

UNIVERSITÉ PARIS DAUPHINE

Département MIDO*

MASTER MIDO

MENTION MMD**

SPÉCIALITÉ ACTUARIAT

Année Universitaire : 2009-2010

Mémoire d'Actuariat présenté le 18 novembre 2010 devant l'Université Paris Dauphine
et l'Institut des Actuaire

Par : Hélène COMPAIN

Tuteur d'entreprise : Régis WEISSLINGER
Tuteur académique : Olivier WINTENBERGER

Sujet : **Analyse du risque de provisionnement non-vie dans le cadre de la réforme Solvabilité II**

Entreprise d'accueil : Autorité de Contrôle Prudentiel

NON CONFIDENTIEL

JURY

Jean-Marie NESSI
Pierre MATHOULIN
Stéphane MENART

* MIDO : Mathématiques, Informatique, Décision, Organisation

** MMD : Mathématiques, Modélisation, Décision

« Ce rapport est basé sur un stage effectué au sein de l'ACP, indépendamment de toute mission de contrôle de l'Autorité. Il n'engage que le stagiaire et en aucun cas l'Autorité et ne constitue pas une communication de l'ACP. Les méthodologies retenues ont été choisies par le stagiaire et ne sont pas à considérer comme une recommandation technique de l'ACP. »

RÉSUMÉ / ABSTRACT

Résumé:

Ce mémoire propose une analyse critique du risque de provisionnement en assurance non-vie dans le cadre de Solvabilité II. Cette étude s'appuie tout d'abord sur l'application pratique de méthodes classiques de calcul de provisions pour établir leur adéquation aux données. La nouvelle réforme incitant à l'élaboration de modèles internes, nous avons cherché à adapter ses méthodes pour le calcul de Best Estimate et de Capital de Solvabilité Requis. Nous avons donc implémenté des procédés gérant la liquidation incomplète des triangles de règlements et les recours. Enfin, nous nous sommes concentrés sur l'évaluation du risque à un an et l'impact de la prise en compte de la dépendance entre branches d'activités par l'utilisation de copules.

Mots clés : provisionnement, modèles stochastiques, bootstrap, dépendance, copules, Solvabilité II.

Abstract:

The aim of this report is to analyze the non-life insurance reserving risk within the framework of Solvency II. First, this study is based on a practical application of classical methods of reserves in order to establish their adequacy to our data. The new reform encourages the development of internal models, we tried to adapt these methods to calculate Best Estimate and Solvency Capital requirement. So we built process managing uncompleted development of claim settlements and recurses. Finally, we tried to evaluate the one-year reserve risk and the impact of dependence between lines of business using copulas.

Key words: reserving, stochastic models, bootstrap, dependence, copulas, Solvency II.

REMERCIEMENTS

Je tiens tout d'abord à remercier mon maître de stage, Régis WEISSLINGER, pour m'avoir proposé ce sujet et fait bénéficier de son expérience et de ses conseils.

Je remercie également l'ensemble de la cellule des modèles internes et collègues de l'ACP pour leur accueil et leur soutien.

Enfin, je remercie ma famille et mes amis qui m'ont soutