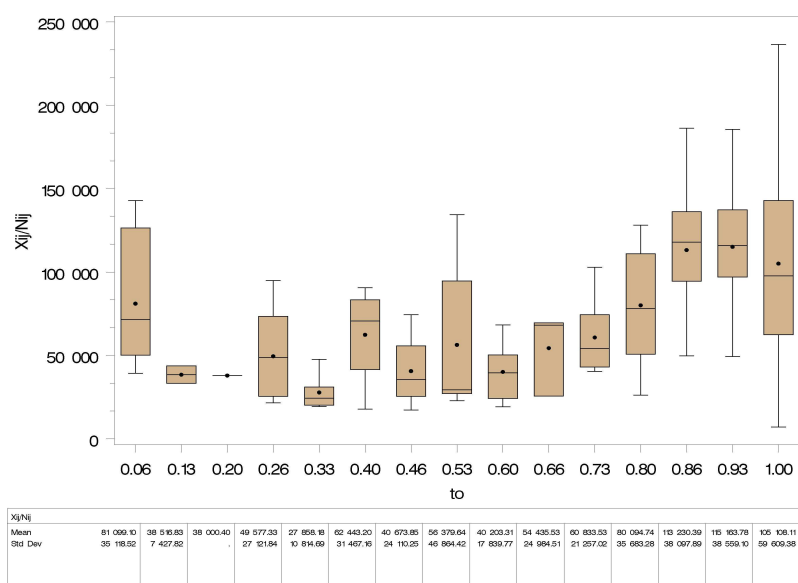


Figure 9 – Coûts moyens par unité de temps opérationnel



III.2 La stationnarité des coûts moyens

L'usage du modèle (2) suppose implicitement la stabilité de la prestation moyenne par unité de temps opérationnel. En particulier, cette prestation est supposée indépendante de l'année de survenance.

Dans le cas étudié, la pertinence de cette hypothèse peut être vérifiée à partir des données du triangle supérieur. Si les coûts moyens de sinistres sont stationnaires par tranche de temps opérationnel, on s'attend à observer avec le temps des fluctuations aléatoires des coûts moyens $S_{i\tau}$ autour de leur espérance S_{τ} .

Pour chaque survenance, on estime des coûts moyens $S_{i\tau}$ aux valeurs de temps opérationnels $v_1, \dots, v_{15}$. Ce calcul est établi par interpolation linéaire à partir des observations de coûts moyens (annexe F) et des valeurs de temps opérationnel associées (annexe G). Les coûts moyens estimés sont présentés aux figures 10 et 11.

Par exemple, le coût moyen de l'année de survenance 1979 (soit $i = 1$) en $\tau = \frac{1}{15}$ est construit à partir des coûts moyens observés $S_{1,0,049}$ et $S_{1,0,233}$. On l'estime à

$$S_{1,\frac{1}{15}} = 130534 + \frac{(94872 - 130534) \times (\frac{1}{15} - 0,049)}{0,233 - 0,049} = 127109$$

Figure 10 – Evolution des coûts moyens $S_{i\tau}$, $\tau = \frac{1}{15}$ à $\tau = \frac{8}{15}$

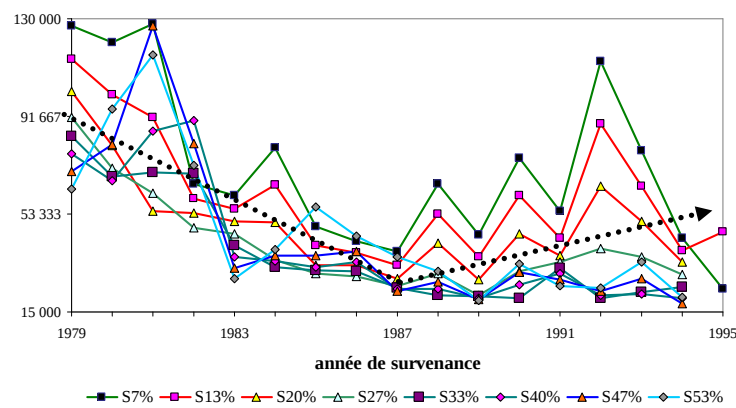
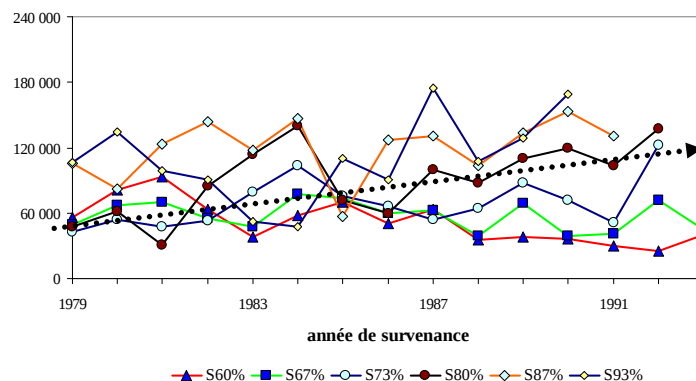


Figure 11 – Evolution des coûts moyens $S_{i\tau}$, $\tau = 0,6$ à 1



Les figures 10 et 11 révèlent l'existence de tendances dans l'évolution temporelle des coûts moyens par sinistre réglé :

- Les coûts moyens en figure 10 suivent la même évolution constituée d'une tendance décroissante depuis 1979 jusqu'en 1987 suivie d'une autre tendance croissante.
- Pour les valeurs élevées de τ présentées à la figure 11, on observe une tendance générale à la hausse des coûts moyens.

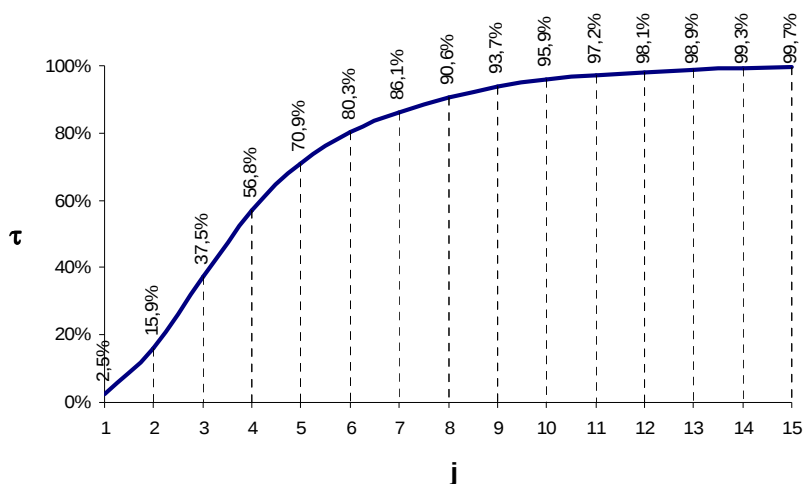
En ce sens, l'hypothèse de stabilité de S_τ est infirmée, dès lors que l'évolution des coûts dans le temps n'est pas stationnaire. Par ailleurs, l'existence de tendances ne peut a priori être imputée à l'inflation, retraitée par Taylor dans le triangle de liquidation. Ce phénomène dit de surinflation des coûts est modélisé par la suite.

III.3 La partition du temps opérationnel

L'approche par les coûts moyens nécessite de transformer les données originales X_{ij} et N_{ij} en un triangle $S_{ij} = X_{ij}/N_{ij}$ de prestations moyennes de sinistres. A ce sujet, le triangle des coûts moyens figurant en annexe G a déjà été présenté dans ce chapitre.

Par ailleurs, l'usage du temps opérationnel requiert d'associer à chacun des coûts moyens un intervalle de temps. La segmentation choisie doit permettre de définir des profils de coûts relativement homogènes à partir d'un nombre suffisant d'observations. Pour nous aider dans notre choix, nous traçons à la figure 12 la courbe du temps opérationnel moyen observé par année de développement.

Figure 12 – Temps opérationnel moyen par année de développement



On remarque un développement du temps opérationnel en deux périodes, avec une inflexion du rythme des règlements en $j = 5$ après une rapide progression. Cette observation rappelle la forte proportion des observations à temps opérationnel élevé dans le triangle de liquidation²³.

²³ Par exemple, 100 des 150 observations de coûts moyens ont une valeur de temps opérationnel supérieure à 70%.

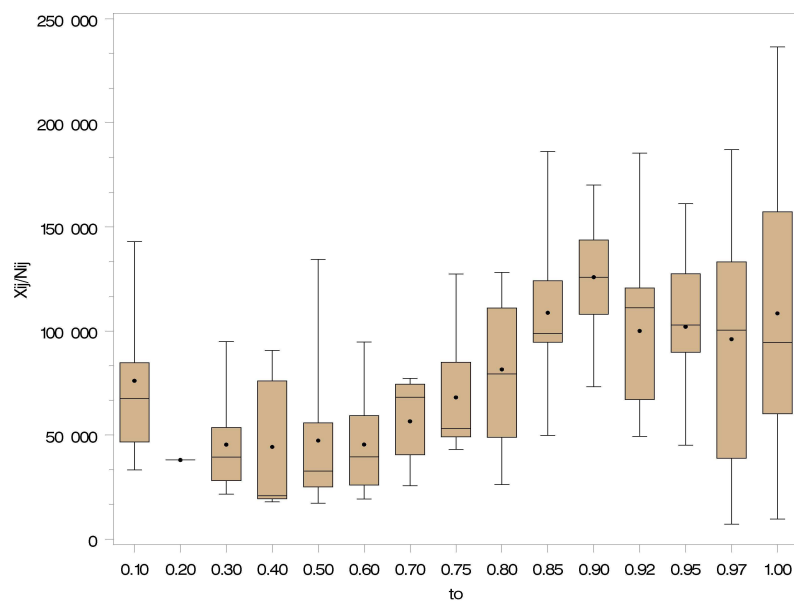
Cette caractéristique permet de segmenter davantage le temps opérationnel pour les valeurs élevées de τ , tout en préservant dans chaque classe de temps un nombre raisonnable d'observations. Cette démarche doit permettre d'affiner l'axe de projection des coûts moyens pour les valeurs élevées de temps opérationnel. En particulier, il s'agit d'estimer à partir de quelle valeur de temps opérationnel les observations de coûts moyens amorcent une décroissance. Une partition jugée satisfaisante est présentée au tableau 4.

Tableau 4 – Intervalles de temps opérationnel

τ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
v_τ	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,75	0,8	0,85	0,9	0,925	0,95	0,975	1

La figure 13 présente le profil des coûts moyens regroupés selon la nouvelle échelle de temps opérationnel. L'hypothèse d'une tendance à la baisse des coûts moyens à partir de la valeur de temps opérationnel de $\tau = 0,9$ est confirmée.

Figure 13 – Coût moyen et temps opérationnel - Echelle finale



III.4 Le lien moyenne - variance

Dans le cadre des GLM, nous estimons les paramètres du modèle de provisionnement par maximum de vraisemblance. Ce calcul est rendu possible en spécifiant une loi de distribution pour chacune des observations de coûts moyens S_{it} . Par les GLM, nous disposons essentiellement des lois dites de type exponentiel liant l'espérance des variables modélisées à leur variance par la relation

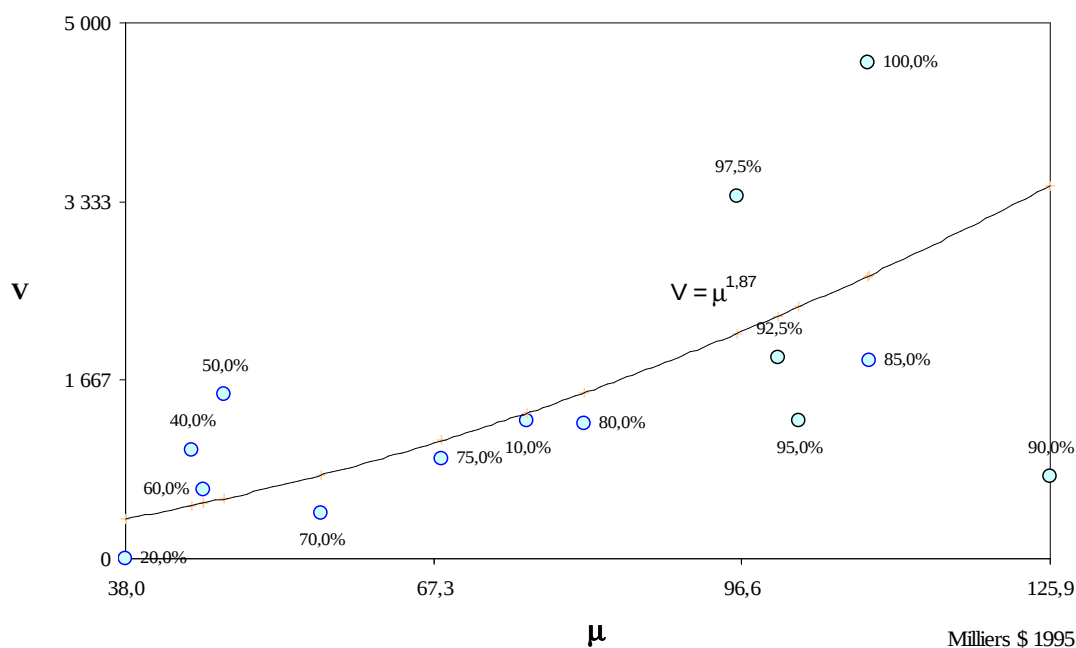
$$\text{Var}[S_{it} | \tau] = E[S_{it} | \tau]^p, \text{ pour } p \in \mathbb{R}^+,$$

soit en simplifiant la notation : $V_\tau = \mu_\tau^p$.

Malgré les tendances temporelles observées sur les coûts moyens, nous supposons sur chaque intervalle de temps que les observations des coûts moyens représentent les réalisations d'une unique variable aléatoire S_τ , dont l'espérance et la variance sont estimées par la moyenne empirique et la variance corrigée.

Ainsi, après avoir associé un intervalle de temps opérationnel à chaque coût moyen S_{ij} , on calcule la moyenne et la variance empirique des coûts moyens de chaque intervalle. Les couples moyenne-variance sont représentés à la figure 14. La liaison croissante entre les deux variables déjà révélée par la figure 9 est confirmée.

Figure 14 – Liaison moyenne - variance (milliers de dollars)



Afin d'éprouver graphiquement l'hypothèse du lien moyenne-variance, nous ajustons, à partir de l'algorithme de Newton²⁴, la courbe $V = \mu^p$ aux couples (V_τ, μ_τ) en minimisant la quantité

$$\sum_{\tau} N_{\tau} \cdot (V_{\tau} - \mu_{\tau}^p)^2,$$

où V_{τ} , μ_{τ} et N_{τ} représentent respectivement la variance corrigée, la moyenne arithmétique et le nombre d'observations associées à la valeur τ . On aboutit à la valeur $p = 1,87$.

IV. Les conséquences sur le choix du modèle

Rappelons les conclusions issues de l'analyse qui vient d'être menée :

- L'utilisation d'une cadence moyenne de paiements a été jugée inadéquate (II.2.1) du fait de l'instabilité des rapports historiques de développement. En illustration, la volatilité des PSAP issue du calcul Chain Ladder a été mesurée (II.2.2).
- L'inadéquation des méthodes par facteurs a été imputée (II.3) à la liaison entre le volume des paiements et le nombre des règlements instables en cadence.
- Fort de ce constat, nous avons choisi de modéliser les incréments à partir des coûts moyens par unité de temps opérationnel. Il a été montré (III.1) que la substitution du temps opérationnel à l'année de développement aboutissait à des coûts moyens plus homogènes.
- Du fait des tendances surinflationnistes observées, l'hypothèse de stationnarité des coûts moyens par intervalle de temps opérationnel n'a pu être retenue (III.2).

Le modèle de provisionnement présenté dans les sections suivantes est développé dans le cadre des modèles généralisés (GLM). Compte tenu des caractéristiques des données qui viennent d'être rappelées, plusieurs points militent en faveur de son adoption.

La possibilité de modéliser les coûts moyens à partir du temps opérationnel constitue la principale opportunité permise par le modèle. En outre, il n'est pas nécessaire de supposer l'hypothèse de stabilité dans le temps des coûts moyens par intervalle de temps opérationnel.

La nature paramétrique du modèle ouvre également des axes d'analyse intéressants. En particulier, si l'introduction du temps opérationnel vise à distinguer des profils de sinistres (en l'occurrence, les petits des gros), il peut être informatif au regard des figures 10 et 11 d'étudier la surinflation des coûts moyens en segmentant l'analyse par niveau de temps opérationnel. Enfin, le calcul nécessaire des facteurs de queue se déduit naturellement du lien modélisé entre les coûts moyens et le temps opérationnel.

²⁴ Les principes de l'algorithme sont rappelés en annexe. Le calcul est mené sous Solveur Excel.

CHAPITRE II. LA MODELISATION DES COÛTS MOYENS

I. La modélisation par les Moindres Carrés Ordinaires

I.1 Le modèle de régression

On considère les $r = 1, \dots, m$ observations de coûts moyens du triangle supérieur, chacune caractérisée par un 3 - upplet $(i_r, t_r = i_r + j_r - 1, \tau_r)$, où i_r , t_r et τ_r sont respectivement l'année de survenance, l'année calendaire et la valeur de temps opérationnel associée à l'observation r .

On définit les variables disjonctives suivantes²⁵ :

- T variables $z_{k,k=1,\dots,T}^{to}$ explicatives de l'intervalle de temps opérationnel associé à l'observation r , telles que $z_k^{to} = 1$, si $u_k < \tau_r \leq v_k$, 0 sinon,
- I variables $z_{k,k=1,\dots,I}^s$ explicatives de la survenance, telles que $z_k^s = 1$ si $i_r = k$, 0 sinon,
- V variables $z_{k,k=1,\dots,V}^c$ liées aux années calendaires, telles que $z_k^c = 1$ si $t_r = k$, 0 sinon.

On modélise le logarithme de la $r^{\text{ème}}$ observation de coût moyen par

$$\ln S_r = \sum_{k=1}^T \xi_k^{to} z_{r,k}^{to} + \sum_{k=1}^I \xi_k^s z_{r,k}^s + \sum_{k=1}^V \xi_k^c z_{r,k}^c + \varepsilon_r \quad (1)$$

Où les termes résiduels ε_r sont supposés normaux, indépendants et d'espérance nulle et $z_{r,k}^{to,c,s}$ désigne la variable $z_k^{to,c,s}$ associée à la r^{e} observation.

Les $m = 150$ équations du modèle s'écrivent

$$\begin{pmatrix} \ln S_1 \\ \ln S_1 \\ \vdots \\ \ln S_m \end{pmatrix} = \begin{pmatrix} \underline{z}_1^{to} & \cdots & \underline{z}_T^{to} & \underline{z}_1^c & \cdots & \underline{z}_V^c & \underline{z}_1^s & \cdots & \underline{z}_I^s \\ z_{1,1}^{to} & \cdots & z_{1,T}^{to} & z_{1,1}^c & \cdots & z_{1,V}^c & z_{1,1}^s & \cdots & z_{1,I}^s \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ z_{m,1}^{to} & \cdots & z_{m,T}^{to} & z_{m,1}^c & \cdots & z_{m,V}^c & z_{m,1}^s & \cdots & z_{m,I}^s \end{pmatrix} \begin{pmatrix} \xi \\ \xi_1^{to} \\ \vdots \\ \xi_1^s \end{pmatrix} + \begin{pmatrix} \varepsilon \\ \varepsilon_1 \\ \vdots \\ \varepsilon_m \end{pmatrix}$$

soit matriciellement,

$$\ln S = Z\xi + \varepsilon = \eta + \varepsilon, \text{ où}$$

- $Z = (Z^{to}, Z^c, Z^s) \in \mathfrak{R}^{m \times p}$ désigne la matrice des $p = T + I + V$ variables explicatives.
- $\xi = (\xi^{to}, \xi^c, \xi^s) \in \mathfrak{R}^p$ désigne le vecteur colonne des paramètres du modèle
- $\eta \in \mathfrak{R}^m$ désigne le prédicteur linéaire.

²⁵ Dans le cas des données étudiées, on a $I = 17$, $V = 17$ et $T = 15$.

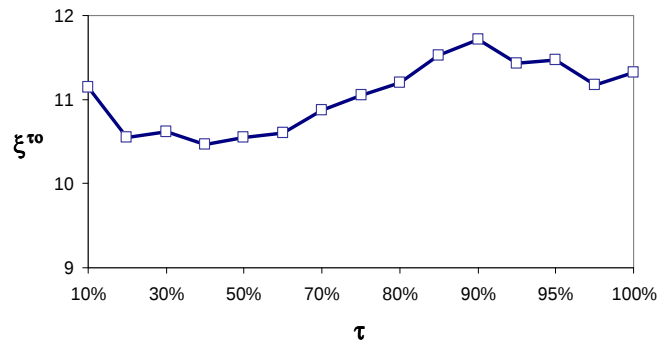
I.2 Les tendances élémentaires

Afin d'identifier les tendances associées au temps opérationnel, on ajuste le modèle

$$\ln S = Z^{\text{to}} \xi^{\text{to}} + \varepsilon, \text{ où } \varepsilon \approx N(0, \sigma^2) \quad (2)$$

aux données du triangle supérieur. Les estimateurs $\hat{\xi}_k^{\text{to}}$ sont représentés sur la figure 15. Pour ce modèle, les estimations des coefficients ξ_k^{to} sont égales à la moyenne arithmétique des $\log S_r$. Il est donc normal de retrouver les tendances observées sur la figure 13.

Figure 15 – Coefficients du temps opérationnel



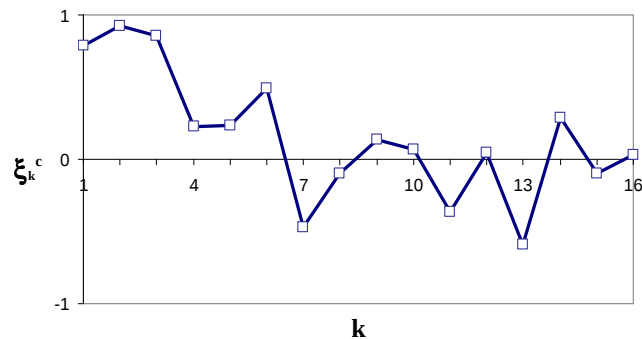
L'ajustement du modèle

$$\ln S = Z^{\text{to}} \xi^{\text{to}} + Z^{\text{c}} \xi^{\text{c}} + \varepsilon \quad (3)$$

permet d'estimer les tendances liées au temps. Les estimations des paramètres ξ_k^{c} présentées à la figure 16 confirment que les coûts moyens restent assujettis à des tendances temporelles, même après l'ajustement des tendances ξ_k^{to} .

Le profil des coefficients²⁶ ξ_k^{c} rappelle l'évolution déjà observée (figure 10) des coûts, constituée d'une première période fortement baissière suivie d'une seconde haussière.

Figure 16 – Coefficients des années calendaires

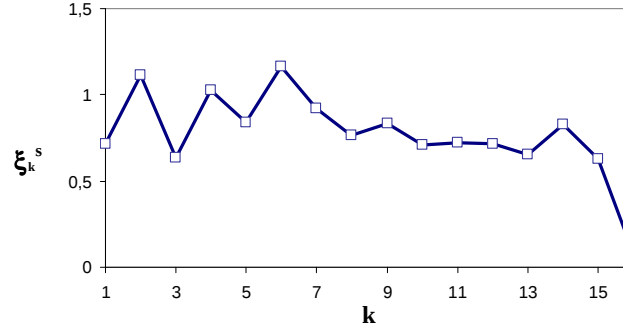


²⁶ $\forall r$, on a $z_{r,17}^{\text{c}} = 1 - \sum_{k=1}^{16} z_{r,k}^{\text{c}}$. Pour des problèmes de colinéarité entre les variables explicatives, on pose $Z_{17}^{\text{c}} = 0$. On a donc 16 variables disjointes explicatives de l'année calendaire d'observation des coûts moyens.

L'ajustement aux données du modèle (4) ne permet pas de distinguer de tendances liées aux années de survénance. Les estimations des paramètres²⁷ ξ_k^s sont présentées à la figure 17

$$\ln S = Z^{\text{to}} \xi^{\text{to}} + Z^s \xi^s + \varepsilon \quad (4)$$

Figure 17 – Coefficients associés aux années de survénances



I.3 Le terme stochastique

Les modèles (2) à (4) supposent les résidus ε_r gaussiens, donc les coûts moyens S_r lognormaux. La conformité de cette hypothèse peut être testée à partir de l'observation du graphe des résidus.

On connaît la variance du résidu observé $e_r = \ln(S_r) - \hat{\mu}_r$:

$$\text{Var}(e_r) = \sigma^2(1 - h_r),$$

où h_r est le coefficient de levier²⁸ de l'observation r et $\hat{\mu}_r$ est l'estimation de $\ln(S_r)$ par le modèle (2). Des résidus bien adaptés à la distribution normale sont les résidus de Student approximativement gaussiens pour de grands échantillons et d'expression

$$e_r^{(\text{st})} = \frac{e_r^2}{s\sqrt{1 - h_r}},$$

où²⁹ $s^2 = \frac{1}{n - p - 1} \cdot \sum_{r=1}^n e_r^2$ est l'estimateur de σ^2 .

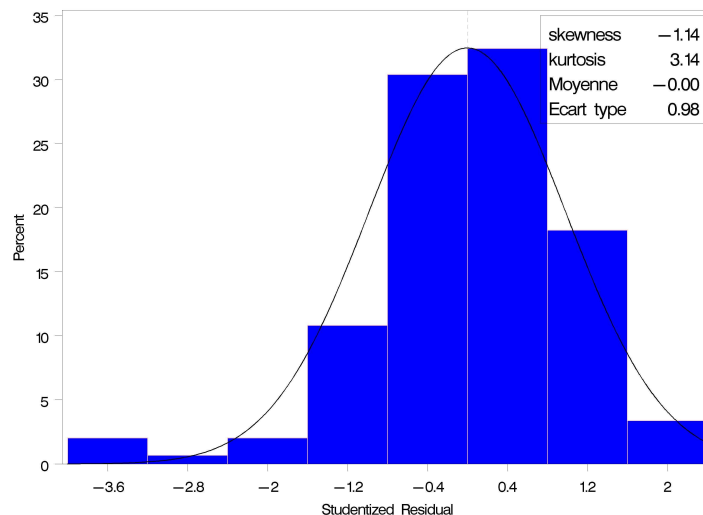
²⁷ La variable disjonctive Z_{17}^s est redondante au sens où $\forall r$, on a $z_{17}^s = \sum_{k=1}^{15} z_k^{\text{to}} - \sum_{k=1}^{16} z_k^s$.

²⁸ Nous explicitons un peu plus loin la forme de h_r dans le cadre des GLM. Pour le moment, retenons simplement que $h_r \in]0, 1[$.

²⁹ Ici, on a $n = 150$ et $p = 15$ (les 15 variables disjonctives Z_k^{to}).

La figure 18 présente les résidus de Student standardisés du modèle (2). Si les coûts moyens S_r sont log normaux, on s'attend à observer des résidus de profil gaussien. Ici, on observe que les résidus sont distribués de façon asymétrique, davantage étendus vers les valeurs inférieures à la moyenne³⁰.

Figure 18 – Répartition des résidus de Student standardisés – modèle (3)



Ce constat suggère que la transformation logarithmique des S_r a « surcorrigé » la queue de distribution des coûts moyens. Autrement dit, les coûts ont une queue de distribution moins longue que celle de la loi log-normale.

II. La modélisation par les Modèles Linéaires Généralisés

II.1 GLM contre MCO

Les modèles linéaires de la section précédente sont développés dans le cadre des Moindres Carrés Ordinaires (MCO). S'ils ont permis d'ajuster les tendances élémentaires des coûts moyens, ils demeurent un outil insuffisant dans l'étude de leur loi de distribution. En effet, le bon usage des MCO repose sur le respect d'une hypothèse difficilement tenable en pratique, celle de la normalité et de l'unicité de la variance des résidus, et par voie de conséquence des variables réponses.

A ce sujet, notons que la transformation logarithmique des données peut dans certains cas être une solution au respect du cadre théorique des MCO. En effet, si les coûts moyens avaient été caractérisés par un même rapport moyenne-variance, le postulat log normal aurait permis de préserver l'hypothèse d'unicité de leur variance³¹.

³⁰ Le coefficient d'asymétrie des résidus est égal à -1,14.

³¹ Par exemple, supposons deux coûts moyens dont les moyennes respectives sont de 1 000 et 30 000, et les écart types de 100 et 6000. La transformation logarithmique permet de considérer deux variables de même variance $\ln[1 + (200/1000)^2] = \ln[1 + (6000/30000)^2] \approx 0,04$

Cependant, dans le cas des données étudiées, la relation croissante décelée entre la moyenne et la volatilité des coûts moyens nous oblige à sortir du cadre des MCO. Pour cette raison, il apparaît plus approprié de recourir aux GLM et aux lois de type exponentiel qui leur sont associées. Ces modèles définissent une relation entre l'espérance du coût moyen et le prédicteur linéaire.

Cette relation est caractérisée par une fonction de lien g telle que

$$E(S_r) = g^{-1} \left(\sum_{k=1}^p z_{r,k} \cdot \xi_k \right), \quad (5)$$

où z_k et ξ_k désignent selon une notation simplifiée les $p = T+I+V$ variables z_k^{to}, z_k^c, z_k^s et les paramètres associés ξ^{to}, ξ^c, ξ^s déjà présentés.

On définit le prédicteur linéaire de la r^e observation

$$\eta_r = \sum_{k=1}^p z_{r,k} \cdot \xi_k.$$

La fonction g permet de relâcher l'hypothèse formulée par les MCO d'une relation linéaire entre $E[S]$ et η . Ou plutôt, la relation reste linéaire, mais sous une certaine échelle. Par exemple, si l'échelle est logarithmique, on passe d'un modèle additif à un autre multiplicatif.

Les modèles linéaires généralisés peuvent être ajustés en utilisant l'algorithme du score de Fisher. L'algorithme maximise la vraisemblance, ou de façon équivalente, minimise la déviance. Cette statistique peut être considérée comme la généralisation de la somme des carrés des résidus d'estimation des MCO³². La généralisation vient du fait que la variance de chaque observation peut être considérée comme une fonction de sa moyenne, évidemment inconnue. L'objectif de l'ajustement du modèle consiste à estimer cette moyenne.

II.2 Introduction aux GLM

II.2.1 La famille des distributions exponentielles

Les coûts moyens S_r sont supposés indépendants et non identiquement distribués. Leur loi de distribution appartient à la famille exponentielle et leur densité est de la forme

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

Les fonctions a , b et c sont définies telles que f soit une densité de probabilité, avec $a(\phi)$ positive et continue, $b(\theta_r)$ convexe et $c(x_r, \phi)$ indépendante de θ_r . Ainsi, chaque observation a un coefficient propre θ_r , mais le paramètre d'échelle ϕ est le même pour toutes les observations.

³² On montre en annexe l'équivalence des deux notions dans le cas de la loi normale.

On suppose par la suite que les fonctions a , b et c sont les mêmes pour toutes les observations. Autrement dit, chaque coût moyen est modélisé à partir de la même loi, mais avec un paramètre canonique propre θ_r lié à sa moyenne.

Les paramètres réels θ_r et ϕ résument l'information relative à la moyenne et à la variance des coûts moyens. En effet, on peut montrer que pour ces distributions, on a

$$E(S_r) = \mu_r = b'(\theta_r) \quad (6)$$

$$\text{Var}(S_r) = \phi \cdot b''(\theta_r) \quad (7)$$

où la valeur de $\phi > 0$, si elle est inconnue, doit être estimée à partir des données³³.

Si comme c'est le cas des lois classiques, l'expression de l'inverse de $b'(\theta_r)$ est connue, on peut exprimer le paramètre θ_r , directement en fonction de la moyenne des coûts avec

$$\theta_r = b'^{-1}(\mu_r)$$

D'après (5), on a également

$$\theta_r = b'^{-1} \left(g^{-1} \left(\sum_{k=1}^p z_{r,k} \cdot \xi_k \right) \right).$$

Cette dernière expression explicite le lien entre la distribution des S_r et les paramètres $\xi_1, \dots, \xi_p$ estimés par maximum de vraisemblance. On remarquera que la fonction c est indépendante des θ_r et n'intervient pas dans l'estimation par vraisemblance. Comme θ_r est fonction de μ_r , on peut considérer que l'équation (7) établit un lien entre l'espérance et la variance des coûts moyens. En ce sens, on réécrit (7) en fonction de la fonction variance V :

$$\text{Var}(S_r) = \phi \cdot V(\mu_r), \quad (8)$$

Les fonctions spécifiques aux distributions de la famille exponentielle sont présentées dans le tableau 5 pour les lois classiques.

Tableau 5 – Fonctions spécifiques d'une densité de type exponentielle

Lois	ϕ	$\mu_r(\theta_r)$	$V(\mu_r)$	$a(\phi)$	$b(\theta_r)$	$c(S_r, \phi/\omega)$
Normale	σ^2	θ_r	1	ϕ	$\theta_r^2/2$	$-\frac{1}{2}(S_r^2/\phi + \ln(2\pi\phi))$
Poisson	1	e^{θ_r}	μ_r	ϕ	e^{θ_r}	$-\ln(S_r!)$
Gamma	$1/\nu$	$-1/\theta_r$	μ_r^2	ϕ	$-\ln(-\theta_r)$	$\frac{1}{\phi} \ln\left(\frac{S_r}{\phi}\right) - \ln(S_r) - \ln\left(\Gamma\left(\frac{1}{\phi}\right)\right)$

³³ On estime ϕ à partir de la déviance ou en ajoutant une contrainte dans le calcul par vraisemblance des valeurs des paramètres du modèle.

En résumé, les lois de type exponentiel possèdent deux caractéristiques majeures. D'une part, chaque distribution est entièrement spécifiée par son espérance et sa variance. D'autre part, la variance de la variable réponse est fonction de son espérance. A partir de (6) et (7), on déduit que l'espérance et la variance du coût moyen sont liées par la relation

$$\text{Var}(S_r) = \phi \cdot E(S_r)^p,$$

où les valeurs $p=0, 1, 2$ correspondent respectivement aux lois normale, poisson et gamma.

II.2.2L'estimation des paramètres

Le vecteur $\underline{\xi} = (\xi_1, \dots, \xi_p)$ des paramètres du modèle est estimé en maximisant la fonction de vraisemblance

$$l(\underline{\theta}, \phi \mid \underline{s}) = \ln L(\underline{\theta}, \phi \mid \underline{s}) = \sum_r \ln[f(S_r, \theta_r, \phi)] = \sum_r l_r$$

où $l_r = \frac{1}{\phi} (S_r \theta_r - b(\theta_r)) + c(S_r, \phi)$ est la contribution du coût S_r à la vraisemblance.

De $\theta_r = b'^{-1} \left(g^{-1} \left(\sum_{k=1}^p z_{r,k} \cdot \xi_k \right) \right)$ obtenu par (5) et (6), on déduit que

$$l(\underline{\theta}, \phi \mid \underline{s}) = \sum_r \ln \left[f \left(S_r, b'^{-1} \left(g^{-1} \left(\sum_{k=1}^p z_{r,k} \cdot \xi_k \right) \right), \phi \right) \right]$$

On remarque donc que le système d'équations constituant la fonction de vraisemblance est à maximiser sur $\underline{\xi}$. On obtient le vecteur $\underline{\xi}$ en annulant le vecteur score $U = \frac{\partial l}{\partial \underline{\xi}}$. Autrement dit, les paramètres $\xi_1, \dots, \xi_p$ sont solutions³⁴ du système d'équations

$$\frac{\partial}{\partial \xi_k} l(\underline{\theta}, \phi \mid \underline{s}) = \sum_r \frac{\partial}{\partial \xi_k} l_r(\theta_r, \phi \mid S_r) = 0 \quad \forall k = 1, \dots, p.$$

On remarque que $l_r(\theta_r, \phi \mid S_r) = l_r \circ \theta_r \circ \mu_r \circ \eta_r(\underline{\xi})$ est une fonction composée, où

$$\begin{cases} \mu_r = b'(\theta_r) \Rightarrow \frac{\partial}{\partial \theta_r} \mu_r = b''(\theta_r) \Rightarrow \frac{\partial \theta_r}{\partial \mu_r} = \frac{1}{b''(\theta_r)} \\ \eta_r = g(\mu_r) \Rightarrow \frac{\partial}{\partial \mu_r} \eta_r = g'(\mu_r) \Rightarrow \frac{\partial \mu_r}{\partial \eta_r} = \frac{1}{g'(\mu_r)} \\ \eta_r = \xi_1 z_{r,1} + \dots + \xi_p z_{r,p} \Rightarrow \frac{\partial \eta_r}{\partial \xi_k} = z_{r,k} \end{cases}$$

³⁴ On peut montrer la stricte concavité de la fonction log-vraisemblance: l'extremum trouvé est donc unique et correspond à un maximum.

En appliquant la règle³⁵ de calcul des dérivées des fonctions composées, on obtient

$$\frac{\partial}{\partial \xi_k} l_r = \frac{\partial l_r}{\partial \theta_r} \cdot \frac{\partial \theta_r}{\partial \mu_r} \cdot \frac{\partial \mu_r}{\partial \xi_k} \cdot \frac{\partial \eta_r}{\partial \xi_k} = \left(\frac{S_r - b'(\theta_r)}{\phi} \right) \cdot \frac{z_{r,k}}{V(\mu_r) \cdot g'(\mu_r)}$$

Comme $\mu_r = b'(\theta_r)$, on obtient l'expression des p équations à résoudre :

$$\frac{\partial l(\underline{\xi}, \phi | \underline{S})}{\partial \xi_k} = \sum_k \frac{\partial l_r}{\partial \xi_k} = \sum_k \frac{(S_r - \mu_r)}{\text{var}(S_r) \cdot g'(\mu_r)} \cdot z_{r,k} = 0, \quad k = 1, \dots, p.$$

Ainsi, en considérant les matrices diagonales W et $N = \frac{\partial \eta}{\partial \underline{\mu}} \in \Re^{m \times m}$ telles que

$$\begin{cases} W(\underline{\xi}) = \left\{ \text{var}(S_r) \cdot g'(\mu_r)^2 \right\}_{r=1, \dots, m} = \left\{ \phi V(\mu_r) g'(\mu_r)^2 \right\}_{r=1, \dots, m} \\ \frac{\partial \eta}{\partial \underline{\mu}} = \left\{ \frac{\partial \eta_r}{\partial \mu_r} \right\}_{r=1, \dots, m} = \left\{ g'(\mu_r) \right\}_{r=1, \dots, m} \end{cases},$$

on remarque que les équations du maximum de vraisemblance pour $\underline{\xi}$ peuvent s'écrire

$$Z' W(\underline{\xi})^{-1} \frac{\partial \eta}{\partial \underline{\mu}} (S - \underline{\mu}) = 0$$

Ce système d'équations n'est pas linéaire en $\underline{\xi}$ qui intervient à la fois dans les matrices W , N , et dans le vecteur $\underline{\mu}$. Une façon de le résoudre consiste à utiliser l'algorithme itératif des scores de Fisher, une variante de l'algorithme plus connu de Newton Raphson³⁶.

De l'expression de $\hat{\underline{\xi}} = (\hat{\xi}_1, \dots, \hat{\xi}_p)'$, on déduit finalement celle de

$$\hat{\underline{\mu}}_r = g^{-1} \left(\sum_{k=1}^p z_{r,k} \cdot \hat{\xi}_k \right),$$

puis de

$$l(\underline{\theta}, \phi | \underline{s}) = \sum_r \ln \left[f \left(S_r, b'^{-1} \left(\hat{\underline{\mu}}_r \right), \phi \right) \right].$$

³⁵ Si $y = f \circ g(x) = f[g(x)]$, on a $f'[g(x)] = f'[g(x)] \times g'(x)$. En posant $u = g(x)$, on écrit $\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$.

³⁶ Nous utilisons la procédure GENMOD du logiciel SAS qui utilise l'algorithme du score de Fisher détaillé en annexe.

II.2.3 La déviance

On construit un modèle *saturé*, de mêmes distribution et fonction de lien que le modèle estimé, mais doté de m variables explicatives. Ce modèle ajuste parfaitement les coûts moyens. On mesure l'adéquation du modèle à partir de la déviance normalisée

$$D^* = \frac{D}{\phi} = 2 \left[\log(\tilde{L}) - \log(\hat{L}) \right],$$

où $\tilde{L}$ est la vraisemblance du modèle saturé et $\hat{L}$ celle du modèle estimé.

Le test de la déviance procède de la même volonté que celui du ratio de vraisemblance : il doit conduire à rejeter les modèles dont le niveau de vraisemblance est « trop éloigné » du niveau maximal associé au modèle saturé.

Autrement dit, le modèle est d'autant mieux ajusté aux données que la déviance est faible. De façon plus formelle, on définit les hypothèses nulle et alternative du test

$$\begin{cases} H_0 : \xi_k = 0 \quad \forall k > p \\ H_1 : \exists k > p \mid \xi_k \neq 0 \end{cases}, k = 1, \dots, n.$$

Si le modèle estimé est le « bon modèle », la déviance sera répartie sur R^+ mais avec plus de chance d'avoir des valeurs proche de zéro. Pour quantifier cette notion, nous utilisons la loi de la déviance sous H_0 .

Comme la déviance est une différence de log-vraisemblances entre deux modèles emboîtés, il découle³⁷ que la statistique D^* suit asymptotiquement une loi du $\chi^2(m - p)$.

Le test se déroule alors classiquement :

- On calcule l'observation D_{obs} de D^* .
- On calcule le quantile de niveau $(1 - \alpha)$ de la loi du $\chi^2(m - p)$ noté $q_{1 - \alpha}$:
 - On rejette l'hypothèse H_0 si $D_{\text{obs}} > q_{1 - \alpha}$
 - Dans le cas inverse, le modèle estimé est considéré « adéquat ».

Notons que cette statistique ne permet la comparaison des modèles que s'ils sont issus d'une même composante aléatoire (V, ϕ) . Dans notre cas, la déviance permettra dans le chapitre suivant de sélectionner le meilleur modèle parmi différentes structures paramétriques.

Par ailleurs, le test de la déviance ne donne qu'une information sur la qualité globale d'ajustement du modèle aux données. En ce sens, il ne dispense pas de l'analyse des résidus qui intervient en parallèle pour obtenir une information plus précise.

³⁷ Ce résultat est principalement lié aux propriétés asymptotiques du maximum de vraisemblance. On pourra se reporter à Saporta (2006) pour une démonstration dans le cas du ratio de vraisemblance.

II.2.4 Les résidus

En un point d'observation r , l'écart entre l'observation et l'estimation de coût moyen est le résidu brut :

$$e_r = S_r - \hat{\mu}_r.$$

Ces résidus e_r sont vus comme les réalisations des « vraies » erreurs inconnues

$$\varepsilon_r = S_r - E(S_r).$$

Dans le calcul des e.m.v des paramètres du modèle, nous avons vu que les hypothèses faites sur les ε_r consistaient à les supposer indépendants et distribués selon un type de loi à définir. Ainsi, si le modèle est approprié aux données, les résidus observés e_r doivent refléter les propriétés des vraies erreurs inconnues ε_r .

Toutefois, l'interprétation des résidus bruts peut être délicate en cas de non normalité des données. En effet, comme ces résidus ne sont qu'une simple transformation linéaire des observations, n'importe quelle caractéristique de non normalité des données se répercute directement sur les e_r . En particulier, les résidus ε_r n'ont pas toujours la même variance.

Pour tenir compte de la variabilité des résidus, on préfère utiliser des quantités standardisées comme les résidus de Pearson définis par

$$e_r^{(p)} = \frac{S_r - \hat{\mu}_r}{\sqrt{V(\hat{\mu}_r)}}.$$

Peter K. Dunn et Gordon Smyth³⁸ ont montré d'une part que la moyenne des résidus de Pearson est nulle, et d'autre part que si leur variance tend vers 0, leur distribution tend vers une loi normale.

Par définition, on standardise le résidu e_r par la variance théorique de S_r , fournie par la fonction de variance $V(\cdot)$ et donc par le modèle. Ce n'est pas la variance de l'estimation $\hat{\mu}_r$ qui est un estimateur donc aléatoire. Pour essayer d'obtenir des résidus de même variance approximative, on utilise les résidus de Pearson ajustés

$$e_r^{(pa)} = \frac{e_r^{(p)}}{\sqrt{V(1 - h_r)}}$$

où h_r est le levier de l'observation r , soit le r^e élément diagonal de la matrice H ³⁹.

³⁸ Dunn P et Smyth Gordon K. (1996), « Randomized Quantile Residuals », Journal of Computational and Graphical Statistics, volume 5, issue 3, 236-244.

³⁹ On rappelle en annexe I que l'algorithme du score équivaut à procéder à une régression linéaire à chaque étape et s'arrête à l'étape finale où le processus est stationnaire : $\underline{\xi} = (Z'W^{-1}Z)^{-1} \cdot Z'W^{-1}\underline{Y}$. Autrement dit, on a $\hat{\underline{\mu}} = Z \cdot \underline{\xi} = H \cdot \underline{Y}$, où H est la matrice de projection de Y sur Z (la forme du vecteur Y est spécifiée en annexe).

Les résidus de déviance standardisés constituent une alternative aux résidus de Pearson. Ils sont définis par

$$e_r^{(DS)} = \frac{\text{sgn}(S_r - \hat{\mu}_r) \cdot \sqrt{d_r}}{\sqrt{\hat{\phi}(1 - h_r)}},$$

où d_r est la contribution de la $r^{\text{ième}}$ observation à la déviance⁴⁰.

Cette définition est moins naturelle. En théorie, ces derniers résidus sont approximativement distribués selon une loi normale standard. Si tel n'est pas le cas, les conclusions du test de déviance peuvent devenir vides de sens.

II.3 Les hypothèses du modèle GLM

II.3.1 Le choix de la distribution des résidus

Pour le moment, nous avons observé une distribution des résidus se situant en termes de queue de distribution en deçà de la loi log normale (section I.3). Pour estimer la distribution la plus adéquate, nous utilisons le modèle (9) :

$$E[S] = \exp[Z^{\text{to}} \xi^{\text{to}} + Z^{\text{c}} \xi^{\text{c}}] \quad (9)$$

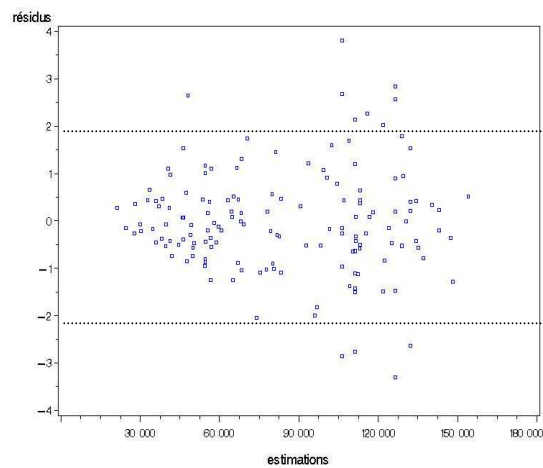
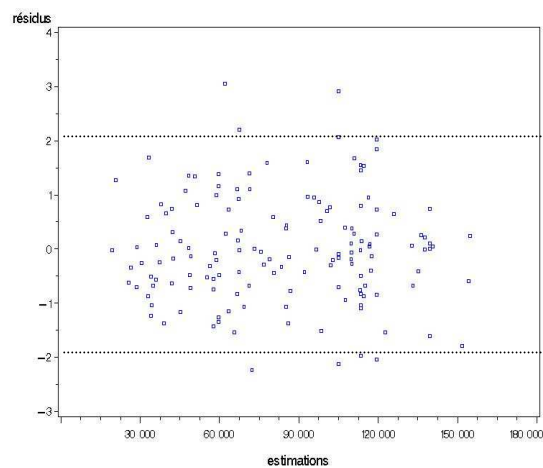
Ce modèle est inspiré du modèle (3), au sens où il partage la même structure paramétrique. En revanche, il ajuste les paramètres du modèle à partir des S_r au lieu des logarithmes népériens des S_r tout en exploitant une fonction de lien logarithmique.

La pertinence de la fonction de lien est considérée dans la section suivante. L'exercice consiste ici en associant le modèle à une loi de distribution à tester l'adéquation de la loi aux données à partir des résidus de Pearson ajustés issus du modèle (9).

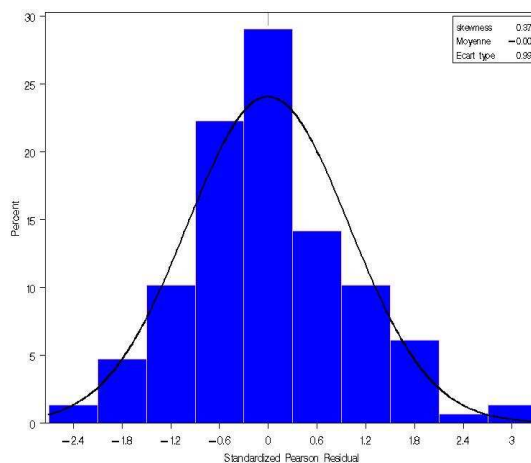
Nous débutons l'expérimentation par la loi Normale. Les résidus issus du modèle GLM - normal sont présentés à la figure 19. Nous continuons l'expérimentation avec la loi gamma et présentons les résidus issus du modèle GLM - gamma à la figure 20. Ces deux figures permettent d'observer empiriquement comment le choix de la fonction variance affecte les résultats des GLM.

Si la forme du modèle est appropriée (distribution, fonction lien), on doit observer des résidus stables en moyenne et en volatilité quelle que soit la valeur des estimations. Hors, il apparaît sur la figure 19 que la volatilité des résidus de Pearson ajustés du modèle GLM - normal croît avec les estimations. Autrement dit, il ne semble pas adéquat d'utiliser une fonction variance gaussienne qui affecte la même volatilité à chaque observation. Ce résultat conforte les conclusions de l'analyse des données établissant un lien croissant entre variance empirique et moyenne des coûts moyens. A l'inverse, l'observation de la figure 20 permet de constater que l'usage de la loi gamma stabilise d'une part les résidus de Pearson ajustés, en moyenne et en volatilité, et d'autre part réduit le nombre et le montant des valeurs élevées.

⁴⁰ L'expression de d_r est rappelée en annexe J.

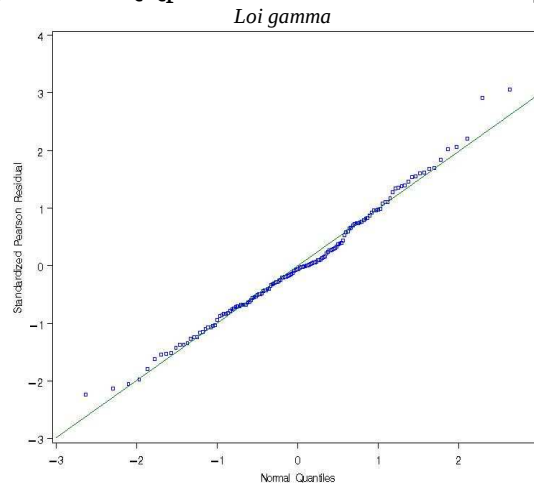
Figure 19 – Résidus en fonction des estimations– loi normale**Figure 20** – Résidus en fonction des estimations – loi gamma

Par ailleurs, on remarque sur la figure 21 que l'utilisation de la loi Gamma corrige le problème d'asymétrie à gauche des résidus détecté en cas d'utilisation de la loi log-normale (figure 18). En témoigne la valeur proche de zéro du coefficient d'asymétrie des résidus. On observe également une distribution des résidus proche de la loi normale standard.

Figure 21 – Histogramme des résidus de Pearson ajustés - Loi gamma

Dans la sélection de la distribution des résidus du modèle, une déviation trop importante des résidus de la distribution normale standard constitue souvent un bon indicateur de l'inadéquation de la distribution choisie. Cette hypothèse est testée graphiquement ci après. Le test est effectué à partir du QQ-plot des résidus présenté à la figure 22. Ce graphique est destiné à vérifier l'adéquation d'une variable à une loi de distribution. Dans notre cas, nous testons donc l'hypothèse de normalité des résidus de Pearson ajustés.

Figure 22 – QQplot des résidus de Pearson ajustés



Le QQ-plot est la représentation graphique des points

$$(\Phi^{-1}(r/m); e_r^{(PS)}),$$

où les $e_r^{(PS)}$ auront été préalablement triés par ordre croissant. Même si on devine sur la figure 22 une queue de distribution des résidus un peu plus épaisse que la loi normale, on observe que l'approximation normale de leur distribution reste satisfaisante.

Comme l'hypothèse de normalité n'est pas infirmée, une nouvelle lecture des figures 19 et 20 est possible. En se fixant un seuil d'acceptation à 5%, on peut en effet considérer que les résidus correspondant à des écarts statistiquement significatifs sont ceux dont la valeur absolue est supérieure à 2. Le choix de cette valeur critique est évidemment motivé par l'approximation de la distribution des résidus ajustés de Pearson par une loi normale centrée réduite. Là encore, l'utilisation de la loi gamma aboutit à des résultats a priori satisfaisants.

Il apparaît donc globalement raisonnable de modéliser les coûts moyens à partir d'une loi gamma. Tout d'abord du fait des bons résultats issus des tests qui viennent d'être présentés. Ensuite, parce que ce choix conforte les conclusions issues de l'analyse des données, où on identifiait la relation $V = \mu^{1,87}$. Enfin, parce que ce choix n'apparaît pas aberrant aux vues des observations de coûts moyens. Ainsi, les incréments du triangle en annexe A et les coûts moyens qui en découlent sont strictement positifs. Il n'apparaît donc pas anormal d'imposer une contrainte de stricte positivité⁴¹ sur les S_r .

⁴¹ Rappelons que ce choix n'est pas trivial, dès lors qu'il est possible d'observer des coûts moyens nuls ou négatifs. Par exemple, dans le cas de traités de réassurance en excédent de sinistres, il est fréquent de ne pas enregistrer de paiements lors d'une année comptable. Il arrive également que les triangles contiennent des incréments négatifs, du fait des recours qui permettent à l'assureur subrogé dans les droits de l'assuré de récupérer la somme versée à son assuré auprès du responsable du dommage.

II.3.2 La fonction de lien

Pour tester l'adéquation de la fonction de lien g , on calcule une approximation de $g(S_r)$. Dans le cas de la fonction de lien logarithmique, on a

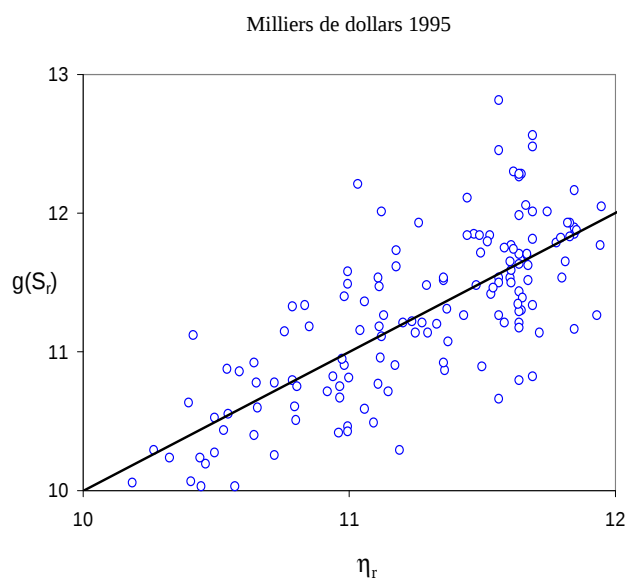
$$g(\mu_r) = \ln \mu_r \Leftrightarrow g'(\mu_r) = \frac{1}{\mu_r}$$

Puisque $\ln \mu_r = \eta_r$, on en déduit l'approximation

$$g(S_r) \cong g(\mu_r) + g'(\mu_r) \cdot (S_r - \mu_r) = \eta_r + \frac{S_r}{\mu_r} - 1 \quad (10)$$

Dans le calcul de l'équation (10), on utilise les estimations $\hat{\eta}_r$ et $\hat{\mu}_r$. Les résultats sont présentés à la figure 23. On observe qu'il existe une relation approximativement linéaire entre l'estimation de $g(S_r)$ et $\hat{\eta}_r$. Le lien logarithmique est donc relativement approprié au modèle.

Figure 23 – Diagnostic sur la fonction de lien logarithmique



CHAPITRE III. LA SPECIFICATION DU MODELE

I. La capture des tendances

I.1 Le rappel du principe de parcimonie

Evidemment, il est impossible de spécifier l'ensemble des variables explicatives de la formation du coût des sinistres. A défaut, les modèles paramétriques de provisionnement utilisent traditionnellement les trois dimensions du triangle : l'année de survenance, l'année calendaire et l'année de développement, remplacée dans notre cas par le temps opérationnel.

Dans le chapitre précédent, l'utilisation de 47 variables disjonctives a permis de modéliser les coûts moyens en fonction de tendances liées à ces trois variables⁴². Malgré la bonne qualité d'ajustement associée à leur utilisation, rappelons que cette approche n'a pas répondu à un principe fondamental en économétrie, celui de parcimonie.

Ainsi, si comme c'est notre cas, le modèle contient « trop » de variables explicatives, les valeurs de ses paramètres capturent de la volatilité, au sens où les données utilisées lors de l'ajustement ne constituent qu'un échantillon du processus à modéliser. On parle de surparamétrisation. La qualité de l'ajustement se fait au détriment de la capacité de prédiction du modèle. A l'inverse, l'utilisation de trop peu de paramètres risque d'aboutir à une estimation biaisée, du fait de l'absence de variables importantes. On parle de sous paramétrisation.

Cette section vise à capturer les tendances du chapitre précédent, tout en réduisant les paramètres sans trop dégrader la qualité d'ajustement du modèle. Dans cette démarche, nous utilisons le modèle de séparation (Zehnwirth, 1994) généralisant le modèle CCI.

I.2 Le modèle de séparation

Considérons une seule année de survenance i . On utilise une notation appropriée à la présentation du modèle de séparation. Les observations (i, j) de coûts moyens sont caractérisées par le couple (t, l) , l désignant le numéro de l'intervalle de temps opérationnel.

On désigne par $y(l) = \ln(S_l)$, pour $l = 2, \dots, T$, le logarithme du coût moyen du l^e intervalle de temps opérationnel et on pose $y(1) = \xi^s$. Selon cette notation, on exprime $y(l)$ par :

$$y(l) = y(1) + y(2) - y(1) + \dots + y(l) - y(l-1) = \xi^s + \sum_{k=1}^{l-1} \xi_k^{to} \quad (10)$$

où $\xi_k^{to} = y(k+1) - y(k)$, pour $k > 0$, désigne ici la différence entre les logarithmes des coûts moyens de deux intervalles de temps opérationnel adjacents.

L'équation (10) définit le modèle déterministe dit *Cap Cod* (CC).

⁴² Les tendances sont représentées aux figures 15 à 17.

L'utilisation de la fonction logarithmique permet d'assimiler la variation ξ_k^{to} au taux de croissance du coût moyen entre les k^e et $(k+1)^e$ intervalles de temps opérationnel :

$$\xi_k^{to} = y(k+1) - y(k) = \ln(S_{k+1}) - \ln(S_k) = \ln\left(\frac{S_{k+1}}{S_k}\right).$$

L'intégration dans l'équation (10) d'une tendance annuelle surinflationniste ξ^c dépendante du temps définit le modèle CCI dit *Cap Cod with Constant Inflation* (CCI) :

$$y(l) = \xi^s + \sum_{k=1}^{l-1} \xi_k^{to} + \xi^c \cdot (t-1) \quad (11)$$

Le modèle dit de séparation est la généralisation probabiliste du modèle CCI. Pour $l > 1$, on modélise le logarithme du coût moyen de la cellule (i, j) par la relation

$$\ln(S_l) = \xi^s + \sum_{k=1}^{l-1} \xi_k^{to} + \sum_{k=1}^{t-1} \xi_k^c + \varepsilon_l, \quad (12)$$

On redéfinit les éléments de la matrice $Z = \{z_{ij}\}$ de la façon suivante :

- pour les colonnes correspondantes aux I paramètres ξ_k^s , on a $z_{ij} = 1$ si $i = k$, 0 sinon.
- pour les colonnes représentant les $(T-1)$ paramètres ξ_k^{to} , on a $z_{ij} = 1$ si $k < l$, 0 sinon.
- pour les colonnes associées aux $(V-1)$ paramètres ξ_k^c , on a $z_{ij} = 1$ si $k < i+j-1$, 0 sinon.

Prenons un exemple pour $I = J = 3$. On suppose que le temps opérationnel est partitionné en trois sous intervalles égaux $]0, \frac{1}{3}]$, $[\frac{1}{3}, \frac{2}{3}]$ et $[\frac{2}{3}, 1]$. On a alors

$$\ln S = \left(\begin{array}{c|c} \ln S_{ij} & \tau_{ij} \\ \hline \ln S_{11} & 0,3 \\ \ln S_{12} & 0,4 \\ \ln S_{13} & 0,5 \\ \ln S_{21} & 0,6 \\ \ln S_{22} & 0,7 \\ \ln S_{13} & 0,1 \end{array} \right) \quad Z = \left(\begin{array}{ccc|cc|cc} \xi_1^s & \xi_2^s & \xi_3^s & \xi_1^{to} & \xi_2^{to} & \xi_1^c & \xi_2^c \\ \hline 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \end{array} \right)$$

Si des tendances sont constantes, il convient de les cumuler en additionnant les colonnes associées de la matrice Z . On peut de cette façon réduire le nombre de paramètres du modèle.

I.3 La recherche des tendances du triangle

La méthode suivie est ascendante. Elle consiste à partir d'un modèle construit sur la base d'un minimum de paramètres, à ajouter autant de variables explicatives pour modéliser les tendances liées aux variables t et τ . A chaque étape, il convient d'être vigilant sur la formation de tendances dans les résidus d'estimation.

Le premier modèle testé est composé de trois variables. Il correspond à l'hypothèse de tendances constantes, où $\xi_k^s = \xi^s$, $\xi_k^{to} = \xi^{to}$ et $\xi_k^c = \xi^c$. Bien que cela soit difficilement concevable, le paramètre ξ^c représentatif des tendances calendaires n'est pas significatif. De ce fait, nous essayons un modèle avec seulement deux paramètres ξ^s et ξ^{to} .

Les résultats sont les suivants :

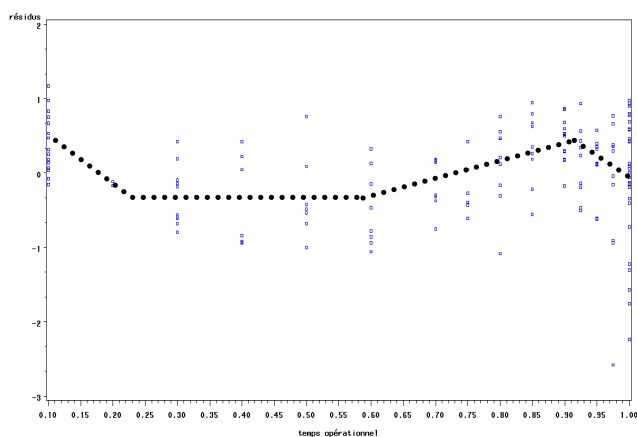
Paramètres	Estimation	Ecart type	p-value
ξ^s	10,7538	0,1050	< 0,0001
ξ^{to}	0,0477	0,0112	< 0,0001

R^2 ajusté= 0,9966 – F – statistique = 21547

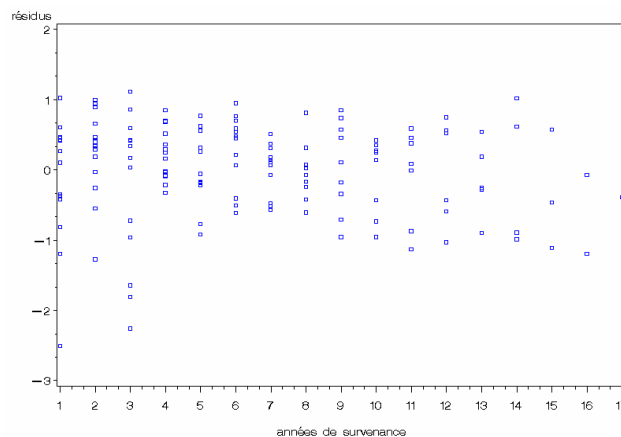
Les tendances observées sur t et τ sont surlignées en figure 24. Sans surprise, elles correspondent à celles révélées par les MCO dans la section précédente (figures 15 et 16).

Figure 24 – Graphiques des résidus – Méthode de Zehnwirth – Etape 2

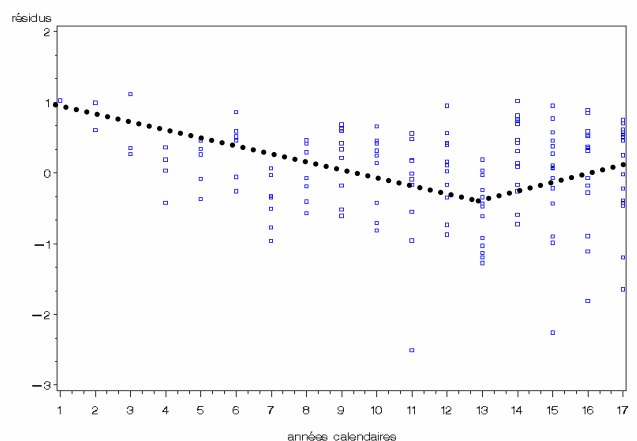
1 : résidus / temps opérationnel



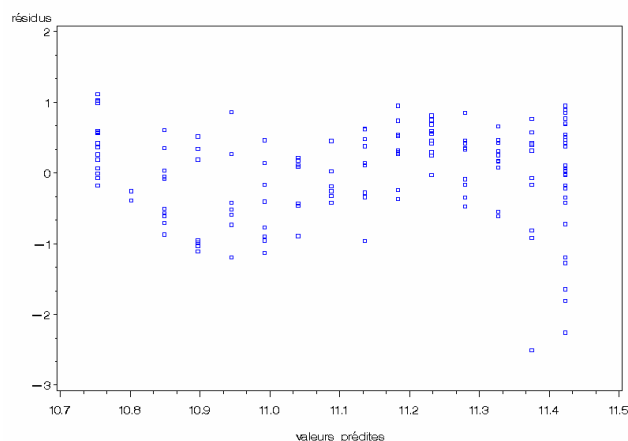
2 : résidus / années de survénance



3 : résidus / années calendaires



4 : résidus / valeurs prédites



Un premier ajustement confirme la nullité et la non significativité du paramètre de régression ξ_2^{to} pour $\tau \in [20\% ; 60\%]$. Après le retrait de la variable associée, on aboutit au résultat final présenté au tableau 6.

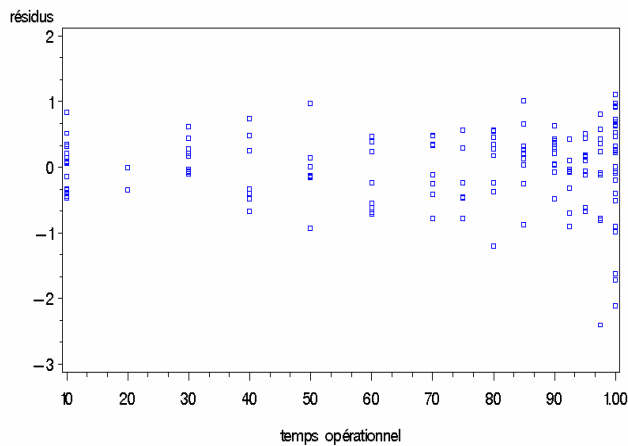
Tableau 6 – modèle de Zehnwirth

Intervalles	Paramètres	Estimateurs	Ecart type	P-value
$i \in]1;17]$	ξ^s	11,5848	0,1923	< 0,0001
$\tau \in]0\%;20\%]$	ξ_1^{to}	-0,5726	0,1724	0,0009
$\tau \in [60\%;90\%]$	ξ_3^{to}	0,2229	0,0305	< 0,0001
$\tau \in]90\% ;100\%]$	ξ_4^{to}	-0,1042	0,0408	0,0107
$t \in]1;13]$	ξ_1^c	-0,0575	0,0189	0,0024
$t \in]13;17]$	ξ_2^c	0,1196	0,0411	0,0036

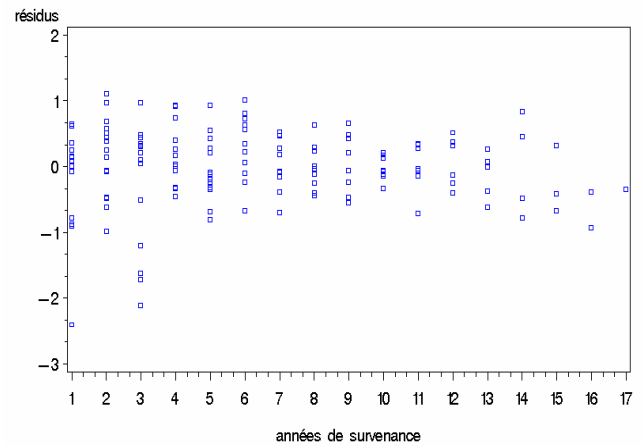
R^2 ajusté. : 0,9974 - F - statistique = 9172,99

Figure 25 – Graphiques des résidus Méthode de Zehnwirth – Etape 3

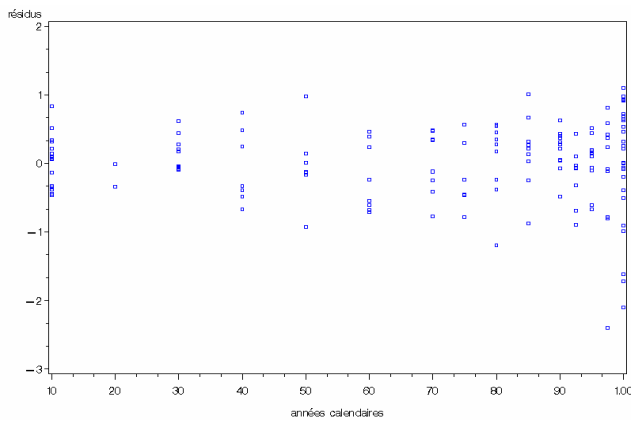
1 : résidus / temps opérationnel



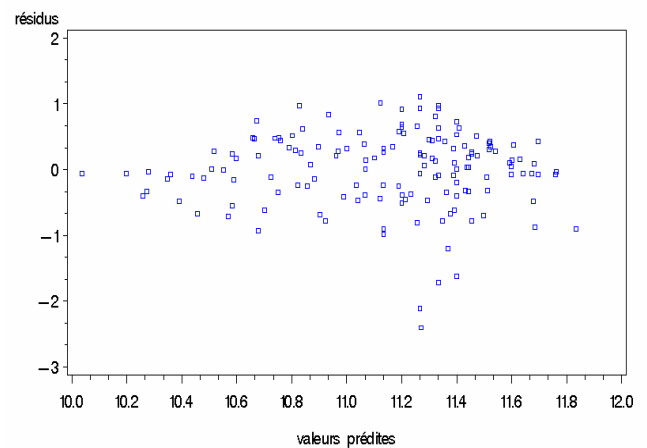
2 : résidus / années de survéance



3 : résidus / années calendaires



4 : résidus / valeurs prédites



La figure 25 suggère que toutes les tendances sur le temps opérationnel, les années de survénance et les années calendaires ont été capturées. En particulier, on estime:

- Une tendance baissière de $e^{-0,1042} - 1 = -9,9\%$ par intervalle de temps opérationnel, pour $\tau > 0,9$.
- Une tendance annuelle haussière de $e^{0,1196} - 1 = +12,7\%$ à partir de 1992 ($t = 13$).

L'identification de ces tendances permet de réécrire l'équation (12) :

$$\begin{aligned} \ln S_l = & + \xi^s + \xi_1^{to} \cdot [1, (l-1)^+]^- + \xi_3^{to} \cdot [5, (l-6)^+]^- + \xi_4^{to} \cdot (l-11)^+ \\ & + \xi_1^c \cdot (12, t-1)^- + \xi_2^c \cdot (t-13)^+ \end{aligned} \quad (13)$$

Par exemple pour la cellule (1,15), on a $\ln(S_{1,15}) = 11,5274$ et $\tau_{1,15} = 0,983$. Selon la partition retenue du temps opérationnel, cette cellule est associée au 15^e intervalle, soit $l = 15$. Par ailleurs, on a $t = 15$.

On estime le logarithme du coût moyen de la cellule (1,15) à partir du couple (t, l) associé et des estimations des paramètres de régression :

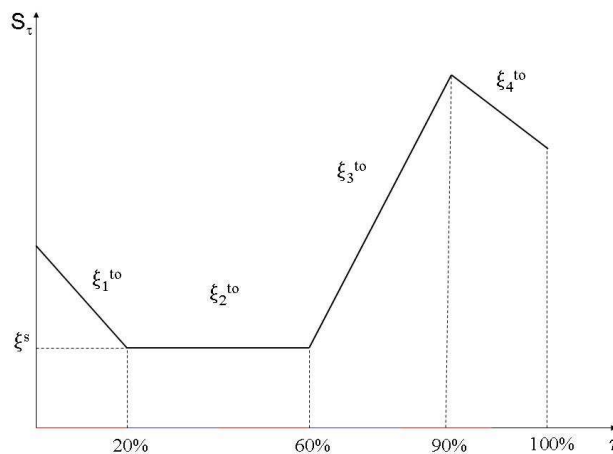
$$\begin{aligned} \ln(S_{1,15}) &= 11,5848 - 0,5726 + 5 \times 0,2229 - 4 \times 0,1042 - 0,0575 \times 12 + 0,1196 \times 2 \\ &= 11,2591. \end{aligned}$$

II. La modélisation des tendances

II.1 Le temps opérationnel

La section précédente a abouti à la détection de quatre tendances relatives au temps opérationnel, et à aucune dans le cas de l'année de survénance. Comme le présente la figure 26, l'évolution des coûts moyens en fonction du temps opérationnel peut être représentée par des droites avec des ruptures de gradient aux points $\tau = 20\%$, 60% et 90% .

Figure 26 – Modélisation des tendances sur i et τ



Autrement dit, il est possible de :

- remplacer les 15 variables constituant la matrice Z^{to} par trois variables continues⁴³.
- remplacer les 16 variables disjonctives constituant la matrice Z^s par le vecteur unité $\underline{1}$.

Reprenons la notation du début de chapitre où τ_r désigne la valeur du temps opérationnel de la $r^{\text{ème}}$ observation de coût moyen. Dans le modèle GLM – gamma, on peut donc trouver une certaine correspondance entre les expressions suivantes :

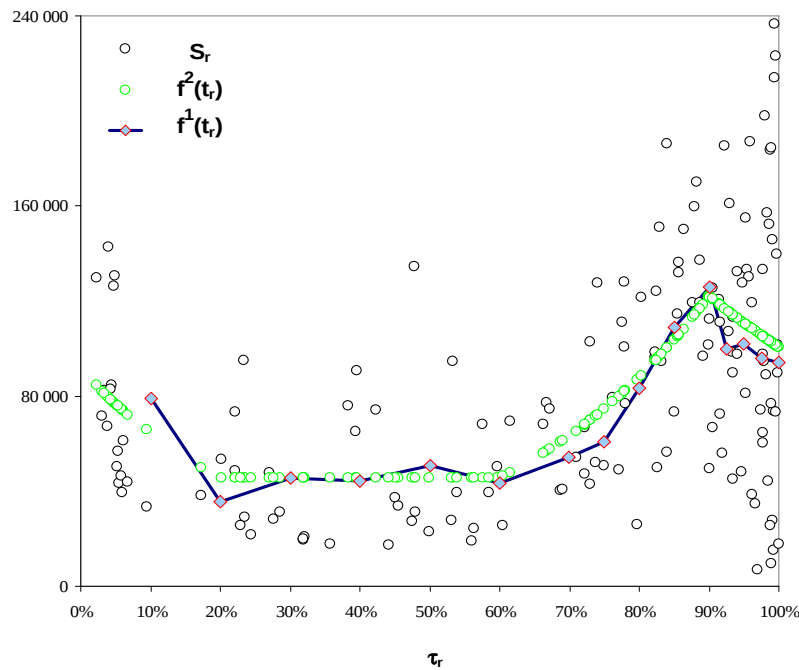
$$E[S_r] = f^1(k_r) = \exp(Z_r^s \underline{\xi}^s + Z_r^{to} \underline{\xi}^{to}), \quad (15)$$

$$E[S_r] = f^2(k_r) = \exp(\xi^s \cdot 1 + \xi_1^{to} \cdot (0,2 - \tau_r)^+ + \xi_3^{to} \cdot (\tau_r - 0,6)^+ + \xi_4^{to} \cdot (\tau_r - 0,9)^+), \quad (16)$$

où Z_r^{to} désigne la $r^{\text{ème}}$ ligne de la matrice Z^{to} relative à l'observation r .

La figure 27 superpose les fonctions $f^1(\tau_r)$ et $f^2(\tau_r)$ ajustées par GLM - gamma. La correspondance entre les deux représentations apparaît satisfaisante.

Figure 27 – Variables continues explicatives du temps opérationnel



Le recours aux variables continues ne dégrade pas significativement la qualité d'ajustement au sens de la déviance, qui passe de 40,70 pour le modèle (15) à 44,52 avec le modèle (16). Par ailleurs, les paramètres de régression sont tous significatifs au sens de leur P-value. Ces résultats indiquent que les quatre paramètres ξ^{to} capturent l'essentiel de l'information contenue par les 15 variables disjonctives du modèle (15).

⁴³ Comme $\hat{\xi}_2^{to} = 0$

II.2 La surinflation

On conçoit que l'inflation affecte différemment chaque branche d'assurance. Autrement dit, il est naturel de penser qu'en plus d'un taux d'inflation général, il existe un facteur additionnel spécifique à chaque branche. Cette notion est appelée par les auteurs surinflation⁴⁴ ou inflation superposée. Dans notre cas, le phénomène de surinflation peut être assimilé à la croissance des observations de coûts moyens calculées sur le triangle déflaté par Taylor. Cette croissance est modélisée à partir des tendances observées sur t . A cet égard, une économie de variables explicatives peut également être effectuée.

Dans cet objectif, deux modèles sont testés. Le premier intègre dans le modèle (16) les tendances révélées par la méthode de Zehnwirth, avec

$$\ln E[S_r] = +\xi^s + \xi_1^{to} \cdot (0,2 - \tau_r)^+ + \xi_3^{to} \cdot (\tau_r - 0,6)^+ + \xi_4^{to} \cdot (\tau_r - 0,9)^+ + \xi_1^c \cdot (13 - t_r)^+ + \xi_2^c \cdot (t_r - 13)^+ \quad (17)$$

On note

$$f^1(t_r) = +\xi_1^c \cdot (13 - t_r)^+ t_r + \xi_2^c \cdot (t_r - 13)^+$$

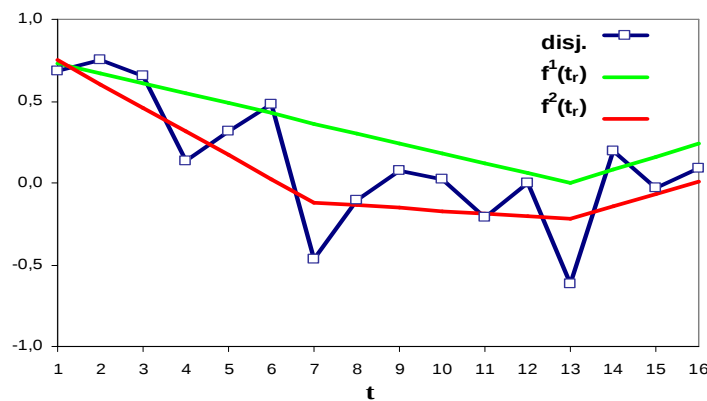
la partie de la forme paramétrique du modèle (17) dédiée à la modélisation de la surinflation. Le modèle (18) teste une structure paramétrique alternative $f^2(\tau_r, t_r)$ avec deux ruptures de gradient aux points $t_r = 7$ et $t_r = 13$. Cela consiste à substituer dans le modèle (17) la fonction $f^1(t_r)$ par la fonction

$$f^2(t_r) = +\xi_1^c \cdot t_r + \xi_2^c \cdot (7 - t_r)^+ + \xi_3^c \cdot (t_r - 13)^+ \quad (18)$$

La figure 28 superpose les fonctions ajustées $f^1(t_r)$ et $f^2(t_r)$ aux coefficients ξ^c estimés à partir des variables disjonctives du modèle (19) :

$$\ln E[S_r] = +\xi^s + \xi_1^{to} \cdot (0,2 - \tau_r)^+ + \xi_3^{to} \cdot (\tau_r - 0,6)^+ + \xi_4^{to} \cdot (\tau_r - 0,9)^+ + Z_r^c \underline{\xi}^c$$

Figure 28 – Tendances surinflationnistes



⁴⁴ Dans le cas des dommages corporels, l'apparition de la surinflation peut être liée à de nombreux phénomènes déjà évoqués tels le renchérissement des frais de soin ou la revalorisation des salaires dans le secteur de l'aide à domicile.

Le modèle (19) produit sensiblement les mêmes estimations des coefficients ξ^c que celles calculées par MCO (figure 16). Les fonctions ajustées $f^1(t_r)$ et $f^2(t_r)$ capturent bien les tendances des coefficients ξ^c estimés à partir des seize variables disjonctives.

La sélection entre $f^1(t_r)$ et $f^2(t_r)$ est faite à partir du critère AIC (Akaike Information Criterion) mettant en balance la précision du modèle évaluée par la déviance D avec la complexité du modèle mesurée par le nombre p de ses variables explicatives.

Avec le critère AIC, on recherche la régression qui minimise la quantité

$$AIC = D + 2 \times p.$$

Le tableau 7 présente les valeurs du critère Akaike des trois modèles considérés. Au sens de cette statistique, on aboutit à deux conclusions. D'une part, l'introduction des seize variables disjonctives n'améliore pas suffisamment la qualité d'ajustement du modèle par rapport aux deux modèles alternatifs. D'autre part, les deux représentations continues sont équivalentes en termes de qualité d'ajustement.

Tableau 7 – AIC des modèles

Représentation des tendances surinflationnistes	Modèle	AIC
Un changement dans le gradient des tendances	17	49,91
Deux changements dans le gradient des tendances	18	49,71
Variables disjonctives	19	72,92

Toutefois, comme la significativité des paramètres ξ^c est meilleure dans le cas du modèle (17), nous retenons ce modèle dont les caractéristiques sont présentées au tableau 8.

Tableau 8 – Modèle avec temps opérationnel, surinflation et année de survénance

Paramètres	Estimateurs	Ecart type	P - value
ξ^s	10,4030	0,1116	< 0,0001
ξ_1^{to}	3,3910	0,9346	0,0003
ξ_3^{to}	3,9140	0,4723	< 0,0001
ξ_4^{to}	-7,8030	1,9464	< 0,0001
ξ_1^c	-0,0607	0,0165	0,0002
ξ_2^c	0,0793	0,0359	0,0273

En passant du modèle composé des 47 variables disjonctives au modèle (17), le recours aux tendances sur t et τ permet de réduire le nombre de variables explicatives de 47 à 6, en améliorant le critère AIC de 123,12 à 49,91.

II.3 L'interaction entre le temps opérationnel et la surinflation

Les tendances temporelles estimées sur les coûts moyens par le modèle (17) peuvent être assimilées à des taux de surinflation. A ce sujet, le modèle estime que les coûts moyens déflatés croissent annuellement de 8,25% par an à partir de 1991 ($t = 13$). Si cette évolution reste envisageable dans le cas des petits sinistres, elle est moins habituelle pour les gros. Alors, si le temps opérationnel permet de distinguer les petits des gros sinistres, il peut être intéressant de tester l'existence d'un lien entre le temps opérationnel et la surinflation.

Dans le cas des données étudiées, nous avons observé⁴⁵ des évolutions de coûts différentes selon le temps opérationnel des sinistres. Une analyse graphique complémentaire⁴⁶ révèle trois profils d'évolution :

- Les coûts moyens associés à des temps opérationnels inférieurs à 20% décroissent depuis 1979 jusqu'en 1987 ($t = 9$) puis augmentent.
- Les coûts moyens associés à des temps opérationnels compris entre 20% et 60% décroissent depuis 1979 jusqu'en 1987 puis se stabilisent.
- Les coûts moyens associés à des temps opérationnel supérieurs à 60% croissent sur l'ensemble de la période observée.

Pour tester la pertinence de ces tendances, on définit trois intervalles⁴⁷ de temps opérationnel $I_{v, v = 1, 2, 3}$:

$$I_1 = \{ \tau \mid \tau \leq 20\% \}, I_2 = \{ \tau \mid 20\% < \tau \leq 60\% \}, I_3 = \{ \tau \mid \tau > 60\% \},$$

auxquels on associe une fonction $f^{1,v}(t_r)$ de structure identique à la fonction $f^1(t_r)$ mais de paramètres propres $\xi_{j,v}^c, j=1,2; v=1,2,3$ ⁴⁸. On fait donc dépendre les paramètres de $f^1(t_r)$ de τ_r .

On construit le modèle (20) :

$$\ln E[S_r] = +\xi_s + \xi_{10} \cdot (0,2 - \tau_r)^+ + \xi_{30} \cdot (\tau_r - 0,6)^+ + \xi_{40} \cdot (\tau_r - 0,9)^+ + \sum_{v=1}^2 I_{\tau_r \in I_v} \times f^{1,v}(t_r) + \xi_3^c \cdot t_r, \quad (20)$$

$$\text{où} \quad f^{1,v}(t_r) = \xi_{1,v}^c \cdot (9 - t_r)^+ + \xi_{2,v}^c \cdot (t_r - 9)^+, I_z = 1 \text{ si } Z \text{ est réalisée, } 0 \text{ sinon.}$$

⁴⁵ Se reporter aux figures 10 et 11.

⁴⁶ Les graphiques sont présentés en annexe K.

⁴⁷ On a $I_1 =]0\%;20\%]$, $I_2 =]20\%;60\%]$ et $I_3 =]60\%;100\%]$.

⁴⁸ Par exemple, $\xi_{1,v=1}^c$ est le paramètre de surinflation sur la période 1979 – 1987 ($t = 9$) pour les coûts moyens associés à des valeurs de τ inférieures à 20%. De par l'unicité de la tendance croissante sur I_3 , on a $\xi_{1,3}^c = \xi_{2,3}^c = \xi_3^c$.

Un premier ajustement aux données conclut à la nullité et à l'absence de significativité du paramètre $\xi_{2,1}^c$. Ce résultat est lié à la redondance des tendances des coûts moyens des ensembles I_1 et I_2 pour la période postérieure à 1987. Ainsi, l'exclusion de cette tendance du modèle conduit à estimer une croissance commune sur cette période pour les deux ensembles.

On aboutit aux résultats suivants :

Tableau 9– Modèle avec temps opérationnel, surinflation x temps opérationnel et survénance

Paramètres	ξ^s	ξ_1^{to}	ξ_3^{to}	ξ_4^{to}	$\xi_{1,1}^c$	$\xi_{1,2}^c$	$\xi_{2,2}^c$	ξ_3^c
Variables	z^s	z_1^{to}	z_2^{to}	z_3^{to}	$z_{1,1}^c$	$z_{1,2}^c$	$z_{2,2}^c$	z_3^c
Estimateurs	10,1975	2,5178	3,5656	-6,8046	0,1388	0,1838	-0,1103	0,0385
Ecart type	0,2004	1,1730	0,5573	1,9438	0,0453	0,0464	0,0346	0,0157
P-value	$< 10^{-4}$	0,0318	$< 10^{-4}$	0,0005	0,0022	$< 0,0001$	0,0014	0,0143

AIC = 51,9213

Les fonctions $f^{l,v}(t_r)$ ajustées sont rapportées à la figure 29, pour des valeurs charnières de temps opérationnel. Les taux de croissance liés à ces tendances sont récapitulés au tableau 10.

Figure 29 – Fonctions $f^{l,v}(t_r)$ pour $\tau = 20\%$, 60% et 90%

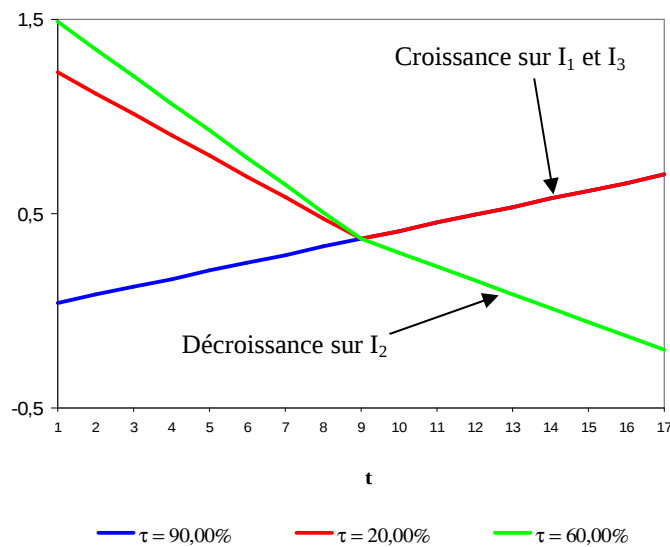


Tableau 10 – Taux de croissance annuels estimés des coûts constants moyens

Intervalles	$t \leq 9$	$t > 9$
I_1	$\exp(-\xi_{1,1}^c + \xi_{1,3}^c) - 1 = -9,54\%$	$\exp(\xi_{1,3}^c) - 1 = +3,93\%$
I_2	$\exp(-\xi_{2,1}^c + \xi_{1,3}^c) - 1 = -13,52\%$	$\exp(\xi_{2,2}^c + \xi_{1,3}^c) - 1 = -6,92\%$
I_3	$\exp(\xi_{1,3}^c) - 1 = +3,93\%$	$\exp(\xi_{1,3}^c) - 1 = +3,93\%$

Les figures 30 à 32 présentent les coûts moyens déflatés dans le temps et leurs estimations par le modèle (20). Elles témoignent de la bonne qualité d'ajustement et de la capture des tendances surinflationnistes, dont les causes restent inconnues.

Dans le cadre du provisionnement, il est légitime de s'interroger sur la pertinence de projeter la tendance baissière observée sur I_2 . Du fait de la forte significativité de $\xi_{2,2}^c$, nous décidons de maintenir cette hypothèse. Aux vues du triangle des temps opérationnel, cette décision influe essentiellement sur le coût futur estimé au titre de la dernière survenance⁴⁹.

Figure 30 – Coûts moyens et estimations - $\tau \leq 20\%$.

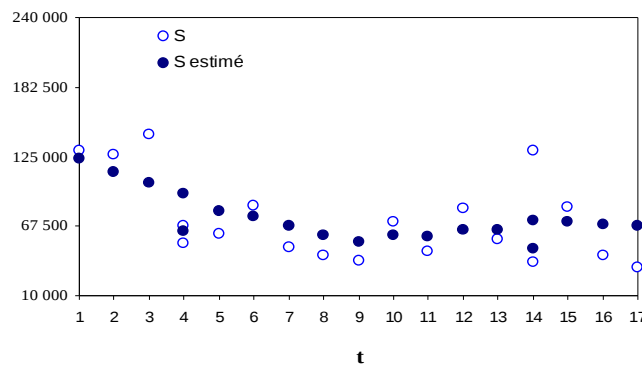


Figure 31 – Coûts moyens et estimations - $20\% < \tau \leq 60\%$.

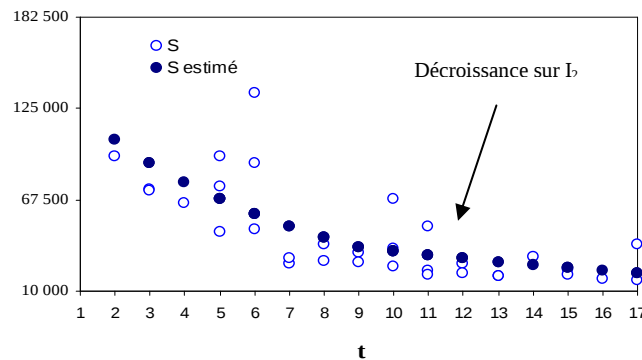
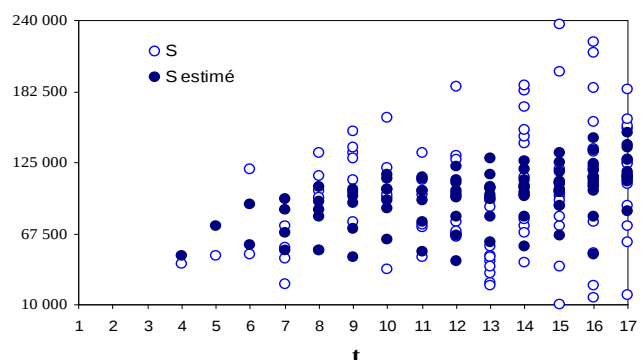


Figure 32 – Coûts moyens et estimations - $\tau > 60\%$.



⁴⁹ Par le triangle des temps opérationnels (annexe F), on voit que τ dépasse 60% en moyenne entre la 3^e et la 4^e année de développement. Pour la dernière survenance, le postulat équivaut donc à anticiper une baisse annuelle des coûts moyens de 6,92% pendant deux à trois ans

III. La modélisation des paiements partiels

III.1 L'exposé du problème

Il n'est pas rigoureux d'assimiler la quantité X_{ij}/N_{ij} à une prestation moyenne par sinistre réglé. En effet, si le règlement de chaque sinistre était effectué au moyen d'un unique paiement, cette quantité représenterait effectivement la prestation moyenne payée en (i,j) . Cependant, un sinistre peut déclencher de nombreux versements. De ce fait, les prestations formant X_{ij} sont versées au titre des N_{ij} sinistres réglés en (i,j) , mais également pour des sinistres toujours ouverts à la fin de cette même période.

Considérons une année de survenance i , dont on réécrit la somme des prestations payées au délai j

$$X_j = X_{j,1} + X_{j,2},$$

comme la somme $X_{j,1}$ des prestations payées l'année de développement j en règlement de sinistres, et de celle $X_{j,2}$ des paiements partiels liés aux sinistres en cours à la fin de l'année j .

On désigne par $N_{j,1}$ et $N_{j,2}$ le nombre de prestations payées durant l'année de développement j respectivement au titre des règlements de sinistres et des paiements partiels. Nous sommes dans le cas le plus courant où seules les quantités X_j et $N_{j,1}$ sont connues. En l'occurrence, la quantité S étudiée jusqu'ici s'exprime en ces termes avec

$$S_j = \frac{X_j}{N_{j,1}}.$$

La prise en compte des paiements partiels complique la modélisation des coûts moyens, au sens où les processus X_1 et X_2 peuvent suivre des développements différents quand τ varie entre 0 et 1. En stipulant sur la dépendance entre N_1 et N_2 , on peut identifier l'influence des deux processus sur la définition du coût moyen S_j .

Considérons tout d'abord N_1 et N_2 comme indépendants. Si N_1 fluctue au delà (en deçà) de son espérance, on peut s'attendre à observer une hausse (baisse) de X_1 sans pour autant que la valeur de N_2 (et a priori celle de X_2) ne soit affectée. Dans ce cas, on s'attend à une baisse (hausse) de S_j par rapport à son espérance. Autrement dit, les variations aléatoires de S autour de son espérance seront corrélées négativement avec celles de N_1 .

Supposons que les quantités N_1 et N_2 soient corrélées positivement. Ce phénomène peut provenir de l'absence de distinction entre règlement et paiement partiel dans les procédures de gestion : les hausses (baisse) concomitantes de N_1 et N_2 traduisent celles de l'activité des paiements au sens large. Dans ce cas, la corrélation négative entre S_j et N_{1j} décrite plus haut est atténuée, voire éliminée si la valeur de N_2 est proportionnelle à celle de N_1 conditionnellement à N_1 . Et dans le cas inverse de corrélation négative entre N_1 et N_2 , l'effet est accentué⁵⁰.

⁵⁰ Si par exemple $a > 1$ et $b < a$, on a bien $\frac{X_1 + X_2}{N_1} > \frac{aX_1 + X_2}{aN_1} > \frac{aX_1 + bX_2}{aN_1}$.

III.2 L'approche de Wright

Considérons une seule année de survenance. Pour l'année de développement j associée aux temps réel et opérationnel t et τ , on désigne par $m_{\tau,t}$ la prestation moyenne d'un sinistre réglé et par L_j le nombre moyen de sinistres encore en cours à la fin de l'année.

L'approche de Wright (1992) repose sur les deux hypothèses fondamentales suivantes :

- Le montant moyen d'un paiement partiel observé durant l'année j peut être exprimé en proportion de $m_{\tau,t}$:

$$E(X_{j,2}) = \Gamma \cdot m_{\tau,t}, \text{ étant donné un scalaire } \Gamma \text{ supposé unique } \forall \tau, t.$$

- Chaque sinistre en cours produit des paiements partiels suivant une loi de Poisson de paramètre p . Sous cette hypothèse, la quantité $N_{j,2}$ est distribué selon une loi $\text{Poi}(p \cdot L_j)$.

En supposant l'indépendance des nombres et des prestations⁵¹, on a

$$\begin{cases} E(X_{j,1}) = N_{j,1} \cdot m_{\tau,t} \\ E(X_{j,2}) = L_j \cdot p \cdot \Gamma \cdot m_{\tau,t} \end{cases} \quad (21)$$

$$(22)$$

A partir de (21) et (22), on réécrit l'espérance de X_j :

$$E(X_j) = E(X_{j,1}) + E(X_{j,2}) = N_{j,1} \cdot m_{\tau,t} \cdot (1 + R_j \cdot \xi^{pp}),$$

où $R_j = \frac{L_j}{N_{j,1}}$ et $\xi^{pp} = p \cdot \Gamma$.

Autrement dit, le coût moyen a pour espérance $E(S_j) = m_{\tau,t} \cdot (1 + R_j \cdot \xi^{pp})$. Sous ces hypothèses, l'introduction des paiements partiels revient à appliquer un facteur multiplicatif $\xi^{pp} \cdot R_j$ au coût moyen d'un sinistre réglé. Comme $E(X_{j,2}) = N_{j,2} \cdot \xi^{pp} \cdot m_{\tau,t}$, on remarque que le coefficient ξ^{pp} exprime l'espérance d'un paiement partiel en proportion de $m_{\tau,t}$.

Si comme dans la plupart des cas⁵², on a $\xi^{pp} \cdot R_j < 1$, on peut estimer $E(S_j)$ par

$$E(S_j) \cong m_{\tau,t} \cdot \exp(R_j \cdot \xi^{pp}) \quad (23)$$

Si les R_j sont connus, Wright préconise d'utiliser l'équation (23) pour estimer ξ^{pp} . Si ξ^{pp} est significatif, on intègre la variable associée z^{pp} dans le modèle. Dans ce cas, il est nécessaire au moment du calcul des provisions de déterminer un modèle de projection des R_j au delà des horizons de temps observés. A ce sujet, Wright propose de modéliser R_j à partir de τ .

⁵¹ Ce résultat élémentaire de théorie du risque est rappelé en annexe L. Sur le sujet, on peut consulter Denuit et al. (2004).

⁵² Le respect de cette condition dépend en fait du rythme des règlements. Si en moyenne, par exemple $p = 50\%$ des sinistres en cours génèrent un paiement partiel correspondant à $\Gamma = 20\%$ du coût d'un sinistre réglé, la condition est respectée si les nombres de règlement $N_{i,j}$ sont supérieurs $\forall j$ à $\xi^{pp} = \Gamma \times p = 10\%$ du nombre des sinistres en cours.

III.3 L'application numérique

Comme le modèle (20) définit une forme linéaire pour $\ln(m_{t,\tau})$, il s'agit simplement d'introduire la variable $z_r^{pp} = R_r = L_r / N_{r,1}$ dans le prédicteur linéaire du modèle (20). Cette action définit le modèle (24). Les valeurs des R_{ij} sont déduites des triangles exposés en annexe B et D.

L'ajustement du modèle (24) conduit à extraire la variable $z_{r,1}^{to}$ devenue non significative, au sens de sa p-value. On obtient les résultats suivants :

Tableau 11 – Modèle avec année de survenance, temps opérationnel, surinflation et paiements partiels

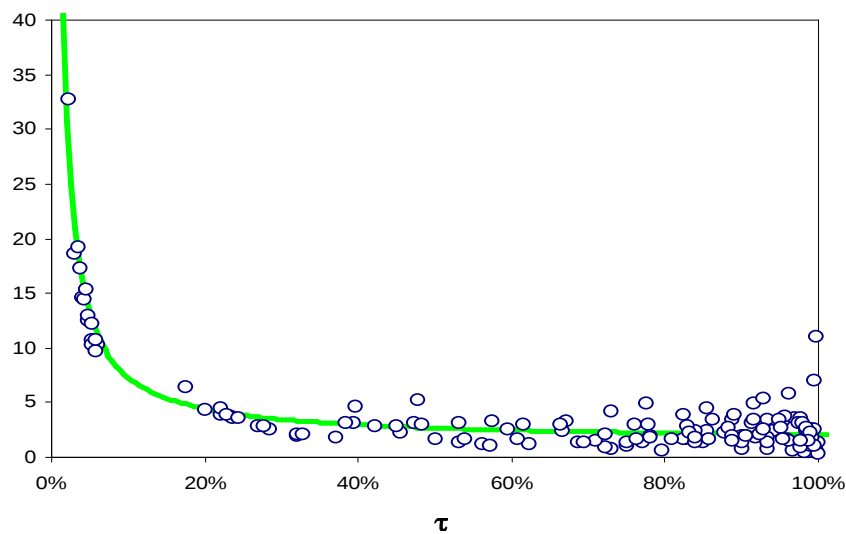
Paramètres	ξ^s	ξ^{pp}	ξ_3^{to}	ξ_4^{to}	$\xi_{1,1}^c$	$\xi_{1,2}^c$	$\xi_{2,2}^c$	$\xi_{1,3}^c$
Estimateurs	9,9955	0,0425	3,9656	-8,0606	0,1448	0,1862	-0,0936	0,0402
Ecart type	0,1991	0,0117	0,5171	1,8890	0,0430	0,0423	0,0328	0,0149
P-value	$< 10^{-4}$	0,0003	$< 10^{-4}$	$< 10^{-4}$	0,0008	$< 10^{-4}$	0,0044	0,0072

AIC = 49,2015

De façon générale, le passage du modèle (20) au modèle (24) améliore la qualité de l'ajustement au sens du critère AIC. En particulier, la significativité du paramètre ξ^{pp} conduit à conclure à l'influence des paiements partiels dans la formation des coûts moyens au sens du modèle de Wright.

Remarquons que le modèle (24) est construit en substituant z_r^{pp} à $z_{r,1}^{to} = (0,2 - \tau_r)^+$ dans le modèle (20). Du fait de la colinéarité des deux variables sur l'intervalle $[0\% ; 20\%]$, cette substitution n'entraîne pas de variation notable dans la valeur des estimations des autres paramètres.

Figure 33 – Ratios R_{ij} des paiements partiels en fonction des τ_{ij}



La figure 33 témoigne du niveau particulièrement bas⁵³ du taux de règlement sur le 1^{er} intervalle de τ , et dans une moindre mesure sur le 2^e intervalle. Ce phénomène peut être lié au temps de traitement des règlements et des déclarations des sinistres.

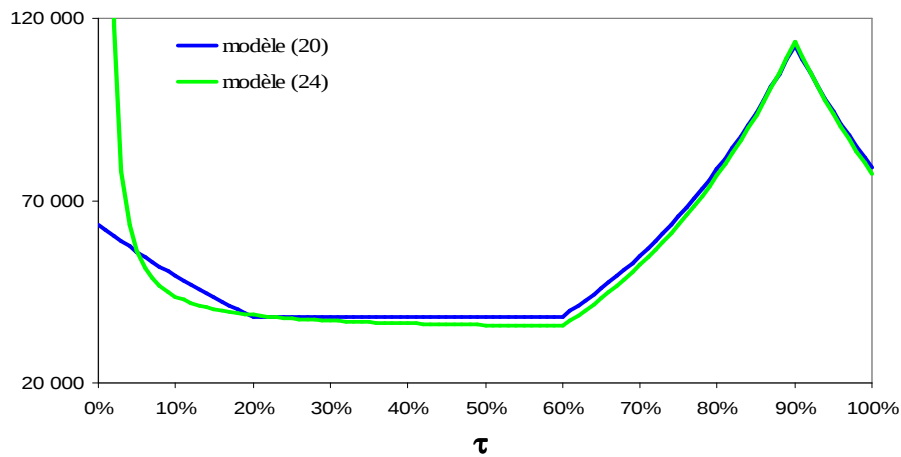
Selon Wright, la significativité du coefficient ξ^{pp} suffit à maintenir les paiements partiels dans le modèle. Si on inclut l'effet des paiements partiels dans les projections de coûts, il est nécessaire d'estimer les valeurs des R_{ij} pour les valeurs futures de temps opérationnel, de telle sorte que l'ensemble des éléments de la matrice Z soient connus. A ce sujet, la forme linéaire $R_\tau = \alpha_0 + \alpha_1/\tau$ ajuste bien les données. La courbe ajustée est présentée à la figure 33 (en vert).

L'estimation des moindres carrés donne

$$\begin{cases} \alpha_0 = 1,5181 \\ \alpha_1 = 0,5818 \end{cases}$$

On observe sur la figure 34 que l'introduction des rachats partiels dans le modèle (24) corrige le profil de la courbe $m_{t,\tau}$ sur l'intervalle $\tau \leq 20\%$. Les coûts moyens estimés sont présentés pour $t = 9$ de sorte à annuler l'effet du temps réel et ne faire ressortir leur évolution qu'en fonction du temps opérationnel.

Figure 34 – Estimation des coûts moyens
modèle (20) contre modèle (24) - $t = 9$



Ce phénomène est lié à la décroissance monotone du facteur $\exp(R_j \cdot \xi^{pp})$ et à la forte baisse des R_j sur cet intervalle.

Notons que la valeur estimée de ξ^{pp} n'est pas en contradiction avec l'interprétation théorique donnée par Wright : l'auteur fait référence à une valeur de $c = 10\%$, associée aux paramètres $\Gamma = 50\%$ et $p = 20\%$. En outre, une interprétation de la baisse des coûts par sinistre réglé est possible sur l'intervalle de temps opérationnel $[0\% ; 20\%]$.

⁵³ Pour $j = 1$, on a en moyenne $R_{11} = 14$, soit un taux de règlement de $1/14 \approx 7\%$. Le même taux passe à 30% en $j = 2$, pour se stabiliser autour de 40% par la suite (soit $R_{ij} \approx 2,5$).

Dans tous les cas, l'introduction des paiements partiels dans la modélisation fournit une interprétation fortuite à la baisse du coût moyen par sinistre réglé observée sur les 1^{ères} valeurs de temps opérationnel :

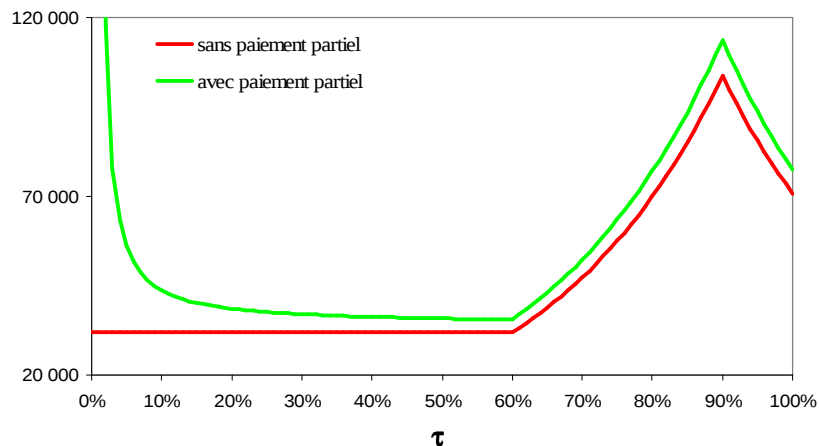
L'observation de la figure 33 rappelle que le taux de règlement 1/R est particulièrement faible au début de l'écoulement du risque. Cette période correspond à la première année de développement, ou de façon équivalente à des valeurs de temps opérationnel inférieures à 20%. Par la suite, le taux de règlement se stabilise au-delà de $\tau \geq 20\%$.

Or, les paiements partiels accroissent le niveau de S_j en augmentant son numérateur X_j , sans contrepartie au dénominateur $N_{j,1}$. Si le taux de règlement croît fortement comme c'est le cas sur l'intervalle $]0\%, 20\%]$, il est possible d'observer une décroissance du coût moyen par règlement, même si la prestation individuelle moyenne $m_{t,\tau}$ reste stable.

Ce phénomène est illustré par la figure 35. Pour une valeur de t donné, elle présente les coûts moyens par sinistre en fonction du temps opérationnel, avec et sans rachats partiels⁵⁴. Les coûts sont estimés à partir du modèle (24) et de valeurs R_τ calculées à partir des estimations des coefficients α_0 et α_1 .

Figure 35 – Estimation des coûts moyens

modèle (24) - $t = 9$



Le modèle (24) explique donc la décroissance des coûts moyens S_j sur les deux 1^{er} intervalles de temps opérationnel par une forte baisse de la proportion des paiements partiels dans l'ensemble des prestations. En début de partie, une deuxième possibilité d'explication avait été énoncée : il s'agissait de considérer que les règlements sans suite intervenaient majoritairement les 1^{ères} années⁵⁵. Aucune information ne permet de valider cette hypothèse.

Au même titre, il est impossible de juger du rôle des paiements partiels dans la baisse des coûts moyens pour les valeurs élevées de τ , au sens où les deux termes de $E(S_j)$ décroissent pour les valeurs élevées de τ . D'une part, le facteur exponentiel décroît du fait de la baisse des R_{ij} avec τ . D'autre part, la figure 35 confirme la baisse de $m_{t,\tau}$ pour les valeurs élevées de τ .

⁵⁴ On modélise les coûts moyens sans rachats partiels en soustrayant le facteur $R_\tau \cdot \xi$ du prédicteur linéaire du modèle (24)

⁵⁵ Les règlements sans suite gonflent le dénominateur de $S_j = X_j / (N_{j,1} + N_{j,2})$, sans contrepartie au numérateur (on a $N_{j,1} = 1$ et $X_j = 0$). Tous paramètres de S_j égaux par ailleurs, on obtient la décroissance de S_j en supposant des règlements sans suite croissant avec j .

IV. La sélection du modèle final

A ce stade, il s'agit de choisir entre les modèles (20) et (24). Comme la substitution de z_r^{pp} à $z_{r,1}^{to}$ ne fait que peu varier les estimations des autres paramètres du modèle, nous sommes tenté de conclure que l'amélioration du critère AIC du modèle (20) au modèle (24) est liée à au meilleur ajustement des coûts moyens sur le premier intervalle de temps opérationnel. Cette conclusion semble confortée par les figures 36 et 37, confrontant ces coûts et leurs estimations par les deux modèles.

Figure 36 – Coûts moyens et leurs estimations
modèle (20) sans paiements partiels – $\tau \leq 20\%$.

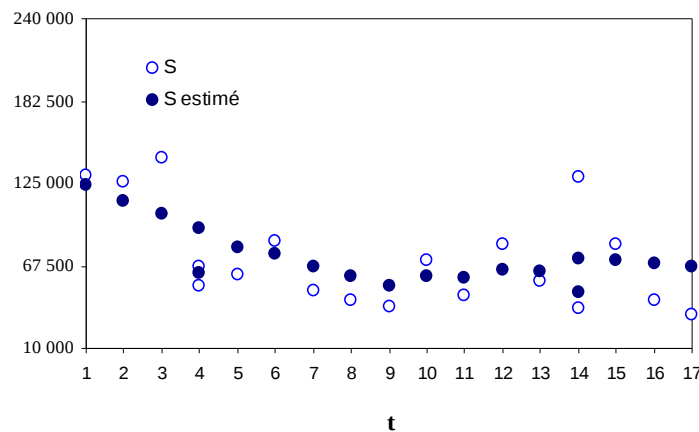
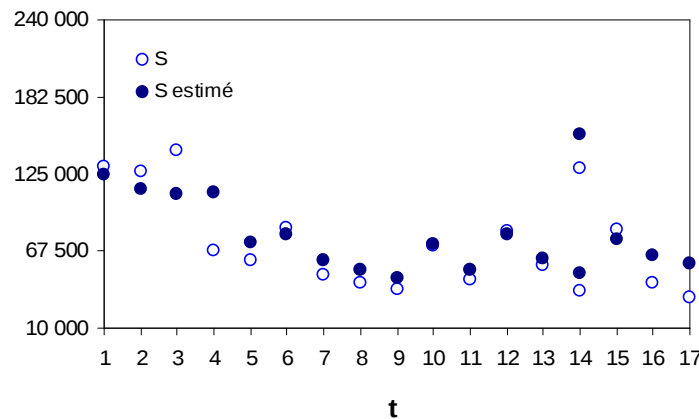


Figure 37 – Coûts moyens et leurs estimations
modèle (24) avec paiements partiels – $\tau \leq 20\%$.



Il convient cependant de relativiser la portée sémantique de ce résultat. En ce sens, Wright tempère l'importance que l'on doit accorder à la vérification des hypothèses sous-jacentes à la modélisation des paiements partiels, qu'il considère avant tout comme un outil d'ajustement à confronter aux données.

Le choix du modèle (20) est motivé par ce même pragmatisme. L'observation de la figure 34 rappelle que les deux modèles partagent sensiblement les mêmes estimations pour les intervalles de temps opérationnel supérieurs à 20%.

Dans le cadre du provisionnement, le meilleur ajustement des « premiers » coûts moyens est de peu d'influence, dès lors que les temps opérationnels des prestations futures à estimer sont dans leur grande majorité au-delà de 20%. Alors, le maintien des paiements partiels dans le modèle alourdirait inutilement la procédure de calcul des provisions, en ce sens qu'il nécessiterait la projection des R_t .

Dans le cadre du provisionnement, nous décidons donc de retenir le modèle (20), dont les caractéristiques sont rappelées au tableau 12 :

Tableau 12 – Variables explicatives et paramètres associés
Modèle (20)

Modélisation	Paramètres	Observations z_{ij}
Année de survenance	ξ^s	$z_r^s = \cdot 1$
Temps opérationnel	ξ_1^{to}	$z_{r,1}^{to} = (0,2 - \tau_{ij})^+$
Temps opérationnel	ξ_2^{to}	$z_{r,3}^{to} = (\tau_{ij} - 0,6)^+$
Temps opérationnel	ξ_3^{to}	$z_{r,4}^{to} = (\tau_{ij} - 0,9)^+$
Surinflation	$\xi_{1,1}^c$	$z_{r,1,1}^c = I_{\tau_r \in I_1} \times (9 - t_r)^+$
Surinflation	$\xi_{1,2}^c$	$z_{r,1,2}^c = I_{\tau_r \in I_2} \times (9 - t_r)^+$
Surinflation	$\xi_{2,2}^c$	$z_{r,2,2}^c = I_{\tau_r \in I_2} \times (t_r - 9)^+$
Surinflation	ξ_3^c	$z_{r,3}^c = t_r$

CHAPITRE IV. VALIDATION DU MODELE PAR DES TESTS STATISTIQUES

I. Les points de validation

Le choix du modèle et les conclusions qui en sont issues dépendent exclusivement des données observées. Une préoccupation majeure de la modélisation est alors de juger de la fiabilité de ces conclusions. Différentes causes peuvent venir jeter un doute sur leur pertinence, dont les plus importantes sont :

- **L'inadéquation des hypothèses formulées sur le processus générateur des données :**

Le premier point traite de la pertinence d'utiliser le modèle retenue au regard des données à modéliser. Si les hypothèses apparaissent cohérentes, on parle de modèle bien spécifié. Sans connaissance a priori du processus étudié, la bonne spécification du modèle ne peut être testée qu'à partir des données disponibles. Ce point est traité dans la section II.

- **Le défaut de « justesse » des données de l'échantillon :**

Les données disponibles peuvent ne pas refléter le processus à modéliser. Autrement dit, il est possible de statuer à la bonne spécification d'un modèle à partir d'un échantillon pas représentatif. Dans notre cas, il est impossible de statuer sur cette problématique étant donné notre méconnaissance du portefeuille. Ce point ne sera donc pas étudié.

- **La sensibilité du modèle à de faibles variations des données :**

Les données historiques ne constituent qu'un échantillon du processus à modéliser. Or, cet échantillon peut inclure des observations qui sans être erronées influent de façon « exagérée » sur l'ajustement du modèle aux données. On peut par exemple s'interroger sur la pertinence de garder dans l'échantillon d'ajustement une observation de valeur extrême et de faible fréquence. A ce sujet sont introduit deux notions : celle de point atypique et celle de point influent. Leur étude est l'objet de la troisième section.

En fin de chapitre, nous testons la crédibilité de notre principal outil de diagnostic dans la construction du modèle de provisionnement : le test de la déviance. Les résidus utilisés dans l'analyse graphique sont les résidus de déviance normalisés.

II. Les tests sur la spécification du modèle

La pertinence d'utiliser le modèle (20) dépend de la crédibilité d'hypothèses formulées a priori sur le processus générateur des coûts moyens. On peut citer trois hypothèses centrales :

- Hypothèse n°1 : La fonction de lien lie l'espérance des coûts moyens avec le prédicteur linéaire. A ce sujet, l'adéquation de la fonction logarithmique a été confirmée en amont.
- Hypothèse n°2 : Le modèle suppose une dépendance entre les coûts moyens et les variables constituant le prédicteur linéaire.
- Hypothèse n°3 : Un lien est supposé entre la variance et l'espérance des coûts moyens.

■ Test de l'hypothèse n°2

Le test de l'hypothèse n°2 est global. Il doit permettre de confirmer qu'il n'existe pas de problème majeur dans la modélisation de la dépendance entre les coûts moyens et les variables explicatives. Ainsi, il s'agit de vérifier globalement tant la pertinence des variables explicatives que la forme de leur dépendance supposée avec la variable réponse.

En pratique, les figures 38 à 41 ne doivent pas faire apparaître de structure non aléatoire qui témoignerait qu'il existe des tendances dans les données non capturées par le modèle. La reconnaissance d'une tendance dans les résidus peut révéler des problèmes très divers⁵⁶. On rappelle que cette contrainte est également d'ordre théorique, dès lors que la qualité globale des résultats dépend en régression de l'indépendance des résidus et des variables explicatives.

Figure 38 – Résidus contre estimations

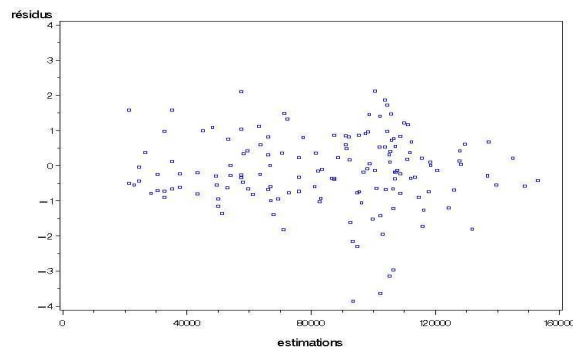


Figure 39 – Résidus contre temps opérationnels

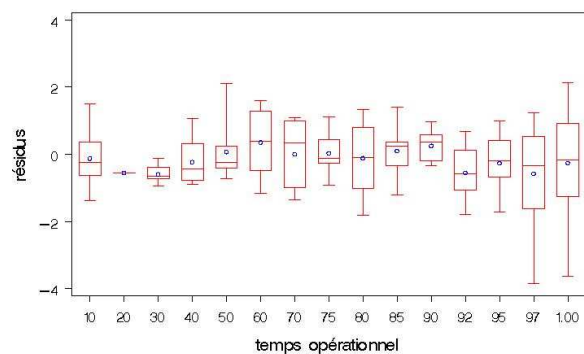


Figure 40 – Résidus contre années de survénance

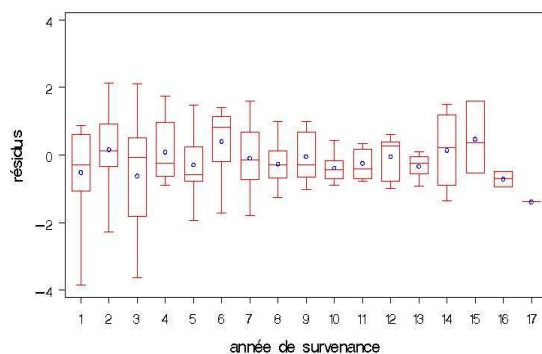
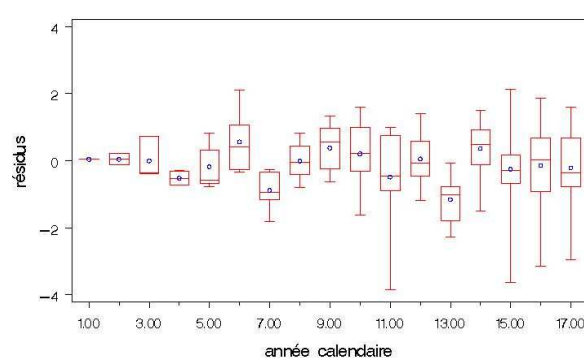


Figure 41 – Résidus contre années calendaires



On constate sur les figures 39 à 41 une horizontalité satisfaisante du centre des boîtes à moustaches⁵⁷, témoignant de l'absence de tendances des résidus en fonction des variables explicatives.

⁵⁶ oubli de variables explicatives, mauvaise définition de la fonction variance et/ou du lien entre la variable réponse et une/des variables explicatives, inadéquation d'un modèle unique sur l'ensemble des valeurs de coûts moyens,

⁵⁷ Les extrémités de la boîte matérialisent les 25^e et 75^e quantiles, celles de chacune des barres les valeurs extrêmes, et le point la médiane.

Par ailleurs, la lecture de la figure 38 de gauche à droite montre que les résidus sont relativement stables en moyenne et en volatilité. Autrement dit, ils semblent indépendants de la valeur des estimations. Ce constat témoigne de la qualité globale de la régression, en ce sens que les estimations ne semblent pas biaisées pour certaines valeurs de coûts moyens. De façon générale, on ne peut donc pas statuer à un problème de spécification du modèle.

▪ Test de l'hypothèse n°3

Quant à la pertinence du choix de la fonction de variance, elle est testée à partir de la propriété d'homoscédasticité des résidus. Rappelons que l'hypothèse d'homoscédasticité des résidus, soit celle de la constance de leur variance, est centrale en économétrie. En théorie, sa transgression biaise les écarts type des estimations des paramètres et altère la fiabilité des tests d'inférence statistique construits à partir des écarts types (t-test, F-test,...).

Dans la pratique des GLM, la constance des écarts types des résidus normalisés dépend de la validité de la relation

$$\text{Var}(S_r) = \phi \cdot V(\mu_r)$$

sur l'étendue de chaque variable explicative. Autrement dit, le lien supposé entre la moyenne et la volatilité des coûts moyens ne doit pas dépendre des variables explicatives. Dans notre cas, ce lien doit en particulier être stable pour toutes les valeurs de τ .

Le test d'homoscédasticité des résidus constitue donc un outil supplémentaire pour conforter le choix de la loi utilisée pour modéliser les S_r . Ce test peut être effectué graphiquement à partir des boîtes à moustaches des résidus de déviance : on statue à leur homoscédasticité si les boîtes ne sont pas de taille « trop » hétérogène. Au regard des figures 39 à 41, le diagnostic n'apparaît pas immédiat.

Pour vérifier l'hypothèse d'homoscédasticité des résidus, on emploie le test de White. Ce test éprouve le pouvoir explicatif des variables explicatives sur les carrés des résidus. Formellement, White définit la régression auxiliaire non contrainte

$$e^2 = a_1 + \sum_{j=1}^p b_j \cdot z_j + \sum_{i=1}^p \sum_{j=1}^p c_{ij} \cdot z_i \cdot z_j \quad (25)$$

où e est le résidu d'estimation issu du modèle (20) et z_j une de ses p variables explicatives.

L'hypothèse d'homoscédasticité testée est

$$H_0 : b_1 = \dots = b_p = c_{11} = \dots = c_{pp} = 0.$$

Le test de White est construit à partir d'un F-test de significativité globale de l'équation (25). Si la p-value de la statistique de test est supérieure au seuil défini, on ne peut rejeter l'hypothèse H_0 d'homoscédasticité des erreurs. Ici, la statistique du test égale à 30,89 est associée à une p-value de 9,83%⁵⁸. Au seuil de 5%, on ne peut rejeter l'hypothèse H_0 .

⁵⁸ On rappelle que la statistique du test est basée sur le coefficient de détermination R^2 de l'équation de régression auxiliaire. Sous H_0 , elle est distribué approximativement selon un $\chi^2_{k, n-k-1}$, où k est le nombre de variables explicatives du modèle auxiliaire.

III. Le test de robustesse

En ce qui concerne les points atypiques, l'observation des figures 38 à 41 permet de constater l'existence de valeurs « anormales » (> 2). Ces résidus présentés au tableau 13 correspondent à des observations de coûts moyens dont les montants sont inhabituellement bas compte tenu des valeurs de temps opérationnel et réel qui leur sont associées.

Tableau 13 – Valeurs atypiques – modèle (20)

i	J	τ	S_r	$\hat{\mu}_r$	$e_r^{(DS)}$
1	11	0,969	7 116,6	93 498 ,66	-3,85
3	13	0,989	9 557	102 247,24	-3,67
3	14	0,992	15 013,33	105 233,43	-3,19
3	15	0,999	17 692,44	106 547,36	-3,02
2	12	0,987	25 627,82	94 934,60	-2,30
1	13	0,991	27 768,77	94 934,60	-2,17
3	4	0,478	134 428,29	93 360,38	2,04
2	14	0,993	236 510	100 573,51	2,08

Cela conduit à s'interroger sur la « justesse » de ces observations plus forcément représentatives du processus analysé. La question est bien illustrée par la notion de point aberrant. Un seul point aberrant peut biaiser le modèle jusqu'à rendre ses prédictions erronées, même dans des régions éloignées de l'observation aberrante. Notre méconnaissance du portefeuille ne permet pas de répondre à cette dernière question. On remarque toutefois que les coûts de montant inhabituel sont associés à d'anciennes survenances.

Afin de statuer sur leur maintien dans l'échantillon d'ajustement, nous étudions leur influence. Par définition, un point influent pèse sur le modèle de telle sorte que si on l'enlève, alors l'estimation des coefficients sera fortement changée. Une mesure classique d'influence est celle de Cook qui est une distance entre le vecteur $\underline{\xi}$ des coefficients estimé avec toutes les observations et celui $\underline{\xi}(-r)$ estimé avec toutes les observations sauf une.

La distance de Cook pour l'observation r se mesure à l'aide de la métrique ($Z'WZ$) :

$$D_r = \frac{1}{p} \cdot \left(\underline{\xi}(-r) - \underline{\xi} \right)' \cdot Z'WZ \cdot \left(\underline{\xi}(-r) - \underline{\xi} \right).$$

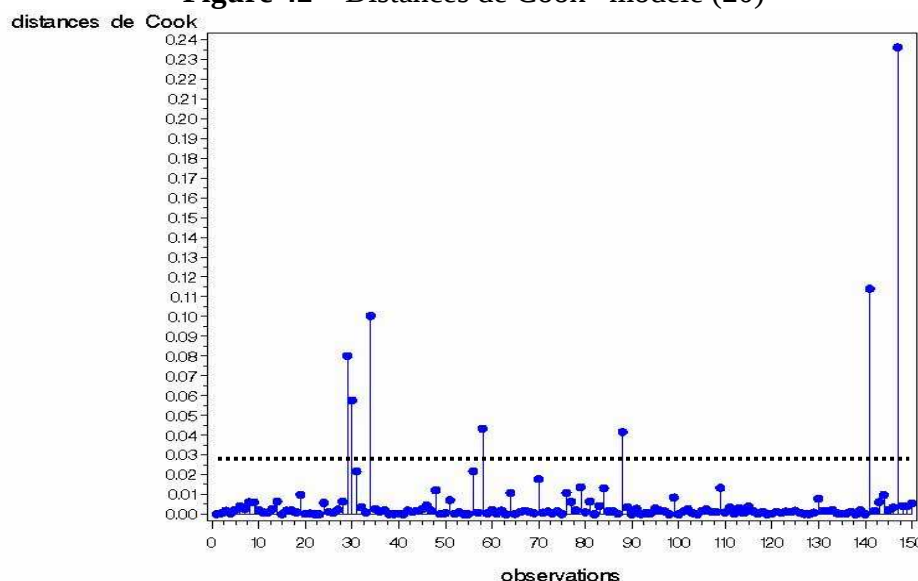
L'expression de la matrice H est rappelé en annexe⁵⁹.

⁵⁹ On rappelle en annexe I que l'algorithme d'estimation fait une régression linéaire à chaque étape et s'arrête à l'étape finale où le processus est stationnaire : $\underline{\xi} = (Z'W^{-1}Z)^{-1} \cdot Z'W^{-1}Y$. Autrement dit, on a $\hat{\underline{\mu}} = Z \cdot \underline{\xi} = H \cdot Y$, où H est la matrice de projection de Y sur Z.

De façon générale, si la distance de Cook est grande pour une observation et/ou très différente des autres, alors le point à l'étude est considéré comme influent. Il convient alors d'analyser les raisons de son influence (erreur de mesure, différence dans la population,...). Pour obtenir des conclusions robustes, il est alors préférable de refaire l'analyse sans ce point.

On calcule les distances de Cook à partir des estimations de ϕ , μ_r , et $V(\mu_r) = \mu_r^2$. Comme la distance de Cook est normée, une valeur supérieure à un est suspecte. Dans le test, nous utilisons une limite plus « sévère » de $4/(n-p-1)$ préconisée par certains auteurs⁶⁰. Compte tenu du seuil de détection fixé à $4/(150-8-1) \cong 0,03$, sept points ressortent comme influents.

Figure 42 – Distances de Cook– modèle (20)



Le tableau 14 présente les observations influentes au sens de Cook et le coût moyen $S_{\tau, \text{moy}}$ observé dans leur intervalle de temps opérationnel depuis $t = 1$.

Tableau 14 – Points influents – modèle (20)

Obs.	i	j	τ	S_r	$S_{\tau, \text{moy}}$	D_r
147	15	3	0,583	39 410,54	47 336,53	0,23
141	14	1	0,022	129 732,60	76 089,43	0,10
34	3	4	0,478	134 428,30	44 295,49	0,10
29	2	14	0,994	236 510,00	108 499,34	0,08
30	2	15	0,996	222 998,50	108 499,34	0,06
58	4	13	0,994	213 777,50	108 499,34	0,04
88	7	4	0,576	68 338,15	47 336,53	0,04

⁶⁰ J. Confais (UPMC-ISUP), M. LeGuen (CNRS-CES-MATISSE-UMR8174) (2009), « Premier pas en régression linéaire avec SAS ».

Les points influents détectés correspondent à des observations dont les valeurs sont anormalement élevées compte tenu du temps réel et du temps opérationnel qui leur sont associés. Des points aberrants du tableau 13, seules deux observations sont considérées influentes. Ces observations apparaissent en gras dans le tableau 14.

Les résultats obtenus en soustrayant de l'échantillon d'ajustement les sept observations influentes sont présentés au tableau 15. On constate une amélioration logique du critère AIC.

Tableau 15 – Modèle (20) – ajustement sans les points influents

Paramètres	ξ^s	ξ_1^{to}	ξ_2^{to}	ξ_3^{to}	$\xi_{1,1}^c$	$\xi_{1,2}^c$	$\xi_{2,2}^c$	$\xi_{1,3}^c$
Estimateurs	10,1639	2,2274	3,9225	-9,2843	0,1528	0,1749	-0,1281	0,0364
Ecart type	0,1930	1,1780	0,5420	1,9302	0,0434	0,0437	0,0362	0,0148
P-value	$< 10^{-4}$	0,0486	$< 10^{-4}$	$< 10^{-4}$	0,0004	$< 10^{-4}$	0,0004	0,0138

AIC = 46,2063 - $\phi = 4,8947^{-1}$

Consécutivement au retrait des sept points influents, on observe que :

- la correction la plus sensible intervient sur le paramètre ξ_3^{to} qui passe de -6,8 à -9,28. Autrement dit, la plus forte correction des paramètres du modèle s'opère sur l'estimation des coûts moyens de fort temps opérationnel. Cette correction est liée à la soustraction de l'échantillon d'ajustement des observations n°29, 30 et 58.
- les taux annuels de surinflation sont révisés à 3,70% sur les intervalles de temps opérationnel I_1 et I_3 et à - 8,75% sur I_2 ⁶¹.

Notons que si nous extrayons les sept points influents de l'échantillon d'ajustement, nous y maintenons les points atypiques non influents⁶² faute de pouvoir justifier de leur retrait.

IV. Le test sur la validité de la déviance

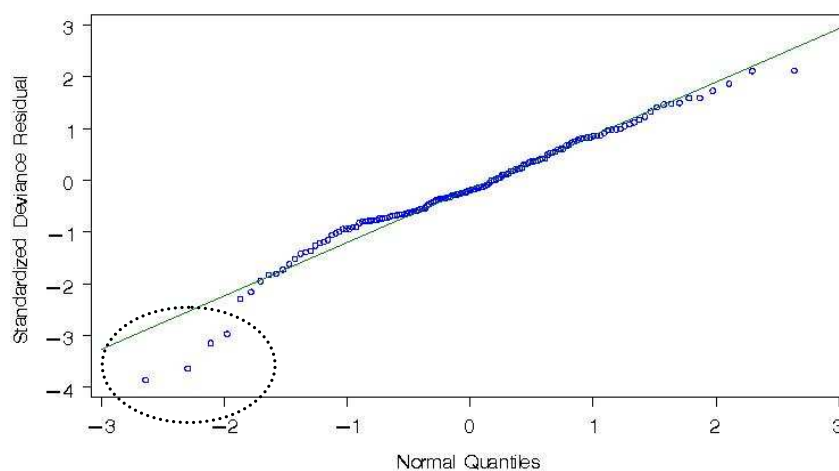
Dans la construction du modèle, nous utilisons essentiellement le test de la déviance comme outil de diagnostic. Ce test utilisé dans la sélection des variables explicatives et la comparaison des modèles fait l'hypothèse que les résidus de déviance ajustés sont issus de distributions normales. Si tel n'est pas le cas, les conclusions qui en sont issues peuvent devenir vides de sens. Nous testons la normalité des erreurs à partir du QQ-plot des résidus.

Mis à part quatre observations en anomalie⁶³, on constate que les points sont alignés autour de la droite passant par les 1^{er} et 3^e quartiles. Les résidus, a priori centrés, passent par le point (0,0). Dans ce cas, il est possible de valider l'hypothèse de normalité.

⁶¹ $e^{0,0364} - 1 = 3,70\%$ et $e^{-0,1281+0,0364} - 1 = -8,75\%$.

⁶² Il s'agit essentiellement des quatre points atypiques soulignés à la figure 43. Ces observations correspondent aux quatre premières valeurs atypiques mentionnées en gras dans le tableau 13

⁶³ Ces observations visionnables à la figure 43 correspondent aux quatre premières valeurs atypiques mentionnées dans le tableau 13.

Figure 43 – QQ plot des résidus de déviance – modèle (20)

V. Conclusion : la validation du modèle

Le tableau 16 présente les résultats du modèle (20) ajusté après le retrait des points influents :

Tableau 16 – Rapports des estimations sur les observations de coûts moyens⁶⁴
modèle (20)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1979	0,98	1,00	1,11	1,16	1,53	0,81	1,73	0,68	0,65	2,23	12,05 ^a	1,32	2,95 ^a	0,59	0,84
1980	0,90	1,13	1,10	0,66	1,16	1,24	1,09	0,94	0,63	2,00	0,72	3,24 ^{a,i}	0,58	0,36 ^{i,a}	0,40 ⁱ
1981	0,72	1,21	0,82	0,41 ^{i,a}	2,64	0,78	0,77	1,00	1,15	0,85	1,01	2,17	10,19 ^a	6,62 ^a	5,53 ^a
1982	1,37	1,31	0,60	1,14	0,86	0,67	1,21	1,59	1,07	0,98	0,48	1,20	0,42 ⁱ	1,03	
1983	1,28	1,12	1,72	0,98	0,57	0,91	0,80	1,89	2,73	1,27	0,48	1,21	1,28		
1984	0,85	1,53	1,04	0,61	0,60	0,69	0,55	2,49	0,54	0,98	0,61	0,53			
1985	1,26	1,41	1,06	0,48 ⁱ	0,69	1,06	2,72	1,03	1,35	0,75	0,65				
1986	1,28	1,28	0,88	0,60	0,97	1,89	0,77	1,35	0,91	1,72					
1987	1,25	1,29	1,31	0,61	1,67	0,70	1,15	0,72	0,76						
1988	0,76	1,44	1,00	1,44	1,02	1,17	1,34	1,26							
1989	1,14	1,27	1,30	0,81	0,90	0,93	1,22								
1990	0,70	1,29	0,73	1,56	0,84	0,90									
1991	1,02	1,21	0,86	1,53	0,94										
1992	0,49 ⁱ	1,03	1,85	0,66											
1993	0,77	1,06	0,44 ⁱ												
1994	1,43	1,00													
1995	1,83														

Rappelons la démarche menant à sa constitution. L'analyse des données a abouti à une conclusion principale, à savoir qu'il existait un lien entre le volume des prestations payées et le nombre des règlements de sinistres. Par ailleurs, l'analyse a conduit à l'instabilité des cadences de règlements dans le temps.

⁶⁴ Les observations annotées d'un exposant « i » et « a » correspondent respectivement à des points influents et à des valeurs atypiques.

Ce constat nous a conduit à intégrer deux notions dans la modélisation des prestations :

- D'une part, nous utilisons une approche par les coûts moyens. Cette méthode nous permet de tenir compte de la dépendance entre les prestations et les nombres de règlements.
- D'autre part, nous modélisons les coûts moyens non pas en fonction du temps mais selon l'ordre d'apparition des règlements. Autrement dit, nous substituons dans la modélisation le temps opérationnel τ à l'année de développement j . Cette méthode nous permet de modéliser des classes de coûts moyens plus homogènes.

Nous avons utilisé la notion de paiement partiel. Il s'est agit de distinguer la prestation moyenne m d'un sinistre réglé du ratio S de coût moyen par sinistre réglé :

Il est naturel d'envisager un lien positif entre m et la durée de règlement des sinistres. Sur ce point, l'introduction des paiements partiels permet de justifier la baisse contre intuitive de S au fur et à mesure que τ pour des valeurs inférieures à 20%. Cette baisse est justifiée par celle de la proportion des paiements partiels dans les prestations totales. Pour les valeurs de τ supérieures à 90%, il n'a en revanche pas été possible de justifier la baisse de S et m .

Au final, le modèle établit une relation entre l'espérance du coût moyen S et deux variables : le temps opérationnel et le taux de surinflation, définissant la croissance de m au-delà de l'inflation générale. Pour plus de robustesse, nous choisissons d'utiliser le moins de variables explicatives possibles et modélisons m en fonction de tendances linéaires.

En ce qui concerne le temps opérationnel τ , nous aboutissons à l'identification de quatre tendances : les coûts moyens baissent avec τ pour les valeurs inférieures à 20%, se stabilisent ensuite pour les valeurs de τ inférieures à 60%, puis augment jusqu'à $\tau = 90\%$ pour rebaisser ensuite.

En ce qui concerne la surinflation, les tendances identifiés par le modèle de Zehwirth ont été révisées en segmentant l'analyse en fonction du temps opérationnel. Autrement dit, une interaction a été détectée entre la surinflation des coûts moyens et leur profil, identifié par leur temps opérationnel. Ainsi, nous avons estimé que les coûts moyens déflatés croissaient de 3,7% par an à partir de 1987, sauf dans le cas des sinistres réglés entre la 3^e et la 5^e année⁶⁵, en baisse de 8,75% par an. Même si il nous est impossible de donner une signification à ce phénomène (sinistres sans suite,...), il caractérise bien les données étudiées⁶⁶.

En fin de partie, des tests de validation du modèle ont été menés. En particulier, ces tests ont conduit au retrait de deux points atypiques, trop influents au sens de la distance de Cook. Cette opération a eu pour résultat d'accentuer la tendance baissière des coûts moyens les plus élevés ($\tau > 90\%$). De façon plus générale, les résultats de ces tests n'ont permis de remettre en cause ni la qualité globale du modèle, ni la fiabilité des outils statistiques utilisés dans sa conception. En effet, d'une part aucun problème de spécification n'a pu être décelé, et d'autre part, les hypothèses sous jacentes au modèle (20) n'ont pu être infirmées sur la base de l'échantillon des coûts moyens étudié.

⁶⁵ Ces sinistres sont réglés dans l'intervalle de temps opérationnel [20%,60%]

⁶⁶ Se reporter aux figures en annexe K.

PARTIE III. LA MISE EN ŒUVRE DES CALCULS

Il s'agit désormais d'utiliser le modèle de provisionnement défini dans la partie II à deux fins : le calcul du Best Estimate des PSAP et celui du risque de provisionnement.

Le premier chapitre est dédié au calcul du Best Estimate de la provision pour sinistres. Son obtention requiert de répondre au préalable aux deux points suivants :

- Le modèle étudié estime le coût moyen des prestations versées à partir de leurs temps réel et opérationnel. Si les temps opérationnels des futurs règlements sont par définition connus, leurs dates restent à estimer.
- Si le modèle intègre le calcul de l'inflation sectorielle contenue dans les prestations futures, il ne prévoit pas l'estimation de l'inflation générale qui nécessite de se doter d'un modèle ad hoc.

Le second chapitre est consacré au calcul du risque de provisionnement. L'approche de calcul privilégiée est de type simulatoire dans un cadre Bootstrap :

- Dans un premier temps, nous adaptons la méthode classique du Bootstrap au modèle de provisionnement et calculons le risque ultime. Ensuite, l'algorithme est adapté au calcul du risque à un an.
 - Enfin, les résultats obtenus sont comparés à celui issu de l'application de la formule standard.
-

CHAPITRE I. LE CALCUL DU BEST ESTIMATE DES PSAP

I. Hypothèses et méthode

I.1 Hypothèses

Nous évaluons le Best Estimate des provisions pour sinistres à partir du modèle (20). Cette section recense les hypothèses sous jacentes à ce calcul.

- **Nombre ultime des sinistres**

Le calcul des PSAP passe par l'estimation de la prestation à payer au titre de chaque sinistre restant à régler. Il requiert donc de connaître de façon certaine le nombre ultime de sinistres pour chaque survenance. A cet égard, nous utilisons les estimations de Taylor (2000). Nous rappelons que nous considérons par simplification le nombre ultime de sinistres comme certain, malgré l'incertitude évidente de cette information en date de calcul.

- **Effet des paiements partiels**

Le modèle (20) est construit à partir des observations de coûts moyens. En d'autres termes, ses projections estiment les coûts moyens futurs et non les prestations futures. Notre utilisation du modèle requiert donc d'assimiler les coûts moyens aux prestations individuelles.

Nous avons vu que les deux notions pouvaient diverger dans le cas de paiements partiels importants. En d'autres termes, assimiler le coût moyen S à la prestation individuelle m suppose de considérer négligeable l'effet des paiements partiels sur les coûts moyens. Dans notre cas, nous justifions cette hypothèse par les résultats de la partie II, incitant à supposer la similitude entre m et S pour les valeurs de τ considérées dans la projection des règlements⁶⁷.

- **Espérances des prestations individuelles**

Dans le calcul des provisions, nous assimilons les extrapolations de coûts moyens du modèle (20) à des prestations individuelles. En omettant l'existence des paiements partiels, réécrivons la somme des prestations payées en j au titre de i :

$$X_{ij} = \sum_{v=1}^{N_{ij}} X_{ij}^v$$

où les X_{ij}^v , $v = 1, \dots, N_{ij}$ désignent les prestations payées au titre des N_{ij} sinistres réglés.

Nous supposons, pour $v = 1, \dots, N_{ij}$, que l'espérance des X_{ij}^v ne dépend que des valeurs t et τ associées au couple (i, j) : $E(X_{ij}^v) = m_{t,\tau} = S_{t,\tau}$. Cette hypothèse est approximative car les N_{ij} sinistres réglés n'ont pas la même espérance. Cependant, elle reste crédible si $m_{t,\tau}$ ne varie pas trop chaque année j . Plus précisément, si τ désigne le temps opérationnel moyen observé durant l'année de développement j , l'hypothèse est exacte si $m_{t,\tau}$ est linéaire en t et τ .

⁶⁷ En particulier, le modèle (24) intégrant les paiements partiels aboutit à des estimations de S et m très proches pour les valeurs de $\tau \geq 0,2$ (se reporter à la figure 35).

- **Indépendance des sinistres**

Supposons que les N_{ij} sinistres soient indépendants et de même espérance $m_{t,\tau}$. Si X_{ij} est la somme de N_{ij} sinistres indépendants, le coût moyen des N_{ij} règlements a pour espérance

$$E(S_{ij}) = E\left(\frac{X_{ij}}{N_{ij}}\right) = m_{t,\tau}$$

Dans le calcul du Best Estimate, nous pouvons alors, au prix des hypothèses formulées, assimiler les estimations des espérances de coûts moyens issues du modèle (20) avec celles des prestations individuelles. On vérifie sous l'hypothèse d'indépendance des sinistres (1) que

$$\begin{aligned} E(S_{ij}) &= E_{N_{ij}}\left(E\left(\frac{X_{ij}}{N_{ij}} \middle| N_{ij}\right)\right) = \sum_{k=0}^{\infty} P(N_{ij} = k) \cdot E\left(\frac{X_{ij}}{k} \middle| N_{ij} = k\right) \\ &\stackrel{(1)}{=} \sum_{k=0}^{\infty} P(N_{ij} = k) \cdot \frac{1}{k} \cdot E\left(\sum_{\nu=1}^k X_{ij}^{\nu}\right) \\ &= \sum_{k=0}^{\infty} P(N_{ij} = k) \cdot \frac{1}{k} \cdot \sum_{\nu=1}^k E(X_{ij}^{\nu}) \\ &= m_{t,\tau} \cdot \sum_{k=0}^{\infty} P(N_{ij} = k) \\ &= m_{t,\tau} \end{aligned}$$

I.2 Les étapes de calcul

En conformité avec les préconisations de l'ACAM (ACAM, 2007), la procédure de calcul du Best Estimate comporte trois étapes :

- l'estimation en monnaie constante des prestations futures
- le calcul de l'inflation contenue dans ces prestations
- l'actualisation des prestations à la courbe des taux sans risque.

- **Meilleures estimations des prestations futures**

Au travers de l'actualisation et du calcul de la croissance (sur)inflationniste des prestations restant à payer, le calcul du Best Estimate dans le cadre du modèle (20) requiert d'estimer la date des règlements futurs. Or, ce modèle n'estime pas cette information. Une solution simple à cette problématique est présentée dans la section suivante. Elle passe par l'approximation d'un lien entre t et τ à partir d'une fonction continue H telle que $t = H(\tau)$.

Ainsi, l'objectif consiste à associer à chaque règlement futur une date $\hat{H}(\tau)$ fonction du temps opérationnel τ qui lui est associé. Une fois la fonction H définie et ses paramètres estimés, le calcul en monnaie constante des prestations futures devient trivial.

En effet, plaçons-nous en t_0 , date de calcul de la provision. On désigne par $N_{i,0}$ le nombre de sinistres réglés à cette date parmi les N_i sinistres survenus au titre de la survenance i et par $\tau_{i,0} = N_{i,0}/N_i$ le temps opérationnel du dernier sinistre réglé.

Les temps opérationnels moyens $\tau_{i,j}$ des $\omega_i = (N_i - N_{i,0})$ sinistres à régler sont⁶⁸

$$\tau_{i,1} = \frac{N_{i,0} + 0,5}{N_i}, \tau_{i,2} = \frac{N_{i,0} + 1,5}{N_i}, \dots, \tau_{i,\omega_i} = \frac{N_i - 0,5}{N_i}, \quad (1)$$

auxquels on associe les temps réels $t_{i,1}, \dots, t_{i,\omega_i}$.

Des $\tau_{i,j}$ et de l'estimation de H , on déduit du modèle (20) les meilleures estimations des prestations associées aux règlements futurs⁶⁹. En effet, l'algorithme du score fournit les estimateurs du maximum de vraisemblance (e.m.v.) du vecteur des paramètres ξ et de ϕ .

Par invariance fonctionnelle du maximum de vraisemblance, on en déduit les e.m.v.

$$\hat{m}_{\tau_{i,j}} = \exp\left(\hat{\eta}\left(\tau_{i,j}, \hat{H}(\tau_{i,j})\right)\right) \text{ des } m_{\tau_{i,j}} = E(S_{\tau_{i,j}}) \quad \text{pour } j = 1, \dots, \omega_i,$$

où $\hat{\eta} = \eta(\hat{\xi}, \hat{\phi})$ désigne le prédicteur linéaire estimé dans la partie II.

Il en résulte un e.m.v de l'espérance de la somme des prestations futures associées aux I années de survenance :

$$\sum_{i=1}^I \sum_{\tau=\tau_{i,1}}^{\tau=\tau_{i,\omega_i}} \hat{m}_{\tau}$$

• Inflation future

Nous appliquons les conditions d'inflation future aux projections de prestations estimées par le modèle (20). Dans cet objectif, nous nous dotons d'un modèle d'inflation ad hoc, à partir duquel il est possible au travers d'un indice $CP(H(\tau))$ en base t_0 , d'estimer l'inflation générale moyenne courant entre chaque date de règlement futur $H(\tau)$ et la date de calcul t_0 .

A partir de $CP(H(\tau))$, on exprime en monnaie réelle les estimations des prestations futures:

$$\hat{m}_{\tau_{i,j}} = \hat{CP}_{\hat{H}(\tau_{i,j})} \hat{m}_{\tau_{i,j}} \quad \text{pour } j = 1, \dots, \omega_i.$$

⁶⁸ Pour $j = 1$, on a bien $\tau_{i,1} = \frac{1}{2} \cdot \left(\frac{N_{i,0}}{N_i} + \frac{N_{i,0}+1}{N_i} \right) = \frac{N_{i,0}+0,5}{N_i}$.

⁶⁹ En supposant que chaque règlement futur déclenche le paiement d'une prestation, on néglige l'effet des sinistres futurs sans suite.

• Actualisation

En conformité avec les préconisations de l'ACAM, nous utilisons une courbe déterministe des taux, en négligeant les dépendances éventuelles entre taux d'intérêt et prestations. Cette dernière étape nous permet d'aboutir à l'expression du Best Estimate en t_0 de la provision pour sinistres⁷⁰ :

$$\hat{E}(R) = \hat{R}^{(0)} = \sum_{i=1}^I \sum_{\tau=\tau_{i,1}}^{\tau=\tau_{i,\omega_i}} \exp\left[-r_{\hat{H}(\tau)} \cdot \hat{H}(\tau)\right] \cdot {}_2\hat{m}_{\tau} . \quad (2)$$

où $r_{\hat{H}(\tau)}$ désigne le taux sans risque d'échéance $H(\tau)$ prévalant en t_0 .

II. La projection en monnaie constante des prestations futures

II.1 La modélisation des prestations

Estimer la valeur réelle des prestations futures requiert de se doter d'un modèle d'inflation. Pour le moment, cette question est reportée tandis que nous nous concentrons sur la méthode d'estimation des paiements futurs exprimés en monnaie constante.

Sous les hypothèses présentées, il est possible d'estimer la prestation moyenne d'un règlement futur à partir du modèle (20). En effet, pour un sinistre réglé en $t > 17$ au temps opérationnel τ , cette quantité est estimée à

$$\begin{aligned} \hat{m}_{t,\tau} &= f(t) \cdot \hat{m}_{\tau}^1 \\ \text{où } \begin{cases} \hat{m}_{\tau}^1 = \exp\left[\hat{\xi}^s + \hat{\xi}_1^{to} \cdot (0,2 - \tau)^+ + \hat{\xi}_3^{to} \cdot (\tau - 0,6)^+ + \hat{\xi}_4^{to} \cdot (\tau - 0,9)^+\right] \\ f(t) = \exp\left[I_{t \in I_2} \cdot \hat{\xi}_{2,2}^c \cdot (t - 9) + \hat{\xi}_3^c \cdot t\right] \end{cases} \end{aligned} \quad (3)$$

L'estimation des prestations espérées futures nécessite la connaissance de leur temps réel et opérationnel. Hors, seul le temps opérationnel de la prestation $m_{t,\tau}$ est connu à partir de l'équation (1). L'estimation du facteur de surinflation $f(t)$ requiert donc que l'on puisse associer une valeur de temps réel à chaque règlement futur à partir de son temps opérationnel.

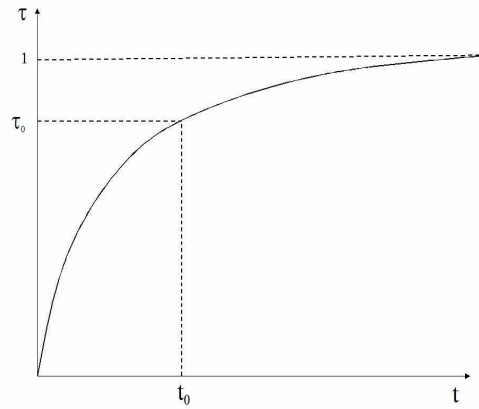
A cet effet, l'estimation du lien entre t et τ peut être effectuée à partir d'une courbe continue. Comme en témoigne la figure 12, cette relation est souvent modélisable à partir d'une fonction de répartition exponentielle⁷¹.

⁷⁰ Le Best Estimate est calculé hors frais de gestion des sinistres.

⁷¹ Dans les premières années, on peut observer un point d'inflexion de la courbe correspondant à une accélération des règlements. Dans ce cas, l'utilisation d'une fonction de type Gamma peut être plus appropriée.

La figure 44 présente une courbe typique.

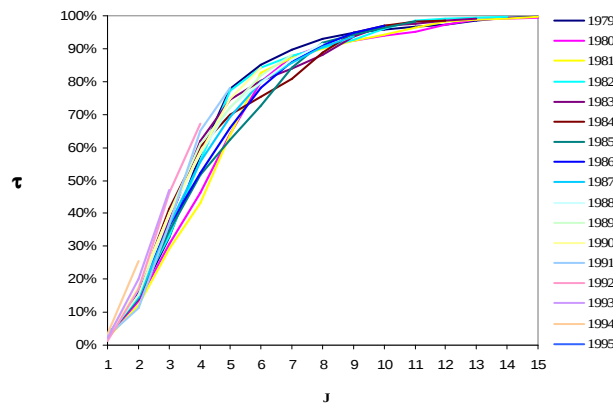
Figure 44 – Liaison type entre temps réel et opérationnel



A ce sujet, l'observation des trajectoires de temps opérationnel observées par survénance permet de considérer raisonnable l'utilisation de la forme exponentielle.

Figure 45 – Trajectoires des temps opérationnels

par année de survénance



Wright suggère la relation suivante entre temps et temps opérationnel⁷² :

$$t - t_{\text{ref}} = H(\tau_{\text{ref}}, \tau) = -\frac{1}{c} \cdot \ln\left(\frac{1 - \tau}{1 - \tau_{\text{ref}}}\right), \quad (4)$$

où t_{ref} désigne le point de départ de comptage du temps réel, et τ_{ref} le temps opérationnel qui lui est associé.

Dans le calcul des coûts moyens, il existe deux points de départ : la mesure du temps réel à partir de $t = 0$ permet d'estimer la surinflation des prestations, celle effectuée à partir de t_0 permet le calcul de leur croissance inflationniste depuis la date de calcul des provisions.

⁷² L'expression de $H(t_0, \tau)$ est justifiée en annexe M.

L'expression en monnaie constante des prestations futures devient

$$m_{t,\tau} = \exp[H(0, \tau) \cdot (I_{\tau \in I_2} \cdot \xi_{2,2}^c + \xi_3^c)] \cdot m_\tau^1 = m_\tau. \quad (5)$$

II.2 L'application numérique

Dans le cas où il existe un lien de type exponentiel entre t et τ , Wright propose une façon aisée de déterminer la meilleure estimation du paramètre c . Il s'agit de remarquer que 50% des sinistres ($\tau = 0,5$) sont réglés au temps réel médian t_m défini par

$$0,5 = 1 - e^{-c \cdot t_m} \Leftrightarrow c = \frac{\ln(2)}{t_m}. \quad (6)$$

La solution consiste à estimer la valeur de t_m à partir de l'observation du triangle des temps opérationnels. Comme les années d'origine sont des années de survénance, il s'agit de remarquer pour $j > 1$, que l'âge moyen des sinistres réglés la $j^{\text{ème}}$ année de développement est approximativement de $j - 1$ années.

Les temps opérationnels moyens en fin de 3^e et 4^e année de développement sont de 37,5% et 56,08%. On estime par interpolation linéaire la valeur médiane t_m pour laquelle $\tau = 0,5$:

$$\frac{0,5608 - 0,5}{3 - t_m} = \frac{0,5 - 0,375}{t_m - 2}, \text{ soit } t_m \cong 2,65 \text{ années.}$$

A partir de l'équation (6), on en déduit la valeur de $c = \ln(2)/2,65 = 0,262$.

On rappelle que le recours au temps opérationnel a été motivé par l'instabilité des cadences de règlements. Dans ce contexte, l'utilisation d'un unique coefficient c pour l'ensemble des survénances peut donc apparaître inadéquate. De ce fait, nous privilégions une solution simple consistant à estimer un temps médian t_m par année de survénance selon la méthode qui vient d'être présentée. Les coefficients estimés sont présentés au tableau 17.

Tableau 17 – Estimation par survénance du lien exponentiel entre temps réel et opérationnel

I	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
t_m	2,69	3,20	3,33	2,74	2,50	2,46	2,90	2,85	2,70	2,55	2,46	2,54	2,46	2,18	2,18	2,18	2,18
c_i	0,25	0,21	0,20	0,25	0,27	0,28	0,23	0,24	0,25	0,27	0,28	0,27	0,28	0,31	0,31	0,31	0,31

La baisse de t_m dans le temps témoigne de l'accélération des règlements sur les trois premières années de développement. Ce phénomène est directement observable depuis le triangle des temps opérationnels. Faute d'observations sur les deux dernières survénances pour pouvoir estimer t_m , nous leur attribuons la courbe d'écoulement des règlements de la survénance 1993 ($i = 14$). L'estimation des écoulements de sinistres par survénance est présentée en annexe M.

L'estimation de la somme des prestations restant à payer est effectuée selon les deux approches : en tenant compte d'un coefficient unique c et à partir d'un coefficient c_i par survenance. Les résultats sont présentés au tableau 18.

On en rappelle la méthode de calcul. Une fois le paramètre c calibré par survenance à partir de l'équation (6) et les temps opérationnels des prestations futures estimés à partir de l'équation (1), on estime les temps réels qui leur sont associés en fonction de l'équation (4). Ensuite, la connaissance des temps réels et opérationnels des sinistres futurs permet d'estimer le montant des prestations à partir de l'équation (5).

On note que l'application d'une seule loi d'écoulement (colonne 2) conduit à des résultats assez différents de ceux issus de l'utilisation d'une cadence par survenance (colonne 3).

Tableau 18 – Estimation des prestations restant à payer

en dollar 1995

i	Nombres de sinistres en cours			Somme des prestations futures		
	Déclarés	IBNR	Total	Chain Ladder (1)	GLM – $c = 0,262$ (2)	GLM – c_i (3)
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	12	0	12	0	1 307 011	1 359 150
4	1	0	1	263 530	100 982	101 318
5	5	0	5	537 194	530 469	526 503
6	11	0	11	1 025 765	1 192 932	1 180 404
7	13	0	13	1 266 194	1 420 650	1 441 796
8	22	0	22	2 384 886	2 466 009	2 495 436
9	40	1	41	3 630 645	4 908 134	4 923 057
10	62	1	63	4 939 162	8 041 863	8 002 984
11	81	3	84	8 607 300	11 446 354	11 337 042
12	91	8	99	14 281 792	14 181 623	14 105 465
13	138	15	153	20 008 266	22 222 160	22 007 342
14	208	26	234	33 439 775	31 973 307	31 110 884
15	332	36	368	36 280 466	42 062 078	40 728 564
16	430	66	496	40 514 274	45 976 992	44 298 879
17	575	243	818	49 070 926	58 158 445	55 901 796
	2 021	399	2 420	216 250 173	245 989 010	239 520 620

Par ailleurs, rappelons-nous avoir observé en début de partie II que le calcul par Chain Ladder conduisait à des résultats très différents selon la profondeur de l'historique pris en compte. Les résultats obtenus ici conduisent à une augmentation de 11% de la charge estimée par Chain Ladder en intégrant l'ensemble des survenances dans le calcul (colonne 1).

La comparaison des deux résultats n'est pas aisée. La seule remarque évidente traite des facteurs de queue. Par définition, la méthode Chain Ladder suppose que le risque des trois premières survenances est totalement écoulé. A l'inverse, la méthode suivie modélise implicitement les facteurs de queue. A cet égard, on estime au titre de la survenance 1981 ($i=3$) une charge de 1,3 millions de dollars relative à 12 sinistres déclarés et dont les règlements interviennent jusqu'à la 28^{ème} année de développement.

III. L'inflation future

III.1 Le modèle d'inflation

Estimer la valeur réelle des prestations futures requiert de se doter d'un modèle d'inflation. De nombreux modèles ont été proposés dans la littérature pour décrire l'inflation. Dans le cadre des travaux actuariels, on peut citer les travaux de Kaufmann (2001), de Wilkie (1995) ou encore d'Ahlgrim et al (2001) dont nous nous inspirons. Dans leur article, les auteurs supposent que le taux d'inflation i_t suit un processus d'Ornstein-Uhlenbeck basé sur l'équation différentielle stochastique

$$di_t = \beta \cdot (\eta - i_t) \cdot dt + \sigma \cdot dW_t. \quad (7)$$

La dynamique décrite par l'équation (7) est caractéristique des modèles dits de « retour vers la moyenne ». Ce phénomène est modélisé par la partie déterministe ou *drift* de l'équation. Par exemple, si le niveau courant i_t de l'inflation est inférieur à sa moyenne de long terme η , le drift devient positif. Ce mécanisme assure la convergence à terme de i_t vers η , convergence dont la vitesse dépend du coefficient $\beta > 0$. Le deuxième terme de l'équation représente l'incertitude liée au processus. La variation du processus de Wiener W_t est distribuée selon une loi normale standard, tandis que le coefficient de diffusion σ détermine l'amplitude de la volatilité du processus.

L'estimation économétrique de ce type d'équation est le plus souvent rendue difficile si comme dans la majorité des cas, on ne dispose que d'observations discrètes. Ainsi, la mise en œuvre de ce processus nécessite dans la pratique sa discrétisation. Il s'agit donc d'estimer les paramètres du modèle à partir du processus discrétisé et non à partir du modèle continu. Pour discrétiser l'équation (7), on peut utiliser :

- **Le schéma d'Euler**⁷³

$$\begin{aligned} \Delta i_t &= i_{t+\Delta t} - i_t = \beta \cdot (\eta - i_t) \Delta t + \varepsilon \sigma \cdot \sqrt{\Delta t} \\ \Leftrightarrow i_{t+\Delta t} &= \beta \Delta t \cdot \eta + (1 - \beta \Delta t) \cdot i_t + \varepsilon \sigma \cdot \sqrt{\Delta t} \end{aligned} \quad (8)$$

où ε est une variable de loi normale centrée réduite et Δt le pas de discrétisation.

Cette approximation est d'autant plus fine que le pas de discrétisation est petit. L'équation (8) définit le taux d'inflation comme une moyenne pondérée du taux de la période précédente et du taux moyen de long terme.

- **La solution exacte**

Gouriéroux et al. (1993) montrent qu'il existe une solution explicite⁷⁴ à l'équation (7), on a

$$i_{t+\Delta t} = i_t \cdot e^{-\beta \Delta t} + \eta \cdot (1 - e^{-\beta \Delta t}) + \sigma \sqrt{\frac{1 - e^{-2\beta \Delta t}}{2\beta}} \cdot \varepsilon \quad (9)$$

⁷³ Pour une présentation du schéma d'Euler, se reporter par exemple au chapitre 5 de Planchet et al. (2005).

⁷⁴ Le cheminement à la solution exacte est rappelé en annexe O.

Selon les deux méthodes, on est donc conduit à considérer une relation linéaire autorégressive du type

$$i_{t+\Delta t} = a \cdot i_t + b + \varepsilon. \quad (10)$$

Dans le cas de la solution exacte, la relation entre l'équation (10) et les paramètres de l'équation (9) est donnée par

$$a = e^{-\beta \Delta t}, \quad b = \eta \cdot (1 - e^{-\beta \Delta t}), \quad \sigma_\varepsilon = \sigma \cdot \sqrt{\frac{1 - e^{-2\beta \Delta t}}{2\beta}}.$$

L'estimation des paramètres a , b et σ_ε peut s'effectuer en recourant soit à la méthode des MCO, soit à celle du maximum de vraisemblance. Les deux méthodes aboutissent aux mêmes résultats. Dans le cas de la régression à laquelle nous avons recours, on est conduit à considérer une équation du type

$$Y = a \cdot X + b + \varepsilon, \quad (11)$$

où Y est le vecteur des taux d'inflation et X celui des taux décalés d'un pas.

Les MCO donne des estimateurs sans biais de a , b et σ_ε dont on rappelle l'expression :

$$\hat{b} = \frac{\sum_{t=1}^m (y_t - \bar{y})(x_t - \bar{x})}{\sum_{t=1}^m (x_t - \bar{x})^2}, \quad \hat{a} = \bar{y} - \hat{b} \bar{x}, \quad \hat{\sigma}_\varepsilon^2 = \frac{\sum_{t=1}^m (\hat{\varepsilon}_t - \bar{\hat{\varepsilon}})^2}{n - 2},$$

où $\hat{\varepsilon}_t = y_t - (\hat{a} x_t + \hat{b})$

A partir des estimations des coefficients de l'équation (11), on déduit celles des paramètres du modèle continu :

$$\hat{\beta} = -\frac{\ln \hat{a}}{\Delta t}, \quad \hat{\eta} = \frac{\hat{b}}{1 - \hat{a}}, \quad \hat{\sigma} = \hat{\sigma}_\varepsilon \cdot \sqrt{-2 \cdot \frac{\ln \hat{\beta}}{\Delta t \cdot (1 - \hat{\beta}^2)}}.$$

Etant donné le taux d'inflation actuel i_t et les paramètres estimés du processus, il est alors possible de projeter l'inflation future et de déduire la trajectoire moyenne des taux d'inflation à utiliser dans le calcul du Best Estimate.

En effet, étant donné i_0 le taux d'inflation observé en t_0 , on peut montrer par le développement d'Itô-Taylor que l'équation (7) définit chaque variable i_t comme gaussienne, d'espérance et de variance

$$\begin{cases} E(i_t) = i_0 \cdot e^{-\beta t} + \eta \cdot (1 - e^{-\beta t}) \\ V(i_t) = \frac{\sigma^2}{2\beta} \cdot (1 - e^{-2\beta t}) \end{cases}.$$

En résumé, le choix de ce modèle d'inflation est avant tout motivé par sa facilité de mise en œuvre. Rappelons pour finir que l'utilisation d'un processus Ornstein-Uhlenbeck influe sur la trajectoire des taux d'inflation à au moins deux niveaux :

- **En ce qui concerne la trajectoire moyenne utilisée dans le calcul du Best Estimate :**

On rappelle que le modèle de Vasicek stabilise la queue de la courbe des taux d'inflation, comme l'ensemble des modèles de retour vers la moyenne. Cette caractéristique n'apparaît a priori pas justifiable économiquement. Pour plus de réalisme, il s'agirait d'utiliser un modèle plus ambitieux incorporant les contraintes macro-économiques pesant sur l'inflation (chômage, lien entre taux d'intérêt et taux d'inflation...).

- **Quant aux trajectoires de taux simulées :**

On rappelle que le processus d'Ornstein-Uhlenbeck n'exclut pas les valeurs négatives ou nulles. Cette caractéristique prend effet dans le calcul du risque de provision, au moment d'estimer l'impact d'un choc d'inflation sur la courbe des taux à un an. Mais si la production de valeurs négatives peut être considérée comme problématique dans un contexte de simulation de taux d'intérêt, elle est en revanche nécessaire dans le cas de l'inflation. En témoignent les trajectoires des taux d'inflation présentées dans la section suivante.

III.2 Le calibrage des paramètres

En nous plaçant en Septembre 2009, nous nous affranchissons de toute cohérence de temps avec les données utilisées⁷⁵ pour projeter l'inflation future. En ce sens, nous considérons comme virtuel le portefeuille utilisé dans la construction du modèle de provisionnement.

Tout comme Ahlgrim et al., nous estimons les paramètres β et η à partir des indices des prix à la consommation produits par le Bureau des Statistiques du Travail américain⁷⁶. Selon les auteurs, l'erreur de mesure courante sur ces indices mensuels peut biaiser l'estimation des paramètres. Afin de limiter ce phénomène, ils préconisent d'opérer une régression annuelle sur les indices.

On est conduit à considérer le rapport

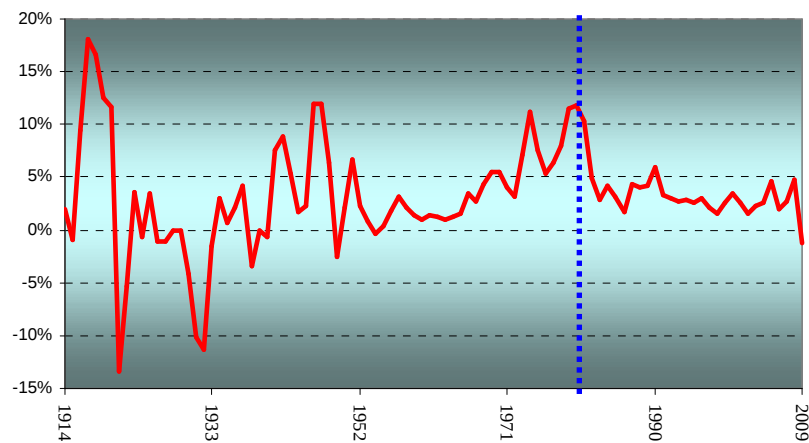
$$i_t = \ln \frac{CPI_t}{CPI_{t-1}}$$

où CPI_t est l'indice des prix à la consommation en t et CPI_{t-1} celui comptabilisé l'année passée sur le même mois.

Les taux annuels d'inflation calculés à partir des indices des prix de septembre 1913 à septembre 2009 sont présentés à la figure 46. On y observe clairement une rupture à partir de 1983, tant en tendance qu'en volatilité.

⁷⁵ Nous rappelons que les données utilisées sont exprimées en dollars américains de 1995, année de la dernière survenance observée.

⁷⁶ La série désaisonnalisée « Consumer Price Index – All Urban Consumers » (ref. CUSR0000SA0) et le descriptif de sa construction sont disponibles sur www.bls.gov.

Figure 46 – Taux annuels d'inflation

Le tableau 19 présente sur différentes périodes la variance empirique σ_i des taux annuels d'inflation et les estimations des coefficients du modèle continu selon le schéma d'Euler et par la solution exacte. On remarque que l'estimation du paramètre β conduit à des résultats très différents selon la méthode retenue. On note en particulier que l'estimation par la solution exacte est davantage sensible aux changements dans l'échantillon d'ajustement⁷⁷.

Tableau 19 – Estimations des paramètres du modèle d'inflation

Echantillon d'ajustement	$\hat{\beta}$		$\hat{\eta}$		$\hat{\sigma}$	σ_i
	<i>Euler</i>	<i>Exact</i>	<i>Euler</i>	<i>Exact</i>	<i>Exact</i>	
1914-2009 (1)	0,40	0,52	3,16%	3,16%	0,21%	4,92%
1947-2009 (2)	0,38	0,48	3,27%	3,27%	0,07%	3,03%
1983-2009 (3)	1,03	- ⁷⁸	2,94%	-	-	1,36%
1983-1995 (4)	0,80	1,63	3,49%	3,49%	0,01%	1,09%
1983-2008 (5)	0,75	1,37	3,13%	3,13%	0,01%	1,09%

La comparaison des résultats issus des régressions (2) et (3) rappelle la rupture de 1983, caractérisée par une augmentation de la force de rappel et une diminution de la variance empirique des taux annuels. La valeur de ces deux paramètres apparaît stable sur la période 1983 - 2008, comme en témoignent les résultats des régressions (4) et (5).

Nous décidons de calibrer les paramètres du modèle d'inflation à partir des résultats de la régression (5) : nous fixons le paramètre de rappel β à 1,4 et la moyenne à long terme η à 3%. En ce sens, nous stipulons le maintien des caractéristiques du processus observé depuis 1983 sur les 23 années⁷⁹ de projection nécessaire à l'écoulement total du risque étudié. Notons que ce choix revient également à considérer le taux négatif de 2009 comme atypique, dès lors qu'il n'est pas incorporé dans l'échantillon d'ajustement.

⁷⁷ Dans le cas du schéma d'Euler, on a $\hat{\beta}_1 = \frac{1-\hat{a}}{\Delta t}$ contre $\hat{\beta}_2 = -\frac{\ln(\hat{a})}{\Delta t}$ pour la solution exacte. Comme $\hat{a} < 1$, on a $\frac{\partial}{\partial \hat{a}} \hat{\beta}_2 > \frac{\partial}{\partial \hat{a}} \hat{\beta}_1$.

⁷⁸ Comme $\hat{a} < 0$, le calcul de $\hat{\beta}$ est impossible.

⁷⁹ Se reporter au tableau 19.

D'autre part, on observe dans le cas de la solution exacte que l'estimation simultanée des paramètres β , η et σ conduit à des taux d'inflation quasi déterministes de par la petitesse de σ . L'absence de réalisme de ce résultat nous conduit à fixer arbitrairement la valeur de σ , par référence à l'écart type empirique. La valeur de 4% retenue pour ce paramètre est justifiée dans la section suivante.

III.3 La validation du modèle par back testing

Au-delà de la précision de la méthode d'estimation des paramètres et de l'actualité de l'échantillon d'ajustement, le test majeur de validité du modèle est celui de la cohérence des résultats. Cette section analyse les caractéristiques des trajectoires de taux simulées et les compare avec les valeurs historiques. L'objectif ne consiste évidemment pas à vérifier la réplication du passé par le modèle mais plutôt à s'assurer de la plausibilité de ses résultats.

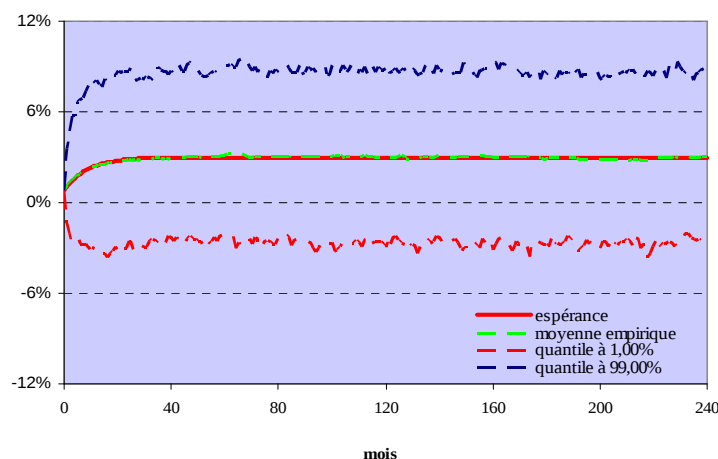
Afin de tester cette plausibilité, on simule 1000 trajectoires de taux mensuels annualisés⁸⁰ sur 20 années à partir du schéma d'Euler pour les paramètres retenus à la section précédente :

$$i_0 = 0,8\% \quad \beta = 1,4 \quad \eta = 3\% \quad \sigma = 4\% \quad \Delta t = \frac{1}{12}.$$

La valeur d'initialisation i_0 correspond au taux mensuel observé en septembre 2009. La valeur du paramètre de diffusion σ est fixée à 4%. Elle correspond à la valeur de l'écart type des taux mensuels annualisés observés sur la période courant de février 1983 à septembre 2009. Nous montrons plus en avant que le niveau de 4% retenu aboutit à une distribution des taux simulés cohérente avec les observations effectuées depuis 1983. Les caractéristiques majeures des trajectoires de taux sont présentées à la figure 47 et au tableau 19.

La figure 47 présente sur les 20 années de projection les trajectoires moyennes (empirique et selon le modèle continu) et les quantiles à 0,5% et 99,5% des taux simulés. On observe par exemple que le niveau retenu des paramètres σ et β entraîne la stabilisation du taux moyen d'inflation autour de son niveau de long terme au bout de trois années.

Figure 47 – Taux d'inflation mensuels sur 20 années



⁸⁰ Si i_m désigne le taux d'inflation mensuel déterminé par le rapport des indices de prix de deux mois successifs, le taux d'inflation mensuel annualisé associé est égal à la quantité $(1+i_m)^{12} - 1$.

Les taux $i_{t,H}$ reportées au tableau 20 sont calculées à partir des taux simulés. On désigne par $i_{t,H}$ le taux moyen géométrique calculé en début de mois t sur un horizon de H mois :

$$i_{t,H} = \left(\prod_{j=1}^H (1 + i_{t+j-1}) \right)^{\frac{1}{H}} - 1$$

Les valeurs moyennes des variables $i_{t,H}$ sont reportées pour les 1^{er}, 13^{ème} et 121^{ème} mois de projection et leurs quantiles à 0,5% et à 99,5% sont indiqués pour le 121^{ème} mois. Par exemple, le tableau 19 montre que le taux mensuel moyen initialisé à 0,8% passe à 2,5% dès le 13^{ème} mois de simulation. De même on observe que les quantiles à 0,5% et à 99,5% des taux mensuels simulés sur le 121^{ème} mois sont respectivement de -5,8% et 9%.

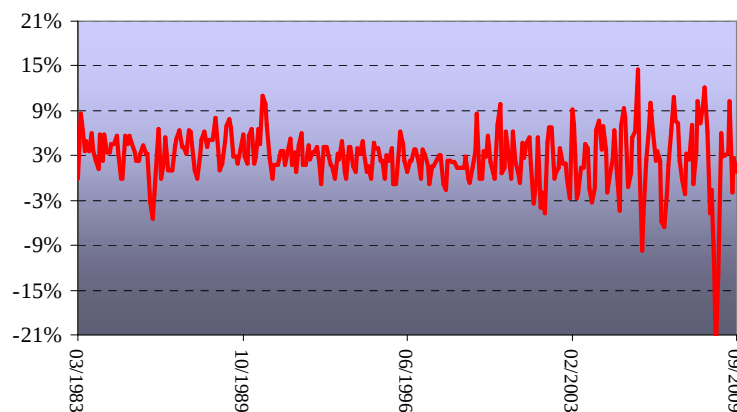
Tableau 20 – Résultats des 1000 simulations de taux

Taux d'inflation	Moyenne			Quantiles	
	t = 1	t = 13	t = 121	t = 121	
				0,5%	99,5%
$i_{t,1}$	1,0%	2,5%	3,0%	-5,8%	9,0%
$i_{t,12}$	1,9%	2,7%	3,0%	-2,6%	8,1%
$i_{t,120}$	2,8%	2,8%	2,9%	0,1%	5,0%

La figure 48 permet d'appréhender le réalisme de ces premiers résultats. Elle présente les taux mensuels annualisés d'inflation⁸¹ enregistrés depuis mars 1983. Comme l'enseigne son observation, il est plausible de simuler des valeurs négatives de taux d'inflation, dès lors qu'il existe des observations négatives. De même, des valeurs excédant 9% sont observées.

Figure 48 – Taux mensuels annualisés d'inflation

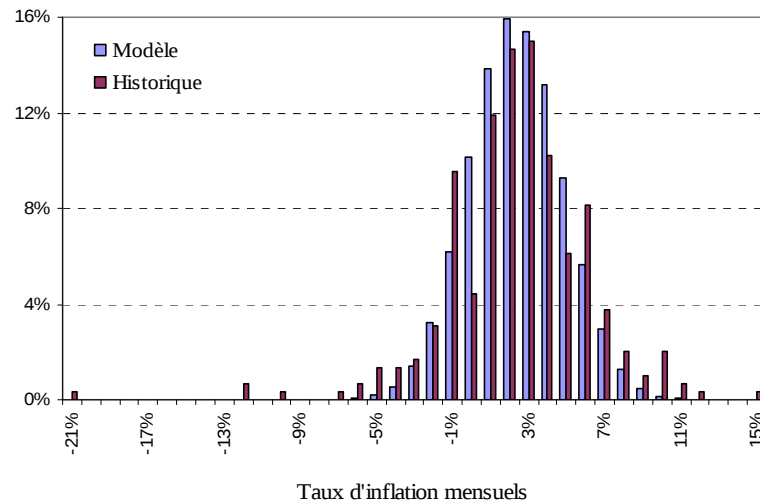
mars 1983 – septembre 2009



⁸¹ Soient les taux $\left(\frac{CP_m}{CP_{m-1}} \right)^{12} - 1$, où CP_m désigne l'indice des prix du mois m .

La figure 49 fournit une indication plus précise sur la plausibilité des simulations. Elle juxtapose l'histogramme des taux mensuels annualisés observés de février 1983 à septembre 2009 et celui des valeurs projetées sur les 20 années de simulation. Il apparaît clairement que la distribution des taux historiques est plus heurtée que celle des taux simulés.

Figure 49 – Taux d'inflation mensuels annualisés historiques et simulations



Ce phénomène provient tout d'abord du fait que l'histogramme résume seulement 294 valeurs historiques contre 1000×240 dans le cas des simulations. Ensuite, les valeurs simulées sont produites par un modèle mathématique tandis que les valeurs historiques résultent de l'interaction d'une multitude de variables.

La question principale est de savoir si le modèle retranscrit de façon réaliste la distribution de probabilité des taux. A ce sujet, on remarque que la distribution historique présente des queues plus épaisses que celle des taux gaussiens simulés. On parle de taux historiques leptokurtiques. En revanche, le cœur de la distribution des simulations apparaît en phase avec celui des valeurs historiques.

L'observation du tableau 19 permet également de considérer les projections de taux comme raisonnables, sur un horizon court comme de long terme :

Sur un horizon annuel, le modèle prévoit une inflation moyenne de 1,9% pour la 1^{ère} année de simulation. Le taux moyen passe à 3% pour la 13^{ème} année de simulation, en accord avec les paramètres du modèle. Dans 0,5% des cas, les taux annuels simulés sur la 11^{ème} année sont inférieurs à -2,6%. A ce sujet, l'observation de la figure 46 nous montre que l'expérimentation d'une déflation sur une année entière reste un phénomène rare pour les années postérieures à 1983. Cependant, cet événement reste envisageable, comme dans le cas de l'année 2009. Il apparaît donc naturel d'attribuer une probabilité non nulle à sa survenance.

Quant à la prévision des taux de long terme, on s'attend à une moindre incertitude du fait du mécanisme de retour à la moyenne. Ce postulat est confirmé par les simulations. Le modèle prévoit un taux moyen annuel de 2,8% pour les 10 premières années de projection, et de 2,9% pour la seconde décennie, niveau proche du taux de long terme. On remarque également par la moindre amplitude des quantiles des taux simulés sur la seconde décennie que le taux moyen géométrique est moins variable sur les horizons longs.

III.4 La modélisation de l'inflation dans les prestations futures

On estime l'évolution de l'inflation à partir des taux simulés. A partir de t_0 , la date de calcul de la provision, on découpe le temps réel en intervalles mensuels.

Autrement dit, on définit le m^e intervalle par l'ensemble des valeurs de temps réel $\left\{ t \mid t > t_0 + \frac{m-1}{12}, t \leq t_0 + \frac{m}{12} \right\}$. Sur le m^e intervalle, on suppose que la variation dCP_t de l'indice des prix CP_t se fait au taux espéré

$$i_m = i_0 \cdot e^{-\hat{\beta} \frac{m}{12}} + \hat{\eta} \left(1 - e^{-\hat{\beta} \frac{m}{12}} \right), \quad \text{pour } m \geq 1.$$

où $\hat{\beta}$ et $\hat{\eta}$ sont les estimations des paramètres du modèle calculées précédemment.

Autrement dit, on a pour $t \in \left] t_0 + \frac{m-1}{12}, t_0 + \frac{m}{12} \right]$:

$$dCP_t = CP_t \cdot i_m \cdot dt.$$

On rappelle que cette équation différentielle s'intègre immédiatement :

$$\begin{aligned} \int_{t=t_0+\frac{m-1}{12}}^{t=t_0+\frac{m}{12}} \frac{dCP_t}{CP_t} &= i_m \int_{t=t_0+\frac{m-1}{12}}^{t=t_0+\frac{m}{12}} dt \\ \Leftrightarrow \ln(CP_{t_0+\frac{m}{12}}) - \ln(CP_{t_0+\frac{m-1}{12}}) &= \frac{i_m}{12} \\ \Leftrightarrow CP_{t_0+\frac{m}{12}} &= CP_{t_0+\frac{m-1}{12}} \cdot e^{\frac{i_m}{12}} \end{aligned}$$

On désigne par $d = 12 \times (t - t_0)$ le nombre de mois écoulés depuis t_0 et par $[.]$ l'opérateur partie entière. En posant $CP_{t_0} = 1$, on obtient par récurrence l'indice d'inflation en base t_0 :

$$\hat{CP}_{t-t_0} = \hat{CP}_{\hat{H}(\tau_0, \tau)} = \exp \left[\frac{1}{12} \cdot \left(I_{\hat{d} \geq 1} \cdot \sum_{j=1}^{\left[\hat{d} \right]} \hat{i}_j + \left(\hat{d} - \left[\hat{d} \right] \right) \cdot \hat{i}_{\left[\hat{d} \right] + 1} \right) \right], \quad (12)$$

L'intégration de l'indice d'inflation dans (5) définit le coût réel d'un sinistre futur :

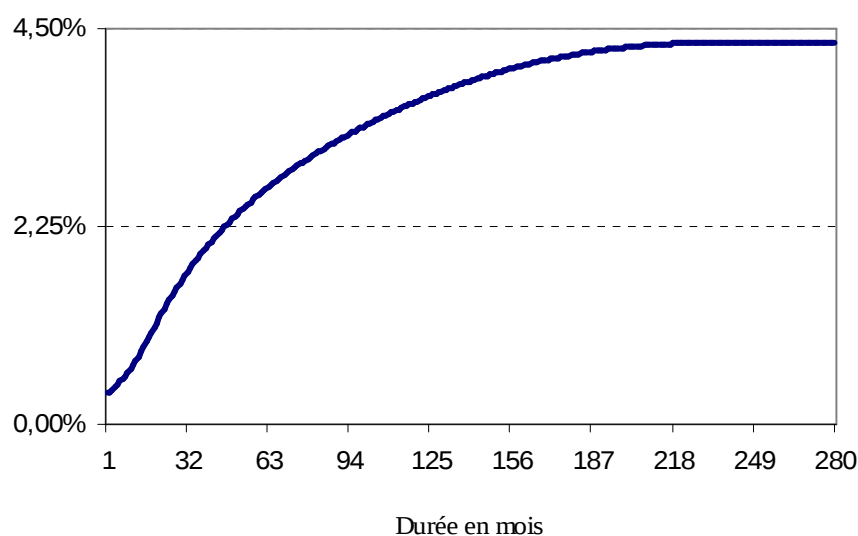
$$m_\tau = \hat{CP}_{\hat{H}(\tau_0, \tau)} \exp \left[\hat{H}(0, \tau) \cdot \left(I_{\tau \in I_2} \cdot \hat{\xi}_{2,2}^c + \hat{\xi}_3^c \right) \right] \cdot \hat{m}_\tau^1 \quad (13)$$

IV. L'évaluation du Best Estimate des PSAP

On définit le Best Estimate des provisions pour sinistres par la somme actualisée à la courbe des taux sans risque des prestations futures définies par l'équation (13).

En ce qui concerne l'actualisation des flux, il s'agit de tenir compte d'un taux sans risque de marché « pertinent⁸² » pour des instruments de même durée. Nous choisissons la courbe des taux zéro coupon publiée par la commission « Indices et références du marché » de l'Institut des Actuaire au 30 Septembre 2009. On remarque que le profil ascendant de la courbe présentée à la figure 50 correspond au cas le plus courant d'anticipations haussières des taux spots⁸³.

Figure 50 – Courbe des taux zéro coupon au 30 septembre 2009



Par ailleurs, si les marchés ne recèlent pas l'information complète, elle doit être extrapolée à partir des taux observés. A ce sujet, les principes MCEV⁸⁴ en voie de convergence avec le référentiel Solvabilité II autorisent les travaux d'interpolation lorsque, comme c'est notre cas, des maturités sont manquantes. Dans l'actualisation des flux, nous utilisons en ce sens les taux zéro coupon interpolés linéairement.

En ce qui concerne les flux constitutifs du Best Estimate, on rappelle que le coût d'un sinistre pour un temps opérationnel donné est supposé évoluer dans le temps selon l'effet conjugué de l'inflation et de la surinflation. L'indice annuel moyen de surinflation a été estimé dans la partie II : il est constant et égal à 3,64%. Insistons une fois encore sur la difficulté d'extrapoler la surinflation des coûts compte tenu des nombreux facteurs juridico socio économiques qui en sont à l'origine. Dans ce contexte, la reproduction dans le futur de la surinflation passée constitue un choix de modélisation parmi d'autres.

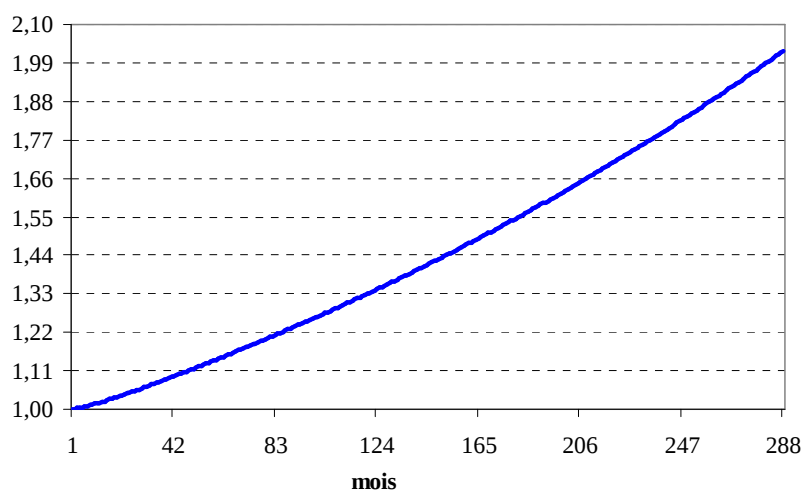
⁸² Article 76 de la Directive Européenne (Parlement Européen, 2009).

⁸³ On rappelle qu'on peut écrire le taux long $r_p(0)$ (à p périodes) comme la moyenne géométrique du taux court $r_1(0)$ (à une période) et des taux forward $f_{i,1}$ à court terme, avec $(1+r_p(0))^p = (1+r_1(0))(1+f_{1,1}(0))\dots(1+f_{p-1,1}(0))$, où le taux $f_{i,1}(0)$ implicite dans la gamme représente à l'équilibre la valeur anticipée aujourd'hui du taux à une période $r_1(p)$ qui prévaudra dans p périodes, majorée d'une prime de liquidité.

⁸⁴ Se reporter au principe 14 « Risk Free Rates » du CFO Forum (2009), « MCEV Principles and Guidance ».

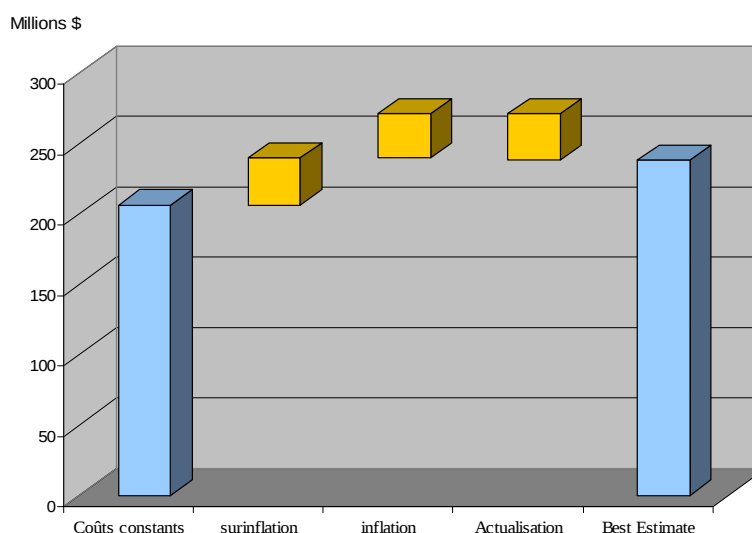
En ce qui concerne l'indice d'inflation en base t_0 , on estime sa progression moyenne à partir de l'équation (12), définissant la meilleure estimation de l'inflation future. Compte tenu des estimations des paramètres β et η , on obtient la figure 51.

Figure 51 – Représentation de l'indice moyen de l'évolution des prix⁸⁵



L'observation de gauche à droite de la figure 52 permet de mesurer l'impact des effets de taux sur l'estimation de la meilleure estimation des PSAP :

Figure 52 - Éléments constitutifs du Best Estimate



Le premier bloc « coûts constants » représente l'estimation de la somme des prestations futures sans prise en compte du phénomène de surinflation. Le Best Estimate des PSAP est constitué de la somme du premier bloc et des blocs en jaune représentatifs des effets de taux. Les résultats sont repris par survenance au tableau 21.

⁸⁵ dont le taux évolue selon le modèle de Vasicek ajusté par discrétisation exacte.

La colonne (1) du tableau 21 correspond au premier bloc « coûts constants » de la figure 52. On remarque que la prise en compte du phénomène de surinflation augmente de 16% la somme des prestations futures. L'intégration de l'inflation conduit à une hausse globale de 13% des prestations surinflatées. Cet effet est compensée par celui de l'actualisation des flux.

Tableau 21 – Estimation du Best Estimate

i	Coûts constants	Effets inflationnistes		Best Estimate
		Surinflation	Inflation	Actualisation
	(1)	(2)	(3)	(4)
1	0	0	0	0
2	0	0	0	0
3	1 135 357	1 359 150	1 588 721	1 328 522
4	91 700	101 318	108 415	103 319
5	463 543	526 503	582 161	526 305
6	1 036 494	1 180 404	1 311 354	1 176 215
7	1 233 804	1 441 796	1 644 261	1 422 876
8	2 140 711	2 495 436	2 843 281	2 463 043
9	4 274 419	4 923 057	5 549 461	4 876 338
10	7 031 223	8 002 984	8 920 753	7 954 372
11	10 050 287	11 337 042	12 530 107	11 294 968
12	12 478 001	14 105 465	15 625 621	14 045 563
13	19 446 332	22 007 342	24 369 346	21 949 841
14	27 611 500	31 110 884	34 223 881	31 143 842
15	35 358 147	40 728 564	45 630 397	40 653 351
16	37 477 628	44 298 879	50 690 620	44 030 834
17	46 282 277	55 901 796	65 332 964	55 022 806
	206 111 422	239 520 620	270 951 344	237 992 195

De façon générale, la méthode retenue allonge sensiblement l'horizon de projection des prestations par rapport à la méthode Chain Ladder. Pour appréhender l'impact des effets de taux sur le niveau des prestations, il est intéressant d'observer les caractéristiques des derniers sinistres. Au tableau 22 sont présentés les estimations des délais de règlement, des facteurs d'inflation, de surinflation et des prestations des trois derniers sinistres réglés.

Tableau 22 – Inflation dans les prestations versées au titre des trois derniers règlements

Délai	Prestation	Facteur de surinflation	Facteur d'inflation	Facteur de croissance
en années	en dollar 1995	$e^{H(0,\tau)\xi_3^c}$	$e^{CP_{H(t_0,r)}}$	
		(1)	(2)	(1)×(2)
18,21	180 026	1,94	1,70	3,30
19,81	189 735	2,06	1,79	3,67
23,27	213 882	2,33	1,98	4,62

On remarque que le modèle estime qu'en vingt trois ans (depuis t_0), le coût du dernier sinistre réglé aura plus que doublé ($\times 2,33$) du fait de la surinflation, et plus que quadruplé ($\times 4,62$) de par l'effet conjugué de l'inflation et de la surinflation.

CHAPITRE II. LE CALCUL DU RISQUE DE PROVISIONNEMENT

I. L'exposé du problème

La problématique du risque d'insuffisance des provisions à couvrir les prestations futures jusqu'à l'écoulement total des sinistres fait l'objet d'une littérature abondante⁸⁶. Pour mesurer ce risque dit ultime, nous nous plaçons aujourd'hui en t_0 dans le cadre du modèle de provisionnement spécifié en seconde partie (équation 20).

Si $R^{(0)}$ désigne la provision estimée aujourd'hui et $X_{i\tau}$, la prestation payée au titre du sinistre de temps opérationnel τ et de survenance i , la mesure du risque ultime passe par l'estimation de la distribution de probabilité de la variable

$$R^{(0)} - \sum_{i=1}^I \sum_{\tau=\tau_{i,1}}^{\tau_{i,\omega_i}} X_{i\tau}.$$

Une fois la distribution estimée, il est par exemple possible à partir d'un de ses quantiles de déterminer un niveau de marge de prudence. On conçoit donc que l'analyse du risque ultime soit une problématique centrale dans l'étude de la solvabilité à long terme de l'assureur.

Toutefois, cette problématique ne correspond pas à celle réglementaire de calcul du capital requis, qui s'appuie sur une mesure de perte (approche VaR) à horizon d'un an. Pour couvrir le risque de provisionnement, le référentiel prudentiel définit ainsi la charge de capital à immobiliser par le quantile à 99,5% de la perte à un an du portefeuille en continuité d'activité.

Si I désigne le montant des intérêts distribués durant l'année à venir, $\lambda_i = N_{i,t_0+2-i}$ le nombre de règlements effectués durant cette période au titre de la survenance i et $R^{(1)}$ la provision estimée dans un an, il s'agit d'estimer la distribution du résultat de liquidation :

$$R^{(0)} + I - \left(\sum_{i=1}^I \sum_{\tau=\tau_{i,1}}^{\tau_{i,\lambda_i}} X_{i\tau} + R^{(1)} \right).$$

Pour calculer la charge en capital au titre du risque de provisionnement et la combiner avec celles d'autres risques de sorte à évaluer le capital global de solvabilité, il est donc nécessaire d'estimer la distribution du résultat de liquidation des sinistres en cours. Dans certains cas, des réponses analytiques existent. Notons en particulier l'approche de M. Wütrich et M. Merz (2007) qui, en extension du modèle stochastique Chain Ladder de Mack (1993), fournit une solution analytique de l'erreur quadratique moyenne de prédiction du résultat de liquidation.

Mais dans la plupart des cas tel celui de notre modèle, où il n'existe pas d'expression analytique de la distribution, les méthodes simulatoires constituent l'unique solution. A ce sujet, Björkwall et al. (2008) décrivent la méthode de calcul de la distribution du résultat à un an dans un cadre Bootstrap. Dans le calcul du risque de provisionnement, nous privilégions ce type d'approche en l'adaptant à notre modèle de provisionnement.

⁸⁶ Se reporter par exemple à Mack (1993) ou England et Verrall (1999).

II. Le calcul du risque à l'ultime

II.1 La méthode du Bootstrap

II.1.1 Notations

Avant d'aborder la procédure de calcul, nous introduisons une notation adaptée à sa présentation. Comme il l'a été précisé en début de partie, l'hypothèse fondamentale du calcul de provisionnement dans le cadre de notre modèle repose sur l'assimilation en espérance des coûts moyens et des prestations individuelles. Ce postulat revient dans la pratique à assimiler les observations de coûts moyens $S_{i\tau}$ à des prestations individuelles $X_{i\tau}$. De ce fait, la notation utilisée dans cette section ne distingue pas les deux notions.

Nous considérons les données sous la forme du triangle de liquidation constitué des observations $\{x_{i\tau}; i, \tau \in \nabla\}$, où ∇ désigne le triangle supérieur. L'objectif consiste à estimer la distribution des prestations futures $\{X_{i\tau}; i, \tau \in \Delta\}$, où $\Delta = \{i = 1, \dots, I; \tau = \tau_{i1}, \dots, \tau_{i\omega_i}\}$.

Nous rappelons que notre modèle donne une estimation $\hat{m}_{i\tau}$ de $m_{i\tau} = E(X_{i\tau})$ pour chaque $X_{i\tau}$ des ensembles Δ et ∇ . Ces estimations sont calculées en maximisant la vraisemblance de l'échantillon gamma $\nabla X = \{X_{i\tau}, i, \tau \in \nabla\}$.

II.1.2 La sélection des résidus

Dans le cadre statistique, la méthode du Bootstrap désigne une technique consistant à obtenir des informations sur une loi à partir d'un échantillon de celle-ci. La technique est basée sur le rééchantillonnage des données de l'échantillon selon un tirage aléatoire avec remise. Son utilisation repose sur l'indépendance et l'équidistribution des variables sur lesquelles on effectue le rééchantillonnage.

Dans notre modèle, on considère les observations du triangle supérieur comme les réalisations de variables aléatoires $X_{i\tau}$. Il est donc naturel de penser à rééchantillonner ces observations. Cependant, si les $X_{i\tau}$ sont supposées indépendantes, ils ne sont pas identiquement distribués, dès lors qu'on les munit des lois gamma $G\left(\frac{1}{\phi}, \frac{1}{\phi \cdot m_{i\tau}}\right)$.

Comme les $X_{i\tau}$ ne sont pas i.i.d, on procède au rééchantillonnage des résidus d'estimation. Dans le cadre du Bootstrap, nous privilégions les résidus de Pearson dont nous rappelons la définition dans le cadre gamma du modèle de provisionnement⁸⁷ :

$$e_{i\tau}^{(p)} = \frac{X_{i\tau} - \hat{E}(X_{i\tau})}{\sqrt{\hat{V}(X_{i\tau})}} = \frac{X_{i\tau} - \hat{m}_{i\tau}}{\sqrt{\phi \cdot \hat{m}_{i\tau}}}, \quad \text{pour } i, \tau \in \nabla.$$

Deux raisons motivent ce choix. Tout d'abord, on montre en annexe P que ces résidus sont équidistribués si comme nous le supposons les $X_{i\tau}$ sont de loi gamma. Ensuite, le calcul des résidus de Pearson est relativement peu consommateurs en temps de calcul.

⁸⁷ L'expression de la variance, de l'espérance et des paramètres de la loi gamma sont justifiés en annexe P.

En suivant les préconisations d'England et Verrall (1999), nous procédons également à un ajustement global de ces résidus. Cet ajustement doit permettre de corriger le biais dans la comparaison des estimations analytiques et Bootstrap de l'erreur de prédiction :

$$e_{i\tau}^{(pa)} = \sqrt{n/n-p} \cdot e_{i\tau}^{(p)}, \quad \text{pour } i, \tau \in \mathbb{V},$$

Où n et p désignent le nombre d'observations et p le nombre de ses variables explicatives. Comme les résidus de Pearson sont en moyenne proches de zéro, cette opération a pour conséquence d'augmenter la variance des résidus, en affectant peu leur moyenne.

II.1.3 L'implémentation du Bootstrap

La procédure de calcul utilise le triangle des résidus ajustés de Pearson comme les données d'un échantillon. On en déduit la distribution prédictive des prestations futures et par extension celle du Best Estimate des PSAP.

La volatilité des estimations de prestations futures ne dépend jusqu'en 2.1.2 que de celle des résidus. Autrement dit, elle ne reflète que l'incertitude liée à la procédure d'estimation. Il s'agit alors de prendre en compte le risque d'estimation.

En complément, on modélise à partir de 2.1.3 l'incertitude liée à la variabilité intrinsèque des $\Delta X_{i\tau}$. Cette sous étape permet de prendre en compte le risque de process. Selon la modélisation retenue, la variabilité des prestations futures dépend de trois facteurs d'incertitude : l'inflation, la date de règlement et le niveau hors inflation de prestation. Ces trois éléments sont supposés indépendants et simulés séparément les uns des autres.

1. Préliminaires

- Estimation des $\nabla x_{i\tau}$ à partir du modèle de provisionnement et calcul des résidus associés $\nabla e_{i\tau}^{(pa)}$, dont on déduit l'estimation du coefficient de dispersion :

$$\hat{\phi} = \frac{1}{n-p} \sum_{\mathbb{V}} (e_{i\tau}^p)^2 = \frac{1}{n} \sum_{\mathbb{V}} (e_{i\tau}^{pa})^2.$$

2. Procédure itérative

Pour chaque itération $b = 1, \dots, B$, le processus se poursuit ainsi :

2.1 Calcul des espérances des prestations futures

o 2.1.1 : Génération d'un triangle de pseudo-données

- Création d'un échantillon de résidus noté $\nabla e_{i\tau}^{*b}$ par tirage aléatoire avec remise parmi les $\nabla e_{i\tau}^{pa}$.
- Détermination de l'échantillon associé des coûts moyens $\nabla x_{i\tau}^{*b}$ par

$$x_{i\tau}^{*b} = \hat{m}_{i\tau} + e_{i\tau}^{*b} \cdot \hat{m}_{i\tau}, \quad \text{pour } i, \tau \in \mathbb{V}.$$

o 2.1.2 : Estimation des paramètres de régression

A partir des ∇X_{it}^{*b} , on détermine les estimateurs du maximum de vraisemblance des paramètres de régression ξ^* du modèle, utilisés en 2.1.4 pour calculer l'espérance des prestations futures.

o 2.1.3 : Simulation des dates de règlement

Par le tirage aléatoire de variables uniformes $U[0;1]$, on simule les dates de règlement

$$t_{ij}^{*b} = t_0 - \frac{1}{c_i} \ln(1-u), \quad \text{pour } i = 1, \dots, I, j = 1, \dots, \omega_i$$

auxquelles on associe par ordre chronologique les temps opérationnels $\tau_{i1}, \dots, \tau_{i\omega_i}$.

o 2.1.4 : Estimation des prestations futures hors inflation

A partir des paramètres ξ^* estimés en 2.1, des t_{ij}^{*b} et des τ_{ij} associés, on calcule les prestations futures (hors inflation) prévues par le modèle :

$$\hat{m}_{i\tau_{ij}}^{*b} = E(X_{i\tau_{ij}}^{*b}) = \exp[\eta(\xi^{*b}, \tau_{ij}, t_{ij}^{*b})], \quad \text{pour } i = 1, \dots, I, j = 1, \dots, \omega_i.$$

2.2 Simulation des prestations futures

o 2.2.1 : Simulation des prestations futures hors inflation

On simule les réalisations x_{it}^{*b} de v.a.r X_{it}^{*b} par un tirage aléatoire selon une distribution

$$G\left(\frac{1}{\hat{\phi}}, \frac{1}{\hat{\phi} \cdot \hat{m}_{i\tau}^{*b}}\right).$$

o 2.2.2 : Simulation de l'inflation future

Par le tirage aléatoire de variables gaussiennes⁸⁸ ε , on simule les taux d'inflation à partir du modèle d'inflation ajusté au chapitre I :

$$i_{t+\frac{1}{12}} = \hat{\eta} \cdot \frac{\hat{\beta}}{12} + \left(1 - \frac{\hat{\beta}}{12}\right) \cdot i_t + \varepsilon \cdot \hat{\sigma} \cdot \sqrt{\frac{1}{12}}.$$

A partir de la trajectoire d'inflation simulée, on en déduit par l'équation (12) les indices d'inflation $CP_{t_{ij}^{*b}}^{*b}$ associés aux dates de règlement t_{ij}^{*b} .

⁸⁸ Les variables uniformes sont simulées à partir de l'algorithme du Tore et les gaussiennes à l'aide de l'algorithme de Gauss-Müller.

o 2.2.3 : Calcul d'une réalisation de R

On suppose que l'effet de l'actualisation est le même dans les scénarios bootstrap et dans le scénario du Best Estimate. En appliquant les indices d'inflation aux prestations futures simulées, et après actualisation à la courbe des taux, on obtient une réalisation de R :

$$R^{*b} = \sum_{\Delta} \hat{m}_{it}^{*b} \cdot CP_{t_{ij}}^{*b} \cdot e^{-t_{ij}^{*b} \cdot r(0, t_{ij}^{*b})}.$$

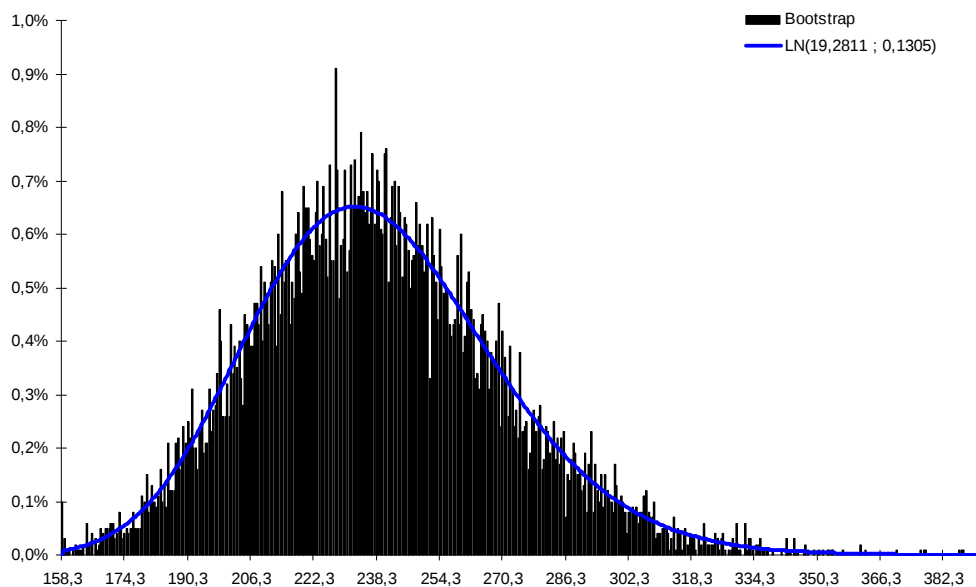
La fonction de répartition et les quantiles empiriques associés à la suite $(R^{*b})_{b=1, \dots, B}$ sont des estimations de la distribution prédictive et des quantiles théoriques de R.

II.2 L'application numérique

Le calcul du risque ultime est mené en deux temps. Nous estimons les paramètres de régression initiaux (étape 1) et les paramètres de régression bootstrapés (étape 2.1) à partir du logiciel SAS⁸⁹. Les calculs relevant de l'étape 2.2 sont menés sous tableur Excel, après récupération des paramètres de régression bootstrapés.

La distribution des provisions est obtenue empiriquement à partir de 10000 simulations. Elle est présentée à la figure 53 sous la forme d'un histogramme.

Figure 53 – Histogramme des provisions



On remarque que la densité empirique de la loi des provisions n'est pas symétrique, avec une queue de distribution étalée vers la droite. L'étude de l'étalement des queues de distribution se fait à l'aide du coefficient de dissymétrie (en anglais, skewness).

⁸⁹ La programmation sous SAS de la procédure figure en annexe Q.

Dans le cadre du Bootstrap, on utilise son estimateur naturel

$$\hat{\gamma}_1 = 1/B \cdot \sum_{b=1}^B \left(\left(R^{(b)} - \hat{E}(R) \right) / \sqrt{\hat{\text{var}}(R)} \right)^3,$$

où $\hat{E}(R)$ et $\hat{\text{var}}(R)$ sont issus de la distribution des $R^{(b)}$.

La valeur de l'estimateur Bootstrap du skewness présentée au tableau 23 témoigne de la dissymétrie à droite de la distribution des provisions⁹⁰. Rappelons que ce résultat correspond à un postulat souvent utilisée en pratique, consistant à supposer que la distribution des provisions admet une queue de distribution étalée à droite. Dans la modélisation de la distribution des provisions, il s'avère essentiel dans notre cas de tenir compte de cette dissymétrie à droite, de sorte à ne pas sous-évaluer le risque de perte mesuré par la VaR.

Dans la pratique, la loi log-normale est ainsi souvent privilégiée pour modéliser la distribution des provisions. Par exemple, cette hypothèse est celle adoptée dans la formule standard du risque de provisionnement décrite dans les spécifications techniques du QIS4. A partir des estimateurs Bootstrap de la moyenne et de l'écart type de R , on calibre les paramètres de $R \approx \text{LN}(m, s^2)$. Pour cela, on inverse les relations

$$\mu = e^{m+0.5s^2}, \quad \sigma^2 = (1 - e^{-s^2}) \cdot e^{2(m+s^2)}.$$

soit $m = \ln(\mu) - 0.5 \cdot s^2$, $s^2 = \ln\left(1 + \frac{\sigma^2}{\mu^2}\right)$, avec $\mu = E(R)$ et $\sigma = \text{Var}(R)$.

On obtient $R \approx \text{LN}(19.28, 0.13)$. Que ce soit graphiquement à partir de la figure 53, où numériquement à partir des quantiles et du skewness⁹¹ du tableau 23, on note la proximité entre la loi log-normale calibrée et la distribution empirique des provisions. On remarque aussi la forte proximité entre la moyenne empirique obtenue par Bootstrap et le Best Estimate.

Tableau 23 – Résultats Bootstrap

	Bootstrap	Log-normale
Moyenne	238 435 175	-
Ecart type	31 260 767	-
VaR _{75%}	258 096 612	258 173 116
VaR _{95%}	293 243 810	293 040 601
VaR _{99,50%}	332 531 877	330 911 134
Skewness	0,417	0,396

La charge en capital pour couvrir le risque de provisionnement correspond à l'excédent de la VaR au seuil de 99,5% par rapport à la provision Best Estimate, soit

$$C_{0,995} = 332\,531\,877 - 237\,992\,195 = 94\,539\,682.$$

⁹⁰ On rappelle que si $\gamma_1 = 0$, la distribution est symétrique; si $\gamma_1 > 0$, la distribution présente une queue étalée vers la droite; sinon, la distribution présente une queue étalée vers la gauche.

⁹¹ On rappelle que le skewness d'une loi log-normal est défini par $(e^{\sigma^2} + 2) \cdot \sqrt{e^{\sigma^2} - 1}$.

III. Le calcul du risque à un an

III.1 La modification de la méthode du Bootstrap

III.1.1 Principes

Dans le calcul du risque à un an, nous suivons les principes présentés par Björkwall et al (2008). Il s'agit de procéder à l'estimation de la distribution du Best Estimate à un an tout en conditionnant les calculs à l'information disponible en t_0 , soit le triangle de liquidation ∇X des coûts moyens vu en t_0 .

L'estimation $\hat{R}^{(0)}$ de la provision en t_0 est considérée fixe, tandis que la distribution des prestations de l'année à venir est obtenue en simulant B fois la nouvelle diagonale X^* du triangle de liquidation. Chaque élément de X^* est égal au coût moyen par année de survenance des prestations simulées sur l'année $\{X_{i\tau}; \tau = \tau_{i1}, \dots, \tau_{i\lambda_i}\}$.

De cette façon, on obtient B pseudo-triangles constitués du triangle ∇X et de la diagonale X^* . Pour chaque simulation, on recalcule la provision dans un an

$$\hat{R}^{(1)*} = \hat{R}^{(1)}(\nabla X \cup X^*).$$

Dans le calcul de $\hat{R}^{(1)*}$, le risque de taux n'est pas pris en compte. En d'autres termes, on suppose que l'effet de l'inflation et de l'actualisation sont les mêmes dans les scénarios bootstrap et dans le scénario du Best Estimate. En ce sens :

- Les indices d'inflation sont calculés en base t_0 à partir de la courbe Best Estimate des taux d'inflation calibrée au chapitre précédent.
- Dans l'actualisation, on utilise les taux forward⁹² déduits de la courbe des taux zéro coupon utilisée dans le calcul du Best Estimate. Autrement dit, les anticipations sous jacentes à la courbe ne sont pas révisées.

Finalement, on obtient B valeurs de

$$\hat{R}^{(0)} + I^* - \sum_{i=1}^I \sum_{\tau=\tau_{i,1}}^{\tau_{i,\lambda_i}^*} X_{i,\tau}^* - \hat{R}^{(1)*} \mid \nabla X,$$

dont on déduit la distribution du résultat de liquidation des sinistres en cours dans un an.

⁹² En désignant par $r_{0,m}$ le taux zéro coupon en t_0 d'échéance m mois, on rappelle que le taux forward $f_{12,m}$ d'échéance m mois dans 12 mois est tel que $(1+f_{12,m})^{\frac{m}{12}} = \frac{(1+r_{0,12+m})^{\frac{12+m}{12}}}{(1+r_{0,12})^{\frac{12}{12}}}$. Autrement dit, placer aujourd'hui sur 12+m mois doit équivaloir à placer aujourd'hui à un an ($m=12$) au taux $r_{0,12}$ et réinvestir dans un an à m mois au taux à terme $f_{12,m}$ anticipé aujourd'hui.

III.1.2 Algorithme

Afin de calculer le risque à un an, l'algorithme du Bootstrap est modifié comme suit :

1. Préliminaires

1.1 Ajustement du modèle

Cette étape fournit les estimations en t_0 des paramètres $\underline{\xi}$ du modèle de provisionnement et du coefficient ϕ de dispersion.

1.2 Calcul de la courbe des taux à un an

De la courbe des taux zéro coupon utilisés dans le calcul du Best Estimate, on déduit les taux forward dans un an tels que

$$(1 + f_{12,m})_{12}^m = \frac{(1 + r_{0,12+m})_{12}^{12+m}}{(1 + r_{0,12})_{12}^{12}},$$

où $r_{0,m}$ désigne le taux zéro coupon aujourd'hui d'échéances m mois et $f_{12,m}$ le taux forward dans un an d'échéance m mois.

2. Procédure itérative

Pour chaque itération $b = 1, \dots, B$, le processus se poursuit ainsi :

2.1 Simulation des prestations de l'année à venir

2.1.1 : Simulation des dates de règlement

- Par le tirage aléatoire de variables uniformes $U[0;1]$, on associe à chaque sinistre restant à régler une date de règlement

$$t_{ij}^{*b} = t_0 - \frac{1}{c_i} \ln(1 - u), \quad \text{pour } i = 1, \dots, I, j \in [1, \dots, \omega_i].$$

- On en déduit les dates de règlement des sinistres réglés dans l'année :

$$t_{i1}^{*b}, \dots, t_{i\lambda_i^{*b}}^{*b} \mid t_{ij}^{*b} \leq t_0 + 1 \forall j, \quad \text{pour } i = 1, \dots, I.$$

et les temps opérationnels associés

$$\tau_{i,1}^{*b} = \frac{N_{i,0} + 0,5}{N_i}, \dots, \tau_{i,\lambda_i^{*b}}^{*b} = \frac{N_{i,0} + \lambda_i^{*b} - 0,5}{N_i}.$$

2.1.2 Calcul des prestations « prévues » hors inflation

A partir du modèle GLM⁹³ et des couples simulés $(\tau_{ij}^{*b}, t_{ij}^{*b})$, on calcule les prestations espérées sur l'année :

$$\hat{m}_{i\tau_{ij}}^{*b} = \hat{E}(X_{i\tau_{ij}}^{*b}) = \exp\left[\hat{\eta}\left(\hat{\xi}, \tau_{ij}^{*b}, t_{ij}^{*b}\right)\right], \quad \text{pour } i = 1, \dots, I, j = 1, \dots, \lambda_i^{*b}.$$

2.1.3 Calcul des prestations réelles

- A partir des espérances $\hat{m}_{i\tau_{ij}}^{*b}$ calculées à l'étape précédente, on simule les réalisations $x_{i\tau_{ij}}^{*b}$ de v.a.r $X_{i\tau_{ij}}^{*b}$ par un tirage aléatoire selon une distribution

$$G\left(\frac{1}{\hat{\phi}}, \frac{1}{\hat{\phi} \cdot \hat{m}_{i\tau_{ij}}^{*b}}\right), \quad \text{pour } i = 1, \dots, I, j = 1, \dots, \lambda_i^{*b}.$$

- Selon la méthode déjà utilisée et à partir d'une trajectoire de taux simulée à l'aide du modèle ad hoc, on revalorise de l'inflation les prestations $x_{i\tau_{ij}}^{*b}$. On note $\tilde{x}_{i\tau_{ij}}^{*b}$ les prestations revalorisées.

2.2 Génération du pseudo triangle

2.2.1 Calcul des éléments de la nouvelle diagonale X^*

On calcule par survénance le coût moyen des prestations hors inflation simulées en 2.1.3 :

$$X_{i,\tau_{i,\lambda_i^{*b}}}^{*b} = \frac{\sum_{j=1}^{\lambda_i^{*b}} x_{i,\tau_{ij}}^{*b}}{\lambda_i^{*b}} \quad \text{pour } i = 1, \dots, I,$$

2.2.2 Génération du pseudo triangle $\nabla X \cup X^*$

On ajoute au triangle de liquidation ∇X les éléments de la nouvelle diagonale X^* composée des coûts moyens simulés en 2.2.1, en leur associant les temps opérationnels $\tau_{i,\lambda_i^{*b}}$ des derniers sinistres réglés⁹⁴.

⁹³ Avec $\hat{\eta}\left(\hat{\xi}, \tau_{ij}^{*b}, t_{ij}^{*b}\right) = \hat{\xi}^s + \hat{\xi}_1^{to} \cdot \left(0,2 - \tau_{ij}^{*b}\right)^+ + \hat{\xi}_3^{to} \cdot \left(\tau_{ij}^{*b} - 0,6\right)^+ + \hat{\xi}_4^{to} \cdot \left(\tau_{ij}^{*b} - 0,9\right)^+ + \hat{\xi}_3^c \cdot t_{ij}^{*b}$.

⁹⁴ De cette façon, on respecte la structure du triangle de liquidation, constitué des coûts moyens calculés à partir des prestations annuelles retraitées de l'inflation par Taylor.

2.3 Calcul du résultat de liquidation des sinistres en cours

2.3.1 Ajustement du modèle

On met à jour le modèle de provisionnement de l'information de l'année : à partir du pseudo triangle $\nabla X \cup X^*$, on ajuste les coefficients de régression ξ^* du modèle.

2.3.2 Calcul de la provision dans un an

Après actualisation à la courbe des taux forward, on obtient une réalisation de $R^{(1)}$:

$$\hat{R}^{(1)*b} = \sum_{\Delta} \hat{m}_{i\tau}^{*b} \cdot CP_{t_{ij}^{*b}} \cdot e^{-t_{ij}^{*b} \cdot f(0, t_{ij}^{*b})},$$

où les indices⁹⁵ d'inflation CP_t sont calculés en base t_0 à partir de la courbe des taux d'inflation Best Estimate.

2.3.3 Calcul des intérêts

Un calcul rigoureux nécessiterait en 2.1.3 de capitaliser les prestations simulées au taux zéro coupon de l'échéance du mois du règlement. Pour des raisons de temps de calcul, les intérêts I^* sont estimés par le produit de la provision moyenne et du zéro coupon à six mois.

2.3.4 Calcul du résultat de liquidation des sinistres en cours

On obtient une réalisation du résultat de liquidation des sinistres en cours :

$$\hat{R}^{(0)} + I^{*b} - \sum_{i=1}^I \sum_{\tau=\tau_{i,1}}^{\tau_{i,\lambda_i^{*b}}} \tilde{x}_{i,\tau}^{*b} - \hat{R}^{(1)*b} \mid \nabla X$$

III.1.3 L'application numérique

Une itération est constituée de quatre étapes principales: la première consiste dans le calcul des prestations réglées durant l'année, la seconde dans celui des coûts moyens constitutifs de la nouvelle diagonale du triangle, la troisième dans le calcul de la provision à un an et la dernière dans celui du résultat de liquidation. Les calculs sont effectués sous Excel, sauf dans le cas du réajustement du modèle au triangle vu à un an qui est mis en œuvre sous SAS.

■ Calcul des prestations de l'année

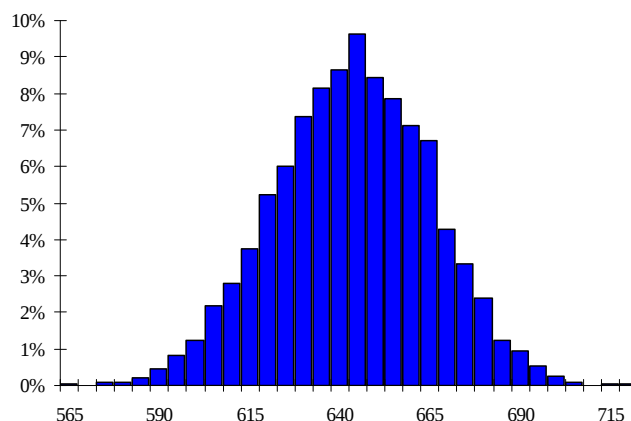
Nous simulons les prestations réglées sur l'année, brutes et nettes d'inflation. Les montants hors inflation sont utilisés dans le calcul des coûts moyens constitutifs de la nouvelle diagonale du triangle. Les montants inflatés correspondent aux montants « réels » payés sur l'année à venir. Ils sont utilisés dans le calcul du risque de provision.

⁹⁵ Se reporter au III.4 du chapitre précédent pour un rappel sur la construction de l'indice.

- **Calcul hors inflation**

La première étape consiste à déterminer le nombre et la date de règlement des sinistres réglés dans l'année à venir, soit entre t_0 et t_0+1 . Cet exercice passe par la simulation des dates de règlement des sinistres restant à régler. Selon la méthode décrite à l'étape 2.1.1, le nombre de sinistres réglés dans l'année correspond pour chaque survenance i aux $(N_i - N_{i,0})$ tirages dans la loi exponentielle de paramètre c_i pour lesquels $t - t_0 \leq 1$. La répartition des nombres de règlements simulés sur l'année est exposée à la figure 54.

Figure 54 - Répartition du nombre des règlements de l'année

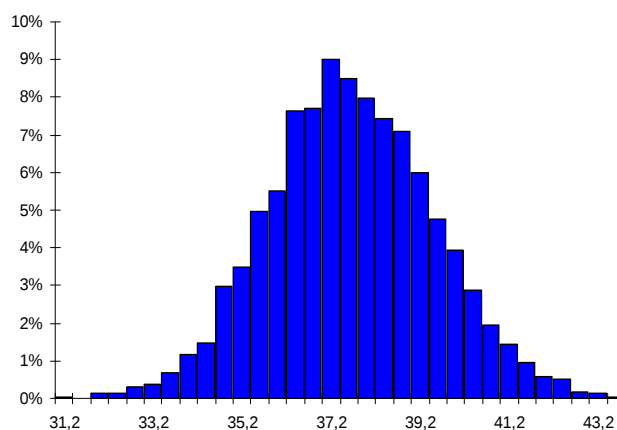


A chaque itération, on suppose connues les espérances des prestations de l'année. Ces espérances sont calculées selon les dates de règlement simulées, et à partir du modèle de provision utilisé dans le calcul du Best Estimate. La connaissance des prestations espérées permet de calibrer les lois gamma à partir desquelles on simule hors inflation les prestations de l'année (1^{ère} partie de l'étape 2.1.3).

- **Calcul avec inflation**

A chaque itération, on simule une trajectoire d'inflation à partir du modèle ad hoc. A partir de cette trajectoire, on revalorise de l'inflation les prestations simulées dans l'année (2^{ème} partie de l'étape 2.1.3). Leur répartition est présentée en figure 55.

Figure 55 - Répartition des prestations de l'année à venir
en millions de dollars



■ Calcul de la nouvelle diagonale

A chacune des B itérations, on forme la matrice des coûts moyens par survenance (étape 2.2.1) et des variables explicatives associées⁹⁶. La matrice est présentée dans la partie gauche du tableau 24 dans le cas de la 1^{ère} simulation. Elle correspond au 1^{er} des B blocs constitutifs de la table « Diag » du programme SAS présenté en annexe S. La partie droite du tableau 24 est présentée à titre informatif. On y voit les prestations avec inflation et le nombre des règlements simulés sur l'année.

Tableau 24 – Coûts moyens, variables explicatives, prestations réelles et nombre de sinistres réglés dans l'année, pour b = 1

n_simu	i	t	τ	z_r^s	$z_{r,1}^{to}$	$z_{r,3}^{to}$	$z_{r,4}^{to}$	$z_{r,1,1}^c$	$z_{r,1,2}^c$	$z_{r,2,1}^c$	$z_{r,2,2}^c$	$z_{r,3}^c$	S_r	$\tilde{x}_{i,\tau}^{*1}$	λ_i^{*1}
1	3	18	0,99	1	0	0,39	0,09	0	0	0	0	18	75 626	230 706	3
1	4	18	1,00	1	0	0,40	0,10	0	0	0	0	18	0	0	0
1	5	18	1,00	1	0	0,40	0,10	0	0	0	0	18	116 700	120 381	1
1	6	18	0,99	1	0	0,39	0,09	0	0	0	0	18	100 307	309 280	3
1	7	18	0,99	1	0	0,39	0,09	0	0	0	0	18	69 230	141 370	2
1	8	18	0,99	1	0	0,39	0,09	0	0	0	0	18	113 854	925 975	8
1	9	18	0,96	1	0	0,36	0,06	0	0	0	0	18	89 035	629 624	7
1	10	18	0,95	1	0	0,35	0,05	0	0	0	0	18	144 748	2 951 845	20
1	11	18	0,92	1	0	0,32	0,02	0	0	0	0	18	131 827	2 415 614	18
1	12	18	0,91	1	0	0,31	0,01	0	0	0	0	18	143 289	3 208 227	22
1	13	18	0,87	1	0	0,27	0,00	0	0	0	0	18	121 211	4 559 636	37
1	14	18	0,81	1	0	0,21	0,00	0	0	0	0	18	104 002	6 983 751	66
1	15	18	0,69	1	0	0,09	0,00	0	0	0	0	18	51 165	4 698 745	90
1	16	18	0,59	1	0	0,00	0,00	0	0	0	9	18	15 411	2 071 060	132
1	17	18	0,33	1	0	0,00	0,00	0	0	0	9	18	33 171	7 282 464	217
														36 528 678	626

La figure 56 présente les éléments de la diagonale dans le cas de la première simulation :

Figure 56 – Nouvelle diagonale du triangle, pour b = 1.

		Année de développement															
Survenance		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	130 534	94 872	74 381	43 125	49 798	119 241	56 186	132 365	133 150	38 841	7 117	64 720	27 769	139 801	101 458	0	
2	126 400	73 450	65 272	94 700	50 724	73 244	101 454	111 192	161 205	48 120	130 199	25 628	145 965	236 510	222 999	0	
3	142 922	53 569	75 969	134 428	26 216	114 488	137 272	120 653	98 592	127 478	94 475	44 363	9 557	15 013	17 692	75 626	
4	67 472	47 675	90 638	47 452	96 535	150 236	96 981	72 585	97 669	89 080	183 468	73 874	213 778	89 698	0		
5	61 019	48 791	27 581	54 227	128 117	98 724	132 142	67 005	34 724	74 001	198 052	76 832	73 487	116 700			
6	84 688	31 068	39 669	77 251	102 854	110 995	186 244	45 039	187 164	97 756	157 318	184 301	100 307				
7	50 255	29 262	33 885	68 338	74 462	76 805	49 374	113 187	81 258	133 414	152 248	69 230					
8	43 593	28 110	37 198	50 362	66 960	56 598	170 193	89 717	119 489	60 228	113 854						
9	39 386	25 522	22 933	69 552	48 843	151 003	119 498	185 511	154 857	89 035							
10	71 524	20 795	27 242	40 475	79 327	94 530	112 488	107 193	144 748								
11	46 621	21 575	19 239	68 157	100 487	136 249	125 275	131 827									
12	82 650	19 338	31 394	40 694	121 338	159 643	143 289										
13	56 671	38 000	24 200	52 006	124 096	121 211											
14	129 733	20 162	25 597	127 427	104 002												
15	83 020	17 894	39 411	51 165													
16	43 769	17 231	15 411														
17	33 265	33 171															

⁹⁶ Les variables explicatives sont celles définies dans le modèle final de provisionnement ayant servi au calcul du Best Estimate. Elles sont présentées au tableau 12. A noter qu'on associe à chaque coût moyen le temps opérationnel du dernier sinistre réglé de la survenance.

■ Calcul de la provision à un an

Le risque de provision a été présenté comme le résultat de deux effets :

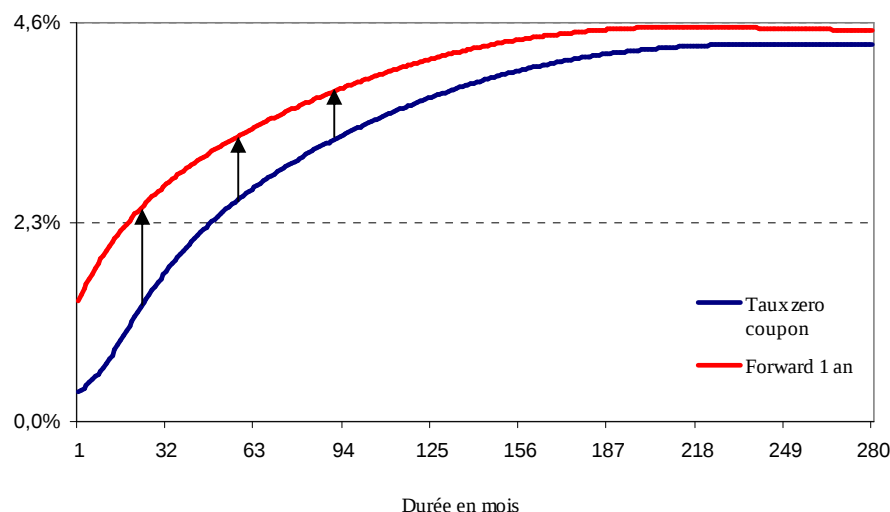
- Celui d'un évènement adverse impactant le niveau des prestations de l'année. Cette période, qualifiée de « période de choc », est simulée dans l'étape précédente.
- Celui d'une « période affectée » débutant en fin d'année, au titre de laquelle l'évènement adverse conduit à réévaluer les provisions. Ce calcul constitue l'objet de cette étape.

La mise en œuvre du calcul de la provision à un an est structurée en deux temps :

- La première étape consiste à intégrer l'information de l'année dans la prévision des prestations futures. En ce sens, on réajuste sous SAS le modèle de provision au triangle de liquidation augmenté de la nouvelle diagonale des coûts moyens.
- Dans un second temps, on calcule la provision selon la méthode décrite au chapitre I, à partir du modèle réajusté et sous les conditions de taux prévalant dans un an. Sur ce dernier point, le risque de taux n'est pas simulé⁹⁷ tandis que les conditions de taux à un an sont extrapolées à partir des trajectoires moyennes estimées en t_0 .

A ce titre, rappelons que la courbe des taux zéro coupon est vue en pleine période de crise financière⁹⁸. La figure 57 superpose la courbe des taux à un an et celle des taux zéro coupon. On observe un écart important entre les deux courbes, illustrant l'anticipation du marché d'une remontée des taux à des niveaux plus habituels.

Figure 57 – Courbe des taux forward dans un an

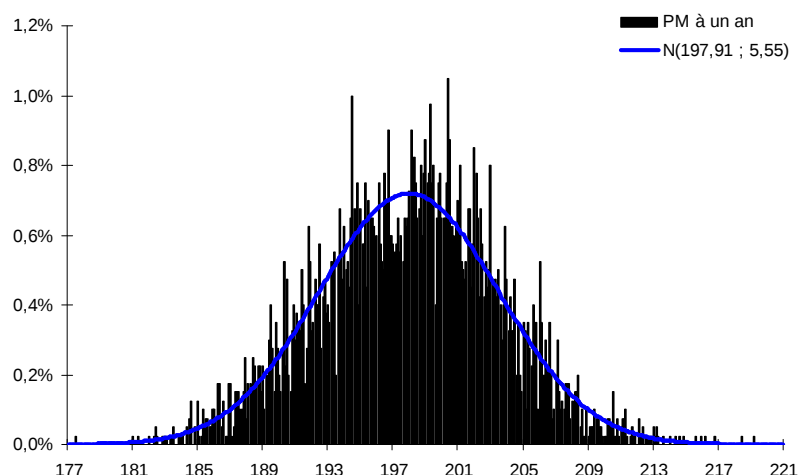


⁹⁷ Les prestations sont actualisées à la courbe des taux forward à un an déduite de la courbe des zéro coupons prévalant en t_0 . Par ailleurs, on revalorise ces prestations de l'inflation selon les indices en base t_0 calculés à partir des taux d'inflation moyens calculée en t_0 .

⁹⁸ On rappelle que la date t_0 correspond au 30/09/2009.

La distribution de la provision à un an est présentée à la figure 58. De façon générale, on observe une réduction de la volatilité de la provision par rapport à celle obtenue dans le calcul du risque ultime. Ainsi, l'écart type empirique de la provision passe de 31,2 millions à 5,5 millions de dollars. En particulier, on remarque que le moindre étalement vers la droite de la distribution de la provision, circonscrite par une loi normale, permet intuitivement d'anticiper des malis moins importants que dans le cas du risque ultime.

Figure 58 – Histogramme de la provision à un an
Millions de dollars



La moindre volatilité de la provision dans le cas du risque à un an résulte des différences de modalité de calcul des deux provisions :

Dans le cas du risque ultime, on simule la volatilité des prestations futures jusqu'à l'écoulement du risque. Dans le calcul du risque à un an, la volatilité de la provision provient de celle des prestations de l'année, au travers du recalcul des paramètres du modèle de provision. Or, la variabilité de ces paramètres est limitée par le faible poids des observations de l'année dans l'ensemble des observations constitutives du triangle vu en t_0 .

Ensuite, le calcul du risque à un an est conditionnel à l'information en t_0 . De ce fait, l'absence de rééchantillonnage du triangle qui en découle limite également la volatilité de la provision par rapport à celle obtenue dans le calcul du risque ultime.

■ Calcul du résultat de liquidation des sinistres en cours

A chaque itération, on obtient selon la procédure décrite un montant de prestations inflatées et une provision à un an. Les intérêts techniques sont calculés à partir du produit du taux zéro coupon à 6 mois et de la demi-somme des provisions d'ouverture et de fermeture.

Dans le calcul de la provision à un an, l'utilisation de la courbe des taux forward à un an introduit un changement de méthode significatif par rapport au Best Estimate, dont les flux sont actualisés à la courbe des zéro coupons. Pour mesurer cet effet dans la constitution du résultat de liquidation des sinistres en cours, nous procédons en deux temps : dans l'actualisation des prestations futures constitutives de la provision à un an, nous continuons d'utiliser les taux zéro coupon avant de recourir aux taux forward à un an.

La distribution de la perte à un an est obtenue empiriquement à partir de $B = 4000$ simulations. Elles sont présentées aux figures 59 et 60 sous la forme d'un histogramme dans le cas des deux scénarios d'actualisation. La charge en capital au titre du risque de provisionnement correspond à la valeur en risque au seuil de 99,5% de la distribution.

Figure 59 – Histogramme de la perte à un an

Actualisation à la courbe des taux zéro coupon

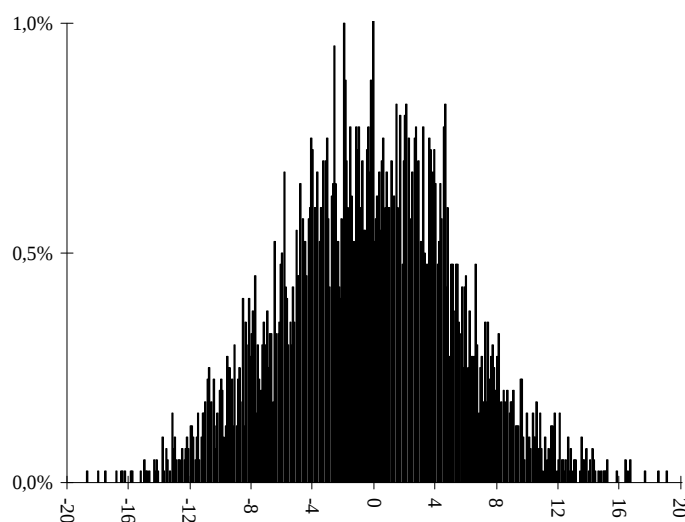
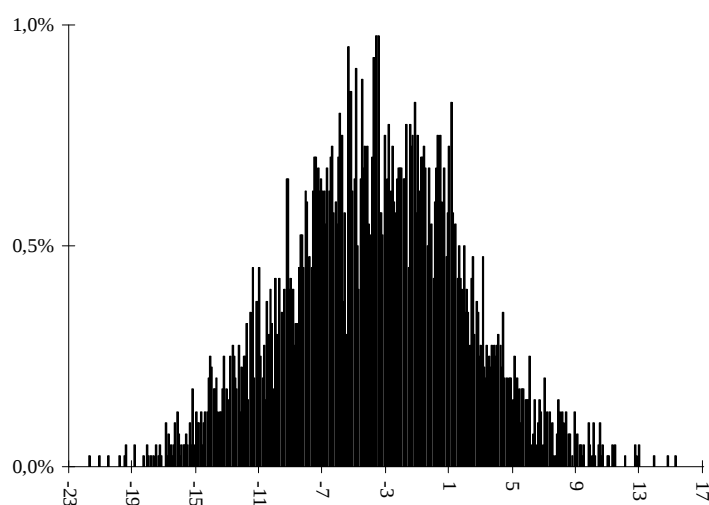


Figure 60 – Histogramme de la perte à un an

Actualisation à la courbe des taux à un an



La mise en parallèle des figures 59 et 60 illustre l'effet du passage au taux forward sur la répartition du résultat de liquidation des sinistres en cours. En effet, si dans le cas du maintien de l'actualisation aux taux zéro coupon, on remarque sur la figure 59 que les simulations aboutissent à une quasi équiprobabilité d'obtenir un boni ou un mali de liquidation sur l'année, on remarque sur la figure 60 que les scénarios à boni sont largement prépondérants⁹⁹. Par ailleurs, l'utilisation de la courbe des taux forward atténue également le montant des malis dans les scénarios à mali.

⁹⁹ 74% des simulations aboutissent à un boni sur liquidation, contre 52% dans le cas de l'actualisation aux taux zéro coupon.

Au final, on obtient une charge en capital de 14,3 millions de dollars par l'actualisation à la courbe des zéros coupons, contre 10,6 millions en utilisant les taux forward. Dans ce dernier cas, le tableau 25 permet de mesurer le poids de chaque survenance dans la formation de la charge en capital.

On remarque également par l'observation du tableau 25 que la somme des VaR des pertes par survenance est inférieure à la VaR du portefeuille. Autrement dit, la mutualisation des pertes par survenance permet, du fait de leur indépendance, de réduire la charge en capital de 13,6 à 10,6 millions de dollars. On parle alors de sous additivité de la VaR¹⁰⁰.

**Tableau 25 – Quantile des pertes à un an par survenance
au seuil de 99,5%**

i	Var de la perte au seuil 99,5%
3	292 861
4	116 297
5	180 723
6	258 348
7	256 843
8	379 937
9	556 817
10	809 591
11	1 025 561
12	1 098 905
13	1 341 864
14	1 456 258
15	1 454 498
16	1 496 474
17	2 954 847
	13 679 825

¹⁰⁰ Approximativement parlant, la VaR n'est une mesure sous additive que si la distribution des pertes est « à peu près » gaussienne. Comme le graphique 60 nous l'enseigne, nous nous trouvons dans ce cas de figure.

III.2 Le calcul par la formule standard du QIS4

III.2.1 Les principes de calcul

La formule standard du QIS4 traite de façon intégrée les risques de prime et de réserve. En négligeant l'effet de l'actualisation des flux, la perte d'un portefeuille sur une période donnée, disons l'année à venir $n+1$, dépend de deux agrégats : les primes P payées sur la période et la charge S des sinistres sur la période.

On définit la quantité S par la somme des prestations payées X et de la variation de provision ΔR sur la période. Selon la notation introduite à la première partie¹⁰¹, S s'écrit :

$$S = \underbrace{\left(X_1 + \tilde{X}_1 \right)}_X + \underbrace{\left(R^{(1)} + \tilde{R}^{(1)} - R^{(0)} \right)}_{\Delta R}$$

En début de période, seuls les montants de deux agrégats sont certains : celui $R^{(0)}$ de la provision Best Estimate d'ouverture, et celui P des primes que l'on suppose connu. Alors, la perte à un an, égale à l'excès de la charge des sinistres sur les primes, est définie en début de période par la variable aléatoire $L = S - P$.

On désigne par $V = R^{(0)} + P$ le montant de ressources disponible en début de période, et on définit le ratio combiné aléatoire Z tel que

$$\frac{L}{V} = \frac{X + R^{(1)} + \tilde{R}^{(1)}}{V} - 1 = Z - 1, \quad Z = \frac{X + R^{(1)} + \tilde{R}^{(1)}}{P + R^{(0)}}.$$

Le modèle de Solvabilité II suppose que le ratio combiné est distribué selon une loi log-normale, avec $Z \approx \text{LN}(m, \sigma)$, où $m = E(\ln Z)$ et $s^2 = \text{Var}(\ln Z)$. Selon le principe d'équivalence actuariel, on suppose que le risque de perte est nul en espérance, soit

$$E(S - P) = 0, \text{ d'où } E(Z) = 1.$$

Par définition de la loi log-normale, on a par ailleurs

$$E(Z) = \mu = e^{\frac{m+s^2}{2}} = 1, \quad \text{Var}(Z) = \text{Var}\left(\frac{L}{V}\right) = \sigma^2 = e^{2(m+s^2)} \cdot (1 - e^{-s^2}).$$

$$\text{d'où} \quad m = -\frac{s^2}{2}, \quad s^2 = \ln[1 + \sigma^2].$$

Au seuil de confiance $\alpha = 99,5\%$, le référentiel Solvabilité II définit la charge de capital C_α au titre des risques de prime et de réserve par la Value-at-risk de la perte à un an :

$$P[S - P < C_\alpha] = P\left[Z < 1 + \frac{C_\alpha}{V}\right] = 1 - \alpha.$$

¹⁰¹ Se reporter au II.2 de la première partie.

Comme la fonction logarithmique est monotone croissante, on peut réécrire cette égalité

$$P \left[\frac{\ln Z - m}{s} < \frac{\ln \left(1 + \frac{C_\alpha}{V} \right) - m}{s} \right] = 1 - \alpha$$

Comme on suppose $\ln Z \approx N(m, s)$, on a

$$\Phi \left(\left\{ \ln \left(1 + \frac{C_\alpha}{V} \right) - m \right\} \cdot s^{-1} \right) = 1 - \alpha$$

où Φ désigne la fonction de répartition de la loi normale standard.

En inversant l'équation, on trouve l'expression de la charge en capital :

$$C_\alpha = V \cdot [\exp(m + s \cdot \Phi^{-1}(1 - \alpha)) - 1] = \rho_\alpha \cdot V.$$

En remplaçant m et s par leur expression en fonction de μ et σ , on retrouve l'expression de ρ_α définie dans les spécifications du QIS4 :

$$\rho_\alpha(\sigma) = \frac{\exp(\Phi^{-1}(1 - \alpha) \cdot \sqrt{\ln(1 + \sigma^2)})}{\sqrt{1 + \sigma^2}} - 1,$$

où $\rho_\alpha(\sigma) \cong 3 \cdot \sigma$.

III.2.2 L'application numérique

Dans le calcul du capital requis, la formule standard tient compte de l'agrégation et de la diversification entre les différentes branches. Le calcul est structuré en deux étapes :

- La première étape consiste à déterminer les écarts types et les mesures de volume V pour chaque branche, pour les risques de primes et de provisions.
- La seconde étape consiste à agréger ces données pour chaque branche pour définir la mesure de volume globale l'écart type global selon une matrice de corrélation entre branches.

Dans notre cas monobranche où seul le risque de provision est étudié, le calcul consiste à paramétrer directement dans la fonction $\rho(\sigma)$ la volatilité de la branche étudiée et à appliquer ce facteur à la provision Best Estimate d'ouverture. A ce sujet, le papier consultatif n°71 (CEIOPS, 2009) préconise dans le cas de la branche RC auto corporelle un niveau de volatilité de $\sigma = 12,5\%$ (contre 12% lors du QIS4), déterminant dans le cas des données étudiées une charge en capital de

$$C_{0,995} = 237\,992\,195 \times 0,367 = 87\,460\,543.$$

III.3 Synthèse des résultats

Pour conclure ce mémoire, le tableau 26 récapitule les charges en capital obtenues à partir des différentes méthodes présentées.

Tableau 26 – Récapitulatif des charges en capital

	Risque à l'ultime	Risque à un an
Bootstrap – zéro coupon	94 539 682	14 311 370
Bootstrap – taux forward		10 679 265
Formule standard du QIS4		87 460 543

On remarque la grande différence entre le calcul à l'ultime aboutissant à une charge en capital de 94,6 millions de dollars, et celui spécifique à horizon un an conduisant à 10,6 millions de dollars de capital requis. Notons également que cette différence aurait été encore plus importante si on avait simulé l'aléa portant sur le nombre des sinistres restant à régler dans le calcul du risque ultime.

Dans la présentation du résultat à un an, nous avons abordé les principales raisons de cet écart d'un point de vue purement méthodologique. D'un point de vue économique, il est clair que les provisions constituées au titre des branches longues varient peu en valeur relative car les nouvelles informations ont un poids relatif plus faible par rapport aux branches courtes. De ce fait, certains auteurs (Ohlson et al, 2008) préconisent de retenir dans les modèles internes un horizon intermédiaire de 3 à 5 ans dans le calcul du risque de provision

Enfin, nous notons dans le cas des données étudiées et compte tenu des méthodes appliquées, que le résultat obtenu par application de la formule standard est davantage comparable au calcul spécifique du risque à l'ultime qu'à celui du risque à un an. A ce sujet, il est certain que la formule standard a été calibrée de sorte à capturer tant le risque de coût que celui de fréquence sur l'horizon d'un an. En ce sens, il n'est pas rigoureux de rapprocher le résultat de son application avec les calculs spécifiques mis en œuvre dans ce mémoire ne simulant que l'aléa sur les coûts.

CONCLUSION

Tout au long de ce mémoire, nous avons présenté un modèle de provisionnement de type coût moyen, construit dans le cadre des modèles linéaires généralisés. Nous avons montré comment en déduire deux exigences quantitatives de Solvabilité II, à savoir le calcul de la meilleure estimation de la provision pour sinistres et celui du risque de provisionnement.

La principale caractéristique du modèle est de supposer qu'il existe un lien invariant dans le temps entre le coût moyen d'un sinistre et son temps opérationnel, soit la proportion vue au moment de son règlement des sinistres finalisés de même année de survenance. Autrement dit, on suppose que les variations du rythme de règlement des sinistres n'ont pour seule conséquence que de compresser ou d'étendre la durée de règlement des sinistres, sans affecter ni l'ordre, ni le montant de leurs prestations. Dans le cas des données étudiées, nous avons observé un lien croissant entre le coût moyen des sinistres et leur ordre de règlement.

La plausibilité de cette hypothèse dite « d'invariance de l'ordre », doit évidemment être vérifiée avant d'utiliser le modèle. En effet, elle peut être remise en cause de nombreuses façons. En particulier, un changement dans la répartition des types de sinistres peut constituer une violation importante. Par exemple dans l'assurance responsabilité civile automobile, la proportion des sinistres graves peut croître du fait d'une décision de justice et ainsi augmenter à partir de cette date le coût moyen des sinistres sur des intervalles de temps opérationnel.

Notons qu'une bonne utilisation du modèle repose également sur la nature des données utilisées. Par exemple, l'information disponible dans la pratique est souvent celle utilisée dans ce mémoire, à savoir les données agrégées du triangle des prestations payées (déflatées) et des nombres de sinistres réglés. Sur ce point, on peut regretter de ne pas avoir disposé des données individuelles, dont l'utilisation aurait permis une modélisation plus fine du lien entre le coût moyen et le temps opérationnel, notamment dans les premiers intervalles de temps opérationnel denses en nombres de règlements. Par ailleurs, et malgré la solution proposée par Wright, on peut également regretter l'absence de retraitement des données des paiements partiels, qui biaise les observations de coûts moyens.

Une autre limite résulte de l'estimation purement statistique des prestations restant à payer. Dans le cas des vieux dossiers, il est clair que cette méthode d'évaluation n'est pas en cohérence avec la réalité de l'activité des compagnies qui privilégient l'évaluation dossier par dossier. Dans notre cas, la méconnaissance du portefeuille ne nous permet pas cette approche.

Enfin, remarquons que le calcul de la provision pour sinistres reste largement soumis à la plausibilité des estimations de sinistres survenus et restant à régler. Sur ce point, il est clair que l'absence de considération de l'aléa sur les nombres de sinistres restant à régler a limité l'écart déjà important entre les charges de capital calculées au titre du risque de provision dans une optique ultime et à un an. A ce sujet, le lecteur intéressé pourra se reporter au papier de Taylor (1981) proposant un modèle d'estimation des sinistres restant à régler.

En résumé, nous avons montré que le modèle présenté peut constituer une solution adaptée en cas d'instabilité de la cadence des règlements, dès lors que l'hypothèse d'invariance de l'ordre est respectée. De ce fait, son utilisation peut être profitable dans le cas des branches longues, exposées à de nombreux facteurs perturbateurs de la stabilité des cadences. Le modèle est également adapté aux branches longues, dès lors qu'il prévoit la modélisation des facteurs de queue et permet d'intégrer la modélisation de l'inflation, soit directement dans sa structure paramétrique, soit à partir d'un modèle dédié, comme c'est le cas dans ce mémoire.

BIBLIOGRAPHIE

- [1] ACAM (2007), « Rapport du groupe de travail sur le calcul du best estimate en assurance dommages ».
- [2] AISAM-ACME (2007), « Aisam-Acme study on non life long tail liabilities. Reserve risk and risk margin assessment under Solvency II ».
- [3] Ahlgrim K.C., D'Arcy S.P., Gorvett R.W. (2001), « Modelling Financial Scenarios : A Framework for the Actuarial Profession », Casualty Actuarial Society Forum.
- [4] Anderson D., Feldblum S., Modlin C., Schirmacher D. et E., Thandi N. (2005), « a Practitioner's guide to Generalised Linear Models, second edition », a Casualty Actuarial Society Study Note.
- [5] Bayazit D. (2004), « Yield Curve Estimation and Prediction with Vasicek Model », A Thesis Submitted to the Graduate School of Applied Mathematics of The Middle East Technical University, Master of Science in the Department of Financial Mathematics, Turkey.
- [6] Björkwall S., Hössjer O., Ohlsson E. (2008), « Non-parametric and parametric bootstrap techniques for arbitrary age-to-age development factor methods in stochastic claims reserving », Research report 2008:2, Mathematical Statistics, Stockholm University.
- [7] CEIOPS (2009), Consultation Paper n° 71 : « Draft CEIOPS'Advice for level 2 implementing measures on Solvency II : SCR standard formula, calibration of non-life underwriting risk », CEIOPS-CP-71-09.
- [8] Denuit M., Charpentier A. (2004), « Mathématiques de l'assurance non vie – tome I : principes fondamentaux de théorie du risque », Economica.
- [9] Dhaene J, Goovaerts M, Van Weert K (2008), « Some comments on QIS 3 ».
- [10] Dhaene J., Laeven R.J.A., Vanduffel S., Darkiewicz G., Goovaerts M.J. (2004), « Can a coherent risk measure be too subadditive ? », Research Report OP 0431, Department of Applied Economics, Katholieke Universiteit Leuven.
- [11] England P., Verrall R. (1999), Analytic and bootstrap estimates of prediction errors in claim reserving », Insurance : Mathematics and Economics, Vol. 25, p 281-293.
- [12] Groupe Consultatif Actuariel Européen (2008), « Valuation of Best Estimate under Solvency II for Non-life Insurance », Interim Report.
- [13] Hürlimann W. (2008), « On the non-life Solvency II model », 38th international Astin colloquium, Manchester, 13-16 July 2008.
- [14] Kaufmann R., Gardner A., Klett R. (2001), « Introduction to Dynamic Financial Analysis », Astin Bulletin, vol. 31, 213-249.

-
- [15] Lacoume A. (2008), « Mesure du risque de réserve sur un horizon de un an ». Mémoire I.S.F.A.
- [16] Mack T. (1993), « Distribution-free calculation of the standard error of Chain-ladder reserve estimates », *ASTIN Bulletin*, 23, pp. 214-225.
- [17] Münchener Rück (2007), « Les préjudices corporels graves en Europe : exemple de l'assurance automobile : responsabilité, indemnisation et inflation », Munich Re Group.
- [18] Ohlsson E., Lauzeningks J. (2008), « The one-year non-life risk », *Astin Colloquium presentation*, Manchester.
- [19] Planchet F., Thérond P., Jacquemin J. (2005), « Modèles Financiers en Assurance – analyses de risque dynamiques », *Assurance – Audit – Actuariat, Economica*.
- [20] Planchet F., Jacquemin J. (2003), « L'utilisation de méthodes de simulation en assurance – partie 1 : générer des nombres aléatoires », *Bulletin français d'actuariat*, Vol. 6, N°11, juin-décembre 2003, pp. 3-35.
- [21] Parlement Européen (2009), « Résolution législative du Parlement européen du 22 avril 2009 sur la proposition modifiée de directive du Parlement européen et du Conseil sur l'accès aux activités de l'assurance directe et de la réassurance et leur exercice (refonte) ».
- [22] Pinheiro P.J.R., Andrade e Silva J.M., Centeno M.D.L. (2000), « Bootstrap Methodology in Claim Reserving », *Center for Applied Maths to Forecasting & Economic Decision*.
- [23] Reid D.H. (1978), « Claim Reserves in General Insurance », *Journal of the Institute of Actuaries*, 105, 211 – 296.
- [24] SAS Institute Inc., *SAS/STAT® User's Guide*, Version 8, Cary, NC: SAS Institute Inc., 1999, Chapter 29. The GENMOD Procedure.
- [25] SAS Institute Inc., Cassell D.L., *Design Pathways*, Corvallis, « Don't Be Loopy : Re-Sampling and Simulation the SAS Way », Paper 183 – 2007.
- [26] Taylor G.C. (1981), « Speed of Finalization of claims and claims runoff analysis », *ES Knight & Co, Consulting Actuaries, Sidney Australia*, *Astin Bulletin* 12 (81-100).
- [27] Taylor G.C. (2000), « Loss Reserving, An Actuarial Perspective », *Huebner International Series on Risk, Insurance and Economic Security*, Kluwer Academic Publishers.
- [28] Taylor G.C. (2004), Mac Guire G., « Loss reserving with GLMs : a case study », *Taylor Fry Consultant Actuaries, Research paper n°113*, May 2004.
- [29] Saporta (2006), « Probabilité, Analyse des données et Statistique », Technip.

-
- [30] Wright T.S. (1992), « Stochastic Claims Reserving When Past Claims Numbers are known », PCAS 1992.
 - [31] Wilkie A.D. (1995), « More on a stochastic asset model for actuarial use », British Actuarial Journal, vol. 1, 777-964.
 - [32] Wütrich M.V., Merz M., Lysenko N. (2007), « Uncertainty in the claims development result in the chain ladder method ». Preprint, ETH, Zurich.
 - [33] Zehnwirth B. (1994), « Probabilistic development factor models with applications to loss reserve variability, prediction intervals and risk based capital », Casualty Actuarial Society Forum.

ANNEXES

A. Triangle des incréments

Tableau 27 - Triangle des incréments en dollars constants – base 31/12/1995

Survénance	J = 1	J=2	J=3	J=4	J=5	J=6	J=7	J=8	J=9	J=10	J=11	J=12	J=13	J=14	J=15
1979	3 785 472	10 341 098	8 330 688	7 848 833	2 838 485	3 577 227	1 404 642	1 720 749	1 065 201	155 365	35 583	258 878	249 919	419 402	101 458
1980	4 676 818	9 989 145	8 746 390	10 227 581	8 572 427	5 786 244	3 855 239	1 445 496	1 612 047	625 563	1 171 794	589 440	437 895	473 020	445 997
1981	5 288 102	8 088 865	12 838 745	11 829 690	7 760 019	6 182 336	4 118 148	3 016 336	1 774 652	1 784 693	2 645 293	266 179	38 228	45 040	159 232
1982	2 294 033	9 868 776	10 242 138	13 808 440	8 784 714	5 408 507	2 424 513	1 596 870	2 148 708	3 295 949	917 338	295 496	427 555	358 790	
1983	3 600 113	7 513 855	8 246 783	9 327 124	8 583 828	4 245 146	4 096 408	3 216 232	2 014 006	592 008	1 188 314	691 492	367 433		
1984	3 641 604	7 394 227	9 838 006	9 733 605	6 376 966	4 883 761	11 919 643	4 188 607	4 491 930	1 759 602	943 909	921 506			
1985	2 462 510	5 033 074	6 980 328	7 722 211	6 701 601	7 834 125	5 579 267	3 621 981	1 300 134	3 068 518	1 370 233				
1986	2 266 831	5 959 341	6 174 882	7 050 748	8 102 196	6 338 960	6 977 904	4 396 156	3 106 706	903 427					
1987	2 008 668	3 700 715	5 297 409	6 885 660	6 496 129	7 550 139	5 855 420	5 750 846	3 871 426						
1988	1 859 624	5 281 998	3 650 396	7 528 393	5 156 246	5 766 328	6 861 793	2 572 620							
1989	2 331 058	3 516 749	5 310 014	6 066 004	10 149 191	9 264 942	5 261 570								
1990	2 314 198	4 486 419	4 112 650	6 999 427	11 163 077	10 057 508									
1991	2 606 858	3 952 042	8 228 051	7 904 933	9 307 208										
1992	2 594 652	5 403 501	6 578 467	15 546 107											
1993	3 154 749	4 974 522	7 960 929												
1994	2 626 146	5 703 493													
1995	2 827 488														
	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}	f_{13}	f_{14}	
f_j	3,130	1,802	1,542	1,298	1,191	1,132	1,069	1,047	1,029	1,021	1,009	1,005	1,006	1,004	
f_j	18,355	5,864	3,255	2,110	1,625	1,365	1,205	1,128	1,077	1,047	1,025	1,016	1,010	1,004	
Dernier ex.	3,172	1,979	2,067	1,410	1,346	1,144	1,071	1,089	1,018	1,027	1,014	1,007	1,006	1,002	
2 derniers ex.	2,847	1,902	1,799	1,504	1,342	1,184	1,113	1,077	1,041	1,020	1,014	1,007	1,003	1,005	
3 derniers ex.	2,920	2,004	1,756	1,529	1,314	1,184	1,109	1,061	1,036	1,021	1,011	1,005	1,005	1,004	
4 derniers ex.	2,824	1,912	1,710	1,470	1,313	1,187	1,103	1,066	1,030	1,020	1,009	1,005	1,006	1,004	
5 derniers ex.	2,844	1,911	1,708	1,450	1,291	1,179	1,096	1,061	1,036	1,024	1,009	1,005	1,006	1,004	

B. Triangle des nombres de sinistres réglés

Tableau 28 - Triangle des nombres de sinistres réglés

[illegible]

C. Triangle des nombres de sinistres déclarés et restant à déclarer

Tableau 29 - Triangle des nombres de sinistres déclarés et restant à déclarer

Survenance	J = 1	J=2	J=3	J=4	J=5	J=6	J=7	J=8	J=9	J=10	J=11	J=12	J=13	J=14	J=15	Sinistres restant à déclarer ¹⁰²
1979	393	151	25	6	4	5	4	1	2	1	0	0	0	0	0	0
1980	517	185	29	17	11	10	8	1	0	0	1	0	0	0	0	0
1981	578	254	49	22	17	6	3	0	1	0	0	0	0	0	0	0
1982	622	206	39	16	3	7	0	1	0	0	0	0	0	0		0
1983	660	243	28	12	12	4	4	1	0	0	0	0	0			0
1984	666	234	53	10	8	4	6	1	0	0	0	0				0
1985	573	266	62	12	5	7	6	5	1	0	1					0
1986	582	281	32	27	12	13	6	2	1	0						0
1987	545	220	43	18	12	9	5	2	0							1
1988	509	266	49	22	15	4	8	0								1
1989	589	210	29	17	12	4	9									3
1990	564	196	23	12	9	5										8
1991	607	203	29	9	7											15
1992	674	169	20	12												26
1993	619	190	41													36
1994	660	161														66
1995	660															243

¹⁰² Les estimations des sinistres tardifs restant à déclarer (IBNR) ont été reprises de Taylor (2000), p 28.

La projection des sinistres tardifs a été reconstituée selon l'approche décrite lors de la présentation des données (II.1.2 de la partie II). On y précise les historiques pris en compte dans le calcul des coefficients de passage par Chain Ladder, ainsi que la valeur des coefficients obtenus à partir du modèle ajusté :

Tableau 29 bis – Extrapolation du nombre des sinistres tardifs (Taylor, 2000)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	IBNR
1979	393	544	569	575	579	584	588	589	591	592	592	592	592	592	592	0
1980	517	702	731	748	759	769	777	778	778	778	779	779	779	779	779	0
1981	578	832	881	903	920	926	929	929	930	930	930	930	930	930	930	0
1982	622	828	867	883	886	893	893	894	894	894	894	894	894	894	894,0	0
1983	660	903	931	943	955	959	963	964	964	964	964	964	964	964,0	964,1	0
1984	666	900	953	963	971	975	981	982	982	982	982	982	982,1	982,1	982,1	0
1985	573	839	901	913	918	925	931	936	937	937	938	938,1	938,2	938,2	938,3	0
1986	582	863	895	922	934	947	953	955	956	956	956,2	956,3	956,4	956,5	956,5	0
1987	545	765	808	826	838	847	852	854	854	854,3	854,5	854,6	854,7	854,7	854,8	1
1988	509	775	824	846	861	865	873	873	873,6	873,9	874,1	874,2	874,3	874,3	874,3	1
1989	589	799	828	845	857	861	870	871,4	872,0	872,3	872,5	872,6	872,7	872,7	872,7	3
1990	564	760	783	795	804	809	814,0	815,3	815,8	816,1	816,3	816,4	816,5	816,5	816,6	8
1991	607	810	839	848	855	861,9	867,2	868,6	869,2	869,5	869,7	869,8	869,9	869,9	869,9	15
1992	674	843	863	875	884,8	892,0	897,5	898,9	899,5	899,9	900,1	900,2	900,2	900,3	900,3	26
1993	619	809	850	861,3	871,0	878,0	883,4	884,8	885,4	885,8	886,0	886,1	886,1	886,2	886,2	36
1994	660	821	851,0	862,3	872,0	879,0	884,5	885,9	886,5	886,8	887,0	887,1	887,2	887,2	887,3	66
1995	660	835,7	866,3	877,8	887,7	894,8	900,3	901,8	902,4	902,7	902,9	903,0	903,1	903,2	903,2	243

Calcul des coefficients de passage

Méthode	Calcul par Chain Ladder						Calcul à partir du modèle ajusté										
Historique utilisé	3 derniers exercices				6 derniers	tout l'historique											
Facteurs	1,26626	1,03656	1,01328	1,01125	1,00806	1,00618	1,00160	1,00065	1,00038	1,00023	1,00013	1,00008	1,00005	1,00003	1,00002		

D. Triangle des nombres de sinistres en cours

Tableau 30 - Triangle des nombres de sinistres déclarés et restant en cours (hors IBNR).

[illegible]

E. Triangle des cadences de règlements

Tableau 31 – Cadence des règlements – estimation par la méthode Chain Ladder[illegible]

F. Triangle des temps opérationnels

Tableau 32 – Triangle des temps opérationnel

[illegible]

G. Triangle des coûts moyens par sinistre réglé

Tableau 33 – coûts moyens par nombre de sinistres réglés[illegible]

H. L'algorithme de Newton Raphson

L'algorithme de Newton Raphson généralise le processus de Newton dans le cas multidimensionnel.

La méthode de Newton consiste pour estimer la racine de $f(x) = 0$ à utiliser un algorithme construit à partir d'une linéarisation de la fonction au voisinage de la solution, sur la base du développement de Taylor à l'ordre 1.

Comme on recherche $x \mid f(x) = 0$ et $f(x) \cong f(x_0) + f'(x_0)(x - x_0)$, on propose la récurrence

$$f'(x_n) = \frac{f(x_n) - 0}{x_n - x_{n+1}} \Leftrightarrow x_{n+1} \cong x_n - \frac{f(x_n)}{f'(x_n)}$$

qui consiste à chaque n^e itération à trouver x_{n+1} désignant la racine de la tangente de f calculé en x_n . On peut montrer que si le premier terme de l'algorithme x_0 est choisi suffisamment proche de la racine, alors $f(x_n)$ tend vers 0 à mesure que n augmente¹⁰³.

Pour une optimisation à plusieurs dimensions, la technique de Newton Raphson conduit à la même formule. Dans notre cas, on utilise comme fonction f la dérivée de la log-vraisemblance par rapport aux coefficients ξ_i avec

$$\begin{pmatrix} \frac{\partial l(\xi^{n+1})}{\partial \xi_1} \\ \vdots \\ \frac{\partial l(\xi^{n+1})}{\partial \xi_p} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix} \cong \underbrace{\begin{pmatrix} \frac{\partial l(\xi^n)}{\partial \xi_1} \\ \vdots \\ \frac{\partial l(\xi^n)}{\partial \xi_p} \end{pmatrix}}_s + \underbrace{\begin{pmatrix} \frac{\partial^2 l(\xi^n)}{\partial \xi_1^2} & \dots & \frac{\partial^2 l(\xi^n)}{\partial \xi_1 \partial \xi_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 l(\xi^n)}{\partial \xi_p \partial \xi_1} & \dots & \frac{\partial^2 l(\xi^n)}{\partial \xi_p^2} \end{pmatrix}}_H \begin{pmatrix} \xi_1^{n+1} - \xi_1^n \\ \vdots \\ \vdots \\ \xi_p^{n+1} - \xi_p^n \end{pmatrix}$$

La formule itérative s'écrit matriciellement $\underline{\xi}^{n+1} = \underline{\xi}^n - H(\underline{\xi}^n)^{-1} \cdot \underline{s}$, où H est la matrice hessienne de la fonction l et s son gradient. Le cheminement de l'algorithme consiste alors à :

- 1- Déterminer un vecteur $\underline{\xi}^0$ proche de la solution.
- 2- Répéter : $\underline{\xi}^{n+1} = \underline{\xi}^n - H(\underline{\xi}^n)^{-1} \cdot \underline{s}$, jusqu'à $|\underline{\xi}^{n+1} - \underline{\xi}^n| < \epsilon$.

¹⁰³ L'algorithme converge quadratiquement vers une racine $\bar{x}$ dans son voisinage si $f'(\bar{x}) \neq 0$, i.e. $|x_n - \bar{x}| \leq C |x_{n-1} - \bar{x}|^2$ pour $C > 0$, n suffisamment grand et x_0 suffisamment proche de $\bar{x}$.

I. L'algorithme du score

La substitution des éléments de la matrice H par leurs espérances simplifie l'expression analytique de l'algorithme. Reprenons l'expression de $\partial l / \partial \xi_k$ introduite dans le 2^e chapitre de la partie II. Les éléments de la matrice hessienne H sont définis par¹⁰⁴

$$\begin{aligned} \frac{\partial^2 l(\underline{s}, \underline{\theta}, \phi)}{\partial \xi_s \partial \xi_k} &= \frac{\partial}{\partial \xi_s} \left[\sum_{r=1}^m \frac{(S_r - \mu_r)}{\text{Var}(S_r)} \cdot \frac{\partial \mu_r}{\partial \eta_r} \cdot z_{r,k} \right] \\ &= \sum_r (S_r - \mu_r) \cdot \frac{\partial}{\partial \xi_s} \left[\frac{1}{\text{Var}(S_r)} \cdot \frac{\partial \mu_r}{\partial \eta_r} \cdot z_{r,k} \right] + \sum_r \frac{\partial}{\partial \xi_s} (S_r - \mu_r) \cdot \left[\frac{1}{\text{Var}(S_r)} \cdot \frac{\partial \mu_r}{\partial \eta_r} \cdot z_{r,k} \right] \quad (1) \end{aligned}$$

Comme $E(S_r) = \mu_r$, l'espérance du 1^{er} terme de l'équation (1) est nulle. Par ailleurs, on a

$$\frac{\partial}{\partial \xi_s} (S_r - \mu_r) = -\frac{\partial \mu_r}{\partial \eta_r} \cdot \frac{\partial \eta_r}{\partial \xi_s} = -\frac{\partial \mu_r}{\partial \eta_r} \cdot z_{r,s}$$

Comme $\frac{\partial \mu_r}{\partial \eta_r} = \frac{1}{g'(\mu_r)}$, on en déduit :

$$E \left(\frac{\partial^2 l(\underline{s}, \underline{\theta}, \phi)}{\partial \xi_s \partial \xi_k} \right) = E \left(- \sum_r \frac{z_{r,l} \cdot z_{r,s}}{\text{Var}(S_r) \cdot (g'(\mu_r))^2} \right) = - \sum_r \frac{z_{r,l} \cdot z_{r,s}}{\text{Var}(S_r) \cdot (g'(\mu_r))^2},$$

de telle sorte que $E \left[- \frac{\partial^2 L_n(\underline{\xi}, \underline{s})}{\partial \xi \partial \xi'} \right] = I_n(\underline{\xi}) = (Z' \cdot W(\underline{\xi}) \cdot Z)$,

où $I_n(\underline{\xi})$ désigne la matrice d'information de Fisher.

En utilisant l'expression de $\partial l(\underline{\xi}) / \partial \underline{\xi}$ et celle de la matrice I_n , l'algorithme du score s'écrit

$$\begin{aligned} \xi^{t+1} &= \xi^t - \left[E \left(\frac{\partial^2 L_t(\underline{\xi}, \underline{S})}{\partial \xi \partial \xi'} \right) \right]^{-1} \cdot \frac{\partial L_t(\underline{\xi}, \underline{S})}{\partial \xi} \\ &= \xi^t + (Z' W(\xi^t)^{-1} Z)^{-1} \cdot Z' W(\xi^t)^{-1} \frac{\partial \eta^t}{\partial \underline{\mu}} (\underline{S} - \underline{\mu}^t) \\ &= (Z' W(\xi^t)^{-1} Z)^{-1} \cdot Z' W(\xi^t)^{-1} Y^t \end{aligned}$$

où $Y^t = Z \xi^t + \frac{\partial \eta^t}{\partial \underline{\mu}} (\underline{S} - \underline{\mu}^t) = \underline{\eta}^t + \frac{\partial \eta^t}{\partial \underline{\mu}} (\underline{S} - \underline{\mu}^t)$.

¹⁰⁴ On montre en II.2.2 du chapitre II de la partie II que $\frac{\partial l(\underline{\xi}, \phi | \underline{S})}{\partial \xi_k} = \sum_k \frac{\partial l_r}{\partial \xi_k} = \sum_k \frac{(S_r - \mu_r)}{\text{var}(S_r) g'(\mu_r)} \cdot z_{r,k}$.

On remarque qu'un développement au 1^{er} ordre de la fonction de lien g en μ donne

$$g(\underline{S}) \approx g(\underline{\mu}) + g'(\underline{\mu}) \cdot (\underline{S} - \underline{\mu}) = \underline{Y}.$$

En ce sens, $y(\xi)$ peut être considéré comme une approximation au 1^{er} ordre de $g(s)$ en $\mu(\xi)$. C'est pourquoi une solution permettant d'obtenir une valeur de $\underline{\xi}^0$ consiste à appliquer une procédure de moindres carrés ordinaires sur $g(s)$.

Par l'introduction de $\underline{Y} = \underline{\eta} + \frac{\partial \eta}{\partial \underline{\mu}}(\underline{S} - \underline{\mu})$ dans les équations de vraisemblance, on aboutit à

$$\underline{Z}' W(\underline{\xi})^{-1} \frac{\partial \eta}{\partial \underline{\mu}}(\underline{X} - \underline{\mu}) = \underline{Z}' W(\underline{\xi})^{-1} (\underline{Y} - \underline{Z} \underline{\xi}) = 0$$

Cette réécriture conduit à deux commentaires :

- L'algorithme des scores peut être décrit en résolvant itérativement ce système d'équations : A chaque itération, la valeur courante de $\underline{\xi}$ permet le calcul de la matrice des poids $W(\underline{\xi})$, du vecteur $\underline{Y}$ et l'obtention d'un nouveau vecteur $\underline{\xi}$ par la résolution de ce système linéarisé.
- Cette nouvelle expression des équations de vraisemblance permet une interprétation de l'algorithme de type linéaire :

A $\underline{\xi}$ fixé, on considère $\underline{Y}$ comme un nouveau vecteur de données et $W(\underline{\xi})$ comme une matrice de poids fixés. On reconnaît dans ce système les équations des moindres carrés généralisés associées au modèle

$$\underline{Y} = \underline{Z} \underline{\xi} + \underline{\varepsilon}, \text{ où } E(\underline{\varepsilon}) = 0 \text{ et } Var(\underline{\varepsilon}) = W(\underline{\xi}).$$

En effet, puisque la $r^{\text{ème}}$ composante du vecteur $\underline{Y}(\underline{\xi})$ a pour variance

$$\text{var} \left(\hat{\eta}_r + \left(x_r - \hat{\mu}_r \right) \frac{d\eta_r}{d\mu_r} \right) = g' \left(\hat{\mu}_r \right)^2 \cdot \text{var}(S_r)$$

et parce que les coûts S_r sont supposés indépendants, $W(\underline{\xi})$ correspond bien à la matrice de variance de $\underline{Y}$. Le modèle a deux particularités. D'une part, le vecteur à expliquer $\underline{Y}$ n'est pas observé : ses valeurs sont calculées à chaque nouvelle valeur de $\underline{\xi}$ à partir des S_r . D'autre part, la matrice W dépend également de $\underline{\xi}$.

A chaque itération, l'estimation $\underline{\xi}^{t+1}$ est issue de l'application des moindres carrés généralisés dans le modèle $\underline{y}(\underline{\xi}^t) = \underline{Z} \underline{\xi} + \underline{\varepsilon}^t$, ce qui revient à estimer $\underline{\xi}$ par vraisemblance en supposant les S_r normaux et indépendants.

J.Expression de la déviance

Pour aboutir à l'expression de la déviance, il s'agit de remarquer que la log-vraisemblance peut être exprimée en fonction des espérances des coûts moyens. Comme $\theta_r = b'^{-1}(\mu_r)$, on a

$$\begin{aligned} l(\underline{\theta}, \phi \mid \underline{S}) &= \sum_r \frac{S_r \theta_r - b(\theta_r)}{\phi} + c(S_r, \phi) = \sum_r \frac{S_r b'^{-1}(\mu_r) - b(b'^{-1}(\mu_r))}{\phi} + c(S_r, \phi) \\ &= l(\underline{\mu}, \phi \mid \underline{S}) \end{aligned}$$

où $l(\hat{\underline{\mu}}, \phi \mid \underline{S})$ désigne la log vraisemblance maximisée du modèle estimé avec $\hat{\mu}_r = g^{-1}\left(\underline{Z}_r' \hat{\underline{\xi}}\right)$, et $l(\underline{S}, \phi \mid \underline{S})$ celle atteinte par le modèle saturé avec $\hat{\mu}_r = S_r \forall r$.

La différence des deux log-vraisemblances définit la déviance normalisée

$$\begin{aligned} \frac{D(\underline{S}, \hat{\underline{\mu}})}{\phi} &= -2 \times \left[l(\hat{\underline{\mu}}, \phi \mid \underline{S}) - l(\underline{S}, \phi \mid \underline{S}) \right], \\ &= \frac{2}{\phi} \times \sum_r \left[S_r \cdot (\tilde{\theta}_r - \hat{\theta}_r) - b(\tilde{\theta}_r) + b(\hat{\theta}_r) \right] \end{aligned}$$

où $\tilde{\theta}_r = b'^{-1}(S_r)$ et $\hat{\theta}_r = b'^{-1}(\hat{\mu}_r)$. Autrement dit, $\tilde{\theta}_r$ est tel que $E\left[S_r \mid \hat{\theta}_r, \phi\right] = S_r$.

On peut exprimer la déviance comme la somme des carrés des résidus de déviance $re_r^{(D)}$:

$$D(\underline{S}, \hat{\underline{\mu}}) = -2\phi \ln \left[\frac{L(\hat{\underline{\mu}}, \phi \mid \underline{S})}{L(\underline{S}, \phi \mid \underline{S})} \right] = \sum_{r=1}^n d_r = \sum_{r=1}^n [re_r^{(D)}]^2$$

où $re_r^{(D)} = \text{sgn}(S_r - \hat{\mu}_r) \sqrt{d_r}$ et $d_r = 2S_r(\tilde{\theta}_r - \hat{\theta}_r) - (b(\tilde{\theta}_r) - b(\hat{\theta}_r))$.

Remarque : Dans le cas de la loi normale, la déviance correspond à la somme des carrés des résidus d'estimation des MCO. En effet, comme

$$l(\underline{\theta}, \phi \mid \underline{S}) = \sum_r \frac{1}{2\sigma^2} [S_r - \mu_r]^2 - \frac{1}{2} \cdot n \cdot \ln(2\pi\sigma^2),$$

on a

$$-2 \times \left[l(\hat{\underline{\mu}}, \phi \mid \underline{S}) - l(\underline{S}, \phi \mid \underline{S}) \right] = \frac{1}{\sigma^2} \sum_r (S_r - \hat{\mu}_r)^2 \Rightarrow D^*(\underline{S}, \hat{\underline{\mu}}) = \sum_r (S_r - \hat{\mu}_r)^2$$

K. Evolution des coûts moyens en monnaie constante

Figure 61 – Evolution dans le temps des coûts moyens
Valeurs basses de temps opérationnel

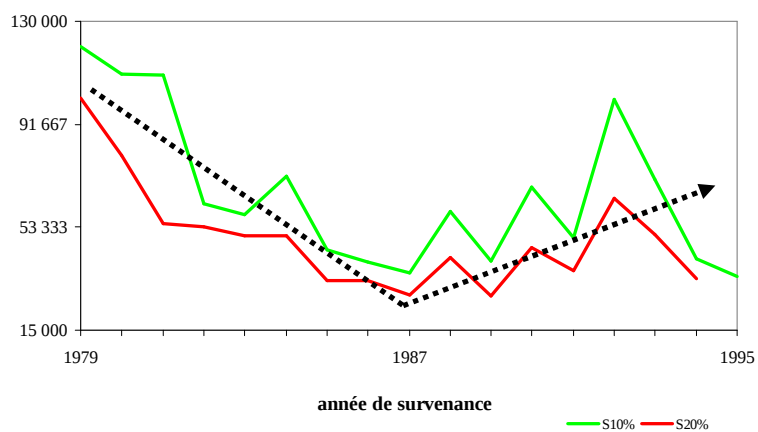


Figure 62 – Evolution dans le temps des coûts moyens
Valeurs intermédiaires de temps opérationnel

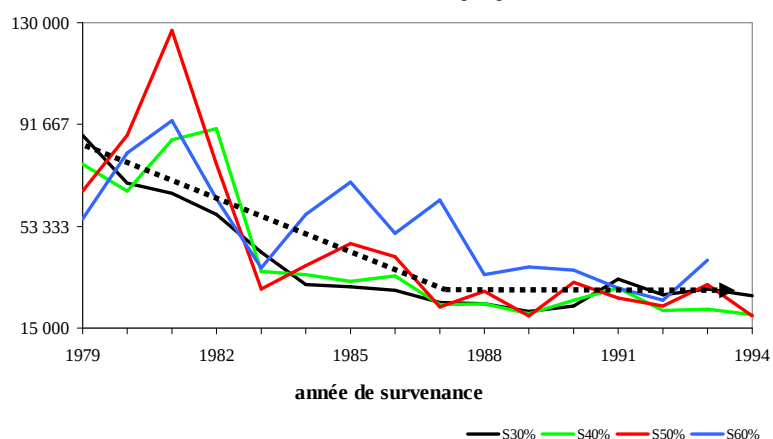
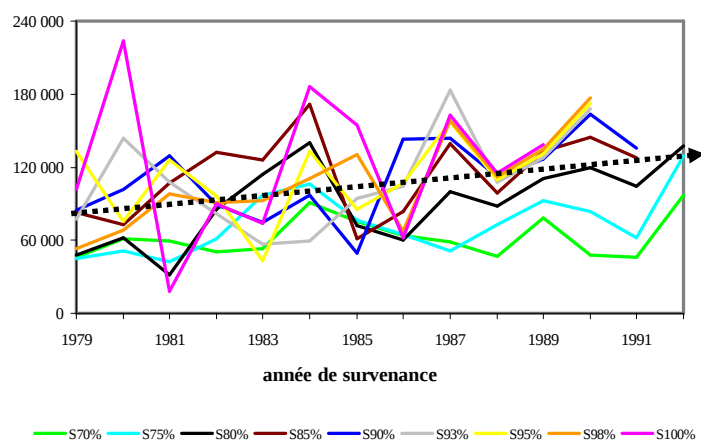


Figure 63 – Evolution dans le temps des coûts moyens
Valeurs élevées de temps opérationnel



L. Rappel élémentaire de théorie du risque

Les résultats rappelés dans cette section concernent les modèles dits de lois composées. Ces modèles probabilistes composent la loi des coûts de sinistres à partir d'une loi de comptage du nombre N de sinistres et d'une loi continue pour les coûts associés X .

Ces résultats sont utilisés dans la modélisation des paiements partiels de Wright présentée dans la partie II. On rappelle que Wright utilise la relation

$$E(X_{ij}) = E(N_{ij}) \cdot E(X_1)$$

où X_{ij} désigne la somme des prestations de montants $X_1, \dots, X_{N_{ij}}$, payées au titre des N_{ij} règlements effectués l'année j ¹⁰⁵.

En d'autres termes, on a

$$X_{ij} = \sum_{v=0}^{N_{ij}} X_v.$$

Comme N est un entier, on a

$$E(N_{ij}) = 0 \cdot P(N_{ij} = 0) + 1 \cdot P(N_{ij} = 1) + \dots + P(N_{ij} = \infty)$$

Au prix de certaines hypothèses¹⁰⁶, on peut écrire

$$\begin{aligned} E(X_{ij}) &= E_{N_{ij}}(E(X_{ij} | N_{ij})) = \sum_{k=0}^{\infty} P(N_{ij} = k) \cdot E\left(\sum_{v=1}^k X_v | N_{ij} = k\right) \\ &\stackrel{(1)}{=} \sum_{k=0}^{\infty} P(N_{ij} = k) \cdot E\left(\sum_{v=1}^k X_v\right) \\ &\stackrel{(2)}{=} \sum_{k=0}^{\infty} P(N_{ij} = k) \cdot k \cdot E(X_1) \\ &= E(N_{ij}) \cdot E(X_1) \end{aligned}$$

où $E(X_1) = E(X_k) \forall k$, en faisant l'hypothèse d'unicité de la loi des X_k .

¹⁰⁵ Notons que cette relation est également utilisée dans la modélisation des paiements partiels.

¹⁰⁶ (1) : indépendance entre les nombres et les montants de prestations, (2) : les montants de prestations sont indépendants et i.i.d.

M. Modélisation des dates de règlements futurs

La méthode de Wright (1992) présentée dans cette section est utilisée dans le calcul du Best Estimate et du risque de provision : elle permet d'associer une valeur de temps réel aux temps opérationnels des règlements futurs.

Le temps opérationnel est modélisé à partir d'une fonction de répartition de type exponentielle. Autrement dit, on suppose la relation

$$\tau = F(t) = \int_{T=0}^{T=t} c \cdot e^{-c \cdot T} dT, \quad 0 \leq \tau \leq 1,$$

où T désigne la variable aléatoire « temps de règlement », de réalisation t associée au temps opérationnel τ .

En intégrant par $T' = T - t_0$, on a

$$\begin{aligned} \tau &= \int_{T'=-t_0}^{T'=t-t_0} c \cdot e^{-c \cdot (T'+t_0)} dT' = c \cdot e^{-c \cdot t_0} \left[-\frac{e^{-c \cdot T'}}{c} \right]_{T'=-t_0}^{T'=t-t_0}, \\ &= e^{-c \cdot t_0} \cdot \left(e^{-c \cdot t_0} - e^{-c \cdot (t-t_0)} \right) = 1 - e^{-c \cdot t_0} \cdot e^{-c \cdot (t-t_0)} \end{aligned}$$

Comme $F(t_0) = \tau_0 = 1 - e^{-c \cdot t_0}$, on retrouve l'expression de Wright

$$\tau = 1 - (1 - \tau_0) \cdot e^{-c \cdot (t-t_0)}.$$

En inversant cette relation, on obtient

$$t - t_0 = -\frac{1}{c} \cdot \ln \left(\frac{1 - \tau}{1 - \tau_0} \right).$$

N. Projection des règlements de sinistres

Tableau 34 – Projections des règlements de sinistres

Année Calendair e	Année de survenance																
	1995	1994	1993	1992	1991	1990	1989	1988	1987	1986	1985	1984	1983	1982	1981	1980	1979
1996	188	114	85	54	35	23	19	15	9	5	3	3	1	0	3	0	0
1997	145	88	65	41	27	17	15	11	8	4	2	1	1	0	2	0	0
1998	112	68	50	32	21	14	12	8	5	3	2	2	1	1	2	0	0
1999	86	52	39	25	16	10	9	7	5	2	1	1	0	0	1	0	0
2000	66	40	30	19	13	8	6	5	3	2	1	1	1	0	1	0	0
2001	51	31	22	14	9	6	6	4	2	1	1	1	0	0	1	0	0
2002	39	24	18	12	8	5	4	3	2	1	1	0	0	0	0	0	0
2003	30	18	14	8	5	4	3	2	2	1	0	1	0	0	1	0	0
2004	23	14	10	7	4	3	2	2	1	1	1	0	1	0	0	0	0
2005	18	11	8	5	4	2	2	1	1	0	0	0	0	0	0	0	0
2006	14	8	6	4	2	1	1	1	1	1	0	0	0	0	0	0	0
2007	11	7	5	3	2	2	1	1	0	0	0	1	0	0	0	0	0
2008	8	4	4	2	2	1	1	1	1	0	1	0	0	0	1	0	0
2009	6	4	3	2	1	0	1	0	0	0	0	0	0	0	0	0	0
2010	5	3	2	1	1	1	0	1	0	1	0	0	0	0	0	0	0
2011	4	2	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0
2012	2	2	2	1	0	1	0	0	1	0	0	0	0	0	0	0	0
2013	3	2	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
2014	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
2015	2	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0
2016	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
2017	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
2018	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2019	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
2020	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total	818	496	368	234	153	99	84	63	41	22	13	11	5	1	12	0	0

O. Caractéristiques du processus d'inflation

Pour obtenir une formule explicite de i_t , on définit le nouveau processus X_t , où $X_t = i_t - \eta$. En introduisant X_t dans l'équation initiale définissant i_t , on voit que X_t est solution de l'équation différentielle stochastique dite de Langevin

$$dX_t = -\beta X_t \cdot dt + \sigma \cdot dW_t \quad (X_0 = x_0, \beta > 0, \sigma > 0).$$

Pour résoudre ce processus, on pose $Y_t = f(t, X_t) = X_t e^{\beta t}$. En appliquant le lemme d'îto, on a

$$dY_t = \left[\frac{\partial f}{\partial t} + \frac{1}{2} \frac{\partial^2 f}{\partial X^2} \sigma^2 \right] \cdot dt + \frac{\partial f}{\partial X} \cdot dX_t$$

Comme $\frac{\partial f}{\partial t} = \beta \cdot X_t \cdot e^{\beta t}$, $\frac{\partial f}{\partial X} = e^{\beta t}$, $\frac{\partial^2 f}{\partial X^2} = 0$, on en déduit

$$\begin{aligned} dY_t &= X_t \cdot \beta \cdot e^{\beta t} dt + e^{\beta t} \cdot dX_t \\ &= \beta \cdot X_t \cdot e^{\beta t} \cdot dt + e^{\beta t} \cdot (-\beta X_t \cdot dt + \sigma \cdot dW_t) \\ &= e^{\beta t} \cdot \sigma \cdot dW_t \end{aligned}$$

Comme $Y_t = x_0 + \int_0^t dY_s$, on a $Y_t = x_0 + \int_0^t \sigma \cdot e^{\beta s} \cdot dW_s$, soit

$$X_t = x_0 \cdot e^{-\beta t} + e^{-\beta t} \cdot \int_0^t \sigma \cdot e^{\beta s} \cdot dW_s,$$

En remplaçant X_t par $i_t - \eta$, on trouve l'expression intégrale du processus :

$$i_t = i_0 \cdot e^{-\beta t} + \eta(1 - e^{-\beta t}) + \sigma \cdot e^{-\beta t} \int_0^t e^{\beta s} dW_s.$$

Et pour $u < t$, on a

$$i_t = i_u \cdot e^{-\beta(t-u)} + \eta(1 - e^{-\beta(t-u)}) + \sigma \cdot \int_u^t e^{-\beta(t-s)} dW_s.$$

Cette dernière solution montre que les valeurs possibles de i_t sont distribuées selon une loi normale standard. Ce résultat découle des propriétés du processus de Wiener dit brownien standard défini par

$$W_t = \varepsilon \times \sqrt{t}, \text{ avec } \varepsilon \sim N(0,1) \text{ et } \text{Cov}(W_t, W_{t-1}) = 0, \text{ soit } W_t \sim N(0,t).$$

Autrement dit, le processus W_t est un processus gaussien à accroissements indépendants : la quantité $W_t - W_s$ ($0 \leq s < t$) suit une loi gaussienne d'espérance nulle et de variance $(t - s)$.

Comme dW_s est gaussien et d'espérance nulle, i_t est gaussien $\forall t$ et d'espérance

$$\mu_r(u, t) = i_u \cdot e^{-\beta(t-u)} + \eta \cdot (1 - e^{-\beta(t-u)}).$$

Ensuite, comme les accroissements dW_s sont supposés iid et de variance égale à ds , on a

$$\begin{aligned} \sigma_r^2(u, t) &= V \left[\sigma \cdot \int_u^t e^{-\beta(t-s)} dW_s \right] \\ &= \sigma^2 \cdot \left(\int_u^t e^{-2\beta(t-s)} \cdot V(dW_s) \right) \\ &= \sigma^2 \cdot \int_u^t e^{-2\beta(t-s)} ds \\ &= \frac{\sigma^2}{2\beta} \cdot (1 - e^{-2\beta(t-u)}) \end{aligned}$$

La discrétisation de i_t est alors immédiate. En désignant par δ le pas de discrétisation retenu, on retrouve l'expression du modèle continu :

$$\begin{aligned} i_{t+\Delta t} &= \mu_r(t, t + \Delta t) + \sigma_r(t, t + \Delta t) \cdot \varepsilon \\ &= i_t \cdot e^{-\beta\Delta t} + \eta \cdot (1 - e^{-\beta\Delta t}) + \sigma \sqrt{\frac{1 - e^{-2\beta\Delta t}}{2\beta}} \cdot \varepsilon \end{aligned}$$

P. Equidistribution des résidus gamma de Pearson

Il s'agit ici de montrer l'équidistribution des résidus de Pearson issus du modèle (20) dans lequel on suppose les coûts moyens S_r de lois gamma. Rappelons tout d'abord l'expression générale de la densité d'une variable gamma de loi $\Gamma(\alpha, \beta)$:

$$f_x(x) = \frac{e^{-\beta x} \cdot \beta^\alpha \cdot x^{\alpha-1}}{\Gamma(\alpha)}. \quad (1)$$

Dans le cadre des modèles GLM, la densité des coûts moyens S_r est de la forme

$$\begin{aligned} f(S_r, \theta_r, \phi) &= \exp\left(\frac{S_r \theta_r + \ln(-\theta_r)}{\phi} + \frac{1}{\phi} \ln\left(\frac{S_r}{\phi}\right) - \ln(S_r) - \ln\left(\Gamma\left(\frac{1}{\phi}\right)\right)\right) \\ &= \frac{\exp\left((S_r \theta_r)^{\frac{1}{\phi}}\right) \cdot (-\theta_r)^{\frac{1}{\phi}} \cdot \left(\frac{S_r}{\phi}\right)^{\frac{1}{\phi}} \cdot S_r^{-1}}{\Gamma\left(\frac{1}{\phi}\right)} \\ &= \frac{\exp\left(\frac{\theta_r}{\phi} S_r\right) \cdot \left(-\frac{\theta_r}{\phi}\right)^{\frac{1}{\phi}} \cdot S_r^{\frac{1}{\phi}-1}}{\Gamma\left(\frac{1}{\phi}\right)} \end{aligned} \quad (2)$$

En posant $X = S_r$, $\alpha = \frac{1}{\phi}$ et $\beta = -\frac{\theta_r}{\phi}$, on retrouve l'équivalence entre les expressions (1) et (2). Par ailleurs, comme $\theta_r = -\frac{1}{m_r}$, on en déduit par analogie que¹⁰⁷

$$S_r \approx \Gamma\left(\frac{1}{\phi}, \frac{1}{\phi \cdot m_r}\right).$$

Comme la fonction génératrice des moments d'une loi $\Gamma(\alpha, \beta)$ est $M(t) = (1 - t/\beta)^{-\alpha}$, on a

$$M_{S_r}(t) = (1 - \phi \cdot m_r \cdot t)^{-\frac{1}{\phi}}$$

On en déduit l'unicité de la fonction génératrice des moments des résidus de Pearson issus du modèle (20). En particulier, on remarque que cette fonction ne dépend pas des m_r :

$$\begin{aligned} M_{e_r(p)}(t) &= E\left[e^{\frac{t \cdot S_r - m_r}{\sqrt{\phi \cdot m_r^2}}}\right] = e^{\frac{-t}{\sqrt{\phi}}} E\left[e^{\frac{t}{\sqrt{\phi \cdot m_r}} S_r}\right] \\ &= e^{\frac{-t}{\sqrt{\phi}}} \cdot M_{S_r}\left[\frac{t}{\sqrt{\phi \cdot m_r}}\right] = e^{\frac{-t}{\sqrt{\phi}}} \cdot (1 - t \cdot \sqrt{\phi})^{-\frac{1}{\phi}} \end{aligned}$$

¹⁰⁷ En vertu des propriétés d'une loi gamma (ν, λ_r) on vérifie que $E(S_r) = \nu/\lambda_r = m_r$ et $V(S_r) = \nu/\lambda_r^2 = \phi \cdot m_r^2$.

Q. Calcul du risque ultime : programmation de la procédure de rééchantillonnage

La macro SAS reçoit en input la table des coûts moyens « INC » utilisée dans la spécification du modèle. La correspondance entre les libellés des variables du programme SAS et la notation utilisée dans le mémoire est exposée dans le tableau ci-dessous :

Variables explicatives	z_r^s	$z_{r,1}^{to}$	$z_{r,3}^{to}$	$z_{r,4}^{to}$	$z_{r,1,1}^c$	$z_{r,2,1}^c$	$z_{r,1,2}^c$	$z_{r,2,2}^c$	$z_{r,3}^c$	S_r
Label SAS	a1	b1	b2	b3	c1	c2	d1	d2	e1	S

Les dix premières lignes de la table « INC » sont présentées ci-dessous.

N_obs	N	i	j	T	a1	b1	b2	b3	c1	c2	d1	d2	e1	S
1	29	1	1	1	1	0,151	0,000	0,000	8	0	0	0	1	130534
2	109	1	2	2	1	0,000	0,000	0,000	0	0	7	0	2	94872
3	112	1	3	3	1	0,000	0,000	0,000	0	0	6	0	3	74381
4	182	1	4	4	1	0,000	0,131	0,000	0	0	0	0	4	43125
5	57	1	5	5	1	0,000	0,227	0,000	0	0	0	0	5	49798
6	30	1	6	6	1	0,000	0,278	0,000	0	0	0	0	6	119241
7	25	1	7	7	1	0,000	0,320	0,020	0	0	0	0	7	56186
8	13	1	8	8	1	0,000	0,342	0,042	0	0	0	0	8	132365
9	8	1	9	9	1	0,000	0,356	0,056	0	0	0	0	9	133150
10	4	1	10	10	1	0,000	0,363	0,063	0	0	0	0	10	38841

A chacune des nb_simu itérations, la macro rééchantillonne le triangle des résidus à partir duquel elle forme un pseudo triangle de coûts moyens. Ensuite, elle ajuste le modèle GLM-gamma sur le pseudo triangle et récupère les estimations des p paramètres de régression (de label a1, b1, b2, b3, c1, d1, d2 et e1). Les valeurs de ces paramètres, stockées dans la table de sortie « Stock_Param » sont ensuite utilisées sous tableur Excel pour simuler les PSAP conformément à la procédure décrite au chapitre II de la troisième partie.

```
%macro bootstrap(p=,m=,nb_Simu=);
ods listing;
%let ddl=%eval(&m-&p);
```

$\Rightarrow$ ddl du coefficient d'ajustement des résidus.

1. Calcul des résidus de Pearson

```
proc genmod data=inc;
  model S= a1 b1-b3 c1 d1-d2 e1 / dist=g link=log noint ;
  output out=residus resraw=res_brut reschi=res_P pred=mu;
run;
```

2. Rééchantillonnage des résidus de Pearson

2.1 Génération du triangle des résidus de Pearson ajustés

```
data T1(keep= a1 b1-b3 c1 d1-d2 e1 mu V_mu N_obs) T2(keep= res_P_aj);
  set work.residus;
  N_Obs+1-1;
  V_mu=(res_brut/res_P)**2;
  res_P_aj=res_P*%sysevalf(&m /&ddl)**0.5;
```

$\Rightarrow$ On attribue un n° aux résidus pour tirage aléatoire
 $\Rightarrow$ On calcule $V(\mu) = \mu \times \phi^{0.5}$ à partir des résidus
 $\Rightarrow$ On calcule les résidus de Pearson ajustés

```
run;
```

2.2 Duplication "nb_Simu" fois des n couples $(V(\mu), \mu)$ du triangle supérieur

⇒ Proc *surveyselect* : assigne un n° « Replicate » à chaque triangle dupliqué
 ⇒ désigne par « numrecs » le nombre d'éléments du triangle (soit n)

```
proc surveyselect data=T1 out=T1 method=srs samprate=1 rep=&nb_Simu;run;
```

2.3 Rééchantillonnage "nb_Simu" fois du triangle des résidus ajustés

```
data T2;
  do replicate=1 to &NB_Simu;
    do N_Obs=1 to numrecs;
      p=ceil(numrecs*ranuni(394747373)); ⇒ On simule le n° du résidu à tirer par tirage uniforme
      set T2 nobs=numrecs point=p;      ⇒ On sélectionne le résidu
      output;
    end;
  end;
stop;
run;
```

2.4 Génération des pseudo coûts moyens

```
data Pseudo_S(keep = replicate a1 b1-b3 c1 d1-d2 e1 S) ;
  merge T1 T2;
  by replicate N_Obs;
  S=mu+res_P_aj*V_mu**0.5; ⇒ Réplication bootstrap des coûts moyens
run;
```

2.5 Régression des "nb_Simu" pseudo triangles

```
%do cpt = 1 %to &NB_Simu;
  dm "clear log;clear output";
  proc SQL; ⇒ Extraction d'un pseudo triangle
    create table work.temp as select *
    from work.Pseudo_S where replicate =&cpt;
  quit;
```

2.5.1 Régression sur le $cpt^{ème}$ pseudo triangle de coûts moyens

```
ods output ParameterEstimates=Param;
proc genmod data=temp;
  model S= a1 b1 b2 b3 c1 d1 d2 e1 / dist=g link=log noint ;
  output out=residus resraw=res_brut reschi=res_P pred=mu;
proc print data=Param; ⇒ Récupération des paramètres de régression
run; quit;
```

2.5.2 Sortie des paramètres de régression

```
data Param(keep=estimate); set Param; run;
proc transpose data=param out=param; run;
%if &cpt=1 %then %do; ⇒ Création de la 1ère ligne de la table
  DATA Stock_Param; SET param ; RUN;
%end;%else %do; ⇒ Ajout des lignes supplémentaires
  DATA Stock_Param;
    set Stock_Param param;
  run;
%end;
%end;
%mend bootstrap;

%bootstrap(p=8,n=143,nb_Simu=10000);
```

R. Calcul du risque de provision : simulations des dates de règlements futurs

Afin d'être en cohérence avec le modèle de provision, nous continuons de supposer que les nombres de sinistres restant à régler sont connus et certains. Ensuite, nous choisissons pour chaque survenance de simuler les dates des règlements futurs à partir d'une loi exponentielle de fonction de répartition

$$P(T \leq t) = F(t) = \int_{T=0}^{T=t} c_i \cdot e^{-c_i \cdot T} dT, \quad i = 1, \dots, I.$$

Autrement dit, il s'agit pour chaque survenance i de simuler $(N_i - N_{i,0})$ réalisations t_{ij} d'une variable exponentielle T dont la loi est de paramètre c_i , et telles que $t_{ij} > t_0$. En utilisant la définition des probabilités conditionnelles, on note que ces variables ont pour f.d.r

$$G(t) = P(T < t \mid T > t_0) = \frac{P(t_0 < T < t)}{P(T > t_0)} = \frac{F(t) - F(t_0)}{1 - F(t_0)}.$$

Comme¹⁰⁸ $F(t) = 1 - e^{-c_i \cdot t}$, on retrouve immédiatement la propriété d'absence de mémoire caractéristique de la loi exponentielle avec

$$G(t) = P(T < t \mid T > t_0) = P(T < t - t_0) = 1 - e^{-c_i \cdot (t - t_0)} = F(t - t_0). \quad (1)$$

En d'autres termes, la probabilité d'un sinistre d'être réglé entre t et $t+dt$ sachant qu'il n'était pas réglé en t_0 est indépendante du temps passé et égale à

$$P(t < T < t + dt \mid T > t_0) = \frac{F(t + dt) - F(t)}{1 - F(t_0)} = 1 - e^{-c_i \cdot dt} \approx c_i \cdot dt$$

Comme G et sa fonction réciproque G^{-1} sont connues, que G est continue et strictement croissante sur $[0;1]$, nous décidons de simuler les t_{ij} par la méthode classique de l'inversion de la fonction de répartition.

Cette méthode repose sur le fait que si G est continue et strictement croissante sur $[0;1]$, la variable aléatoire $G^{-1}(U)$ a également pour fonction de répartition G ¹⁰⁹. Pour simuler un n-échantillon d'une loi de fonction de répartition G , il suffit alors de simuler n réalisations d'une v.a.r. uniforme U sur l'intervalle $[0;1]$ et d'appliquer la fonction G^{-1} à chacune de ces valeurs.

Par l'équation (1), on obtient facilement l'expression de la fonction G^{-1} :

$$t = t_0 - \frac{1}{c_i} \ln(1 - G(t)) = G^{-1}(G(t)).$$

¹⁰⁸ Se reporter à l'annexe M.

¹⁰⁹ En désignant par U une variable uniforme sur l'intervalle $[0;1]$, on a par définition de U : $P[G^{-1}(U) < z] = P[U < G(z)] = G(z)$.

S. Calcul du risque de provision : programmation de la régression à un an

A chaque itération de la procédure Bootstrap¹¹⁰, la macro SAS réajuste le modèle de provisionnement à partir du triangle de liquidation vu à un an.

■ Entrées

La macro SAS prend pour entrées :

- La table initiale des coûts moyens « INC » utilisée dans la spécification du modèle Best Estimate et correspondant au triangle de liquidation initial vu en t_0 .
- La table « Diag » de Bx15 lignes. Elle est constituée de B blocs, composés des coûts moyens (hors inflation) de l'année simulés pour chaque scénario sous Excel pour les survenances $i = 3$ à 17^{111} , et des variables explicatives associées. Chacun des B scénarios est identifié par la variable « n_Simu ».

Les coûts moyens et les variables associées de la diagonale correspondant au 1^{er} scénario sont présentés ci-dessous :

N_Simu	i	t	T_ope	a1	b1	b2	b3	c1	c2	d1	d2	e1	S
1	3	18	0,99	1	0	0,39	0,09	0	0	0	0	18	75 626
1	4	18	1,00	1	0	0,40	0,10	0	0	0	0	18	0
1	5	18	1,00	1	0	0,40	0,10	0	0	0	0	18	116 700
1	6	18	0,99	1	0	0,39	0,09	0	0	0	0	18	100 307
1	7	18	0,99	1	0	0,39	0,09	0	0	0	0	18	69 230
1	8	18	0,99	1	0	0,39	0,09	0	0	0	0	18	113 854
1	9	18	0,96	1	0	0,36	0,06	0	0	0	0	18	89 035
1	10	18	0,95	1	0	0,35	0,05	0	0	0	0	18	144 748
1	11	18	0,92	1	0	0,32	0,02	0	0	0	0	18	131 827
1	12	18	0,91	1	0	0,31	0,01	0	0	0	0	18	143 289
1	13	18	0,87	1	0	0,27	0,00	0	0	0	0	18	121 211
1	14	18	0,81	1	0	0,21	0,00	0	0	0	0	18	104 002
1	15	18	0,69	1	0	0,09	0,00	0	0	0	0	18	51 165
1	16	18	0,59	1	0	0,00	0,00	0	0	0	9	18	15 411
1	17	18	0,33	1	0	0,00	0,00	0	0	0	9	18	33 171

■ Sorties

La 1^{ère} ligne de la table de sortie « Stock_Param » est présentée ci-dessous. Elle est constituée des paramètres de régression estimés. Les valeurs des paramètres sont utilisées sous Excel pour calculer la provision à un an.

N_Simu	a1	b1	b2	b3	C1	d1	d2	e1
1	10,1313	2,4163	3,9809	-8,7095	0,1538	0,1818	-0,1023	0,0366

¹¹⁰ Se reporter à la procédure de calcul du risque à un an décrite au chapitre II de la partie III.

¹¹¹ Pour les survenances $i = 1$ et 2 , l'ensemble des sinistres sont supposés réglés (d'après les estimations de Taylor).

■ Macro SAS

A chacune des « nb_simu » itérations, l'échantillon ajusté par le modèle GLM-gamma est constitué du triangle original « INC » et d'une diagonale de coûts moyens de l'année à venir :

```
%macro bootstrap_1an(nb_Simu=);
ods listing;
%do cpt = 1 %to &NB_Simu;                                ⇒ Ajustement des « nb_simu » pseudo triangles
    dm "clear log;clear output";
    proc SQL;                                                ⇒ Sélection d'une diagonale
        create table work.temp as select *
        from work.Diag where n_Simu =&cpt;
    quit;
    data T(drop=n_simu);                                    ⇒ Ajout de la diagonale au triangle original
        set INC Temp; run;

    ods output ParameterEstimates=Param;                    ⇒ Ajustement du modèle GLM-gamma
    proc genmod data=T;
    model S= a1 b1 b2 b3 c1 d1 d2 e1 / dist=g link=log noint ;
    output out=residus resraw=res_brut reschi=res_P pred=mu;
    proc print data=Param;                                  ⇒ Sortie ODS des coefficients de régression
    run; quit;

    data Temp(keep=estimate t_ope);                        ⇒ lecture des paramètres de régression
        set Param temp; run;

    data Param(keep=estimate);
        set temp;
        if estimate ne . then estimate = estimate;
        if estimate = . then estimate=t_ope; run;

    proc transpose data=param out=param; run;

%if &cpt=1 %then %do;                                       ⇒ Stockage des paramètres de régression
    DATA Stock_Param;
        SET param ; run;
%end;%else %do;
    DATA Stock_Param;
        set Stock_Param param; run;
%end;
%end;

data Stock_Param(keep=a1 b1-b3 c1 d1 d2 e1 phi t3-t17);
    set Stock_Param;
    rename
        COL2=a1 COL3 = b1 COL4 = b2 COL5=b3
        COL6=c1 COL7 = d1 COL8=d2 COL9 = e1;
run;
%mend bootstrap_1an;

%bootstrap_1an(nb_Simu=4000);
```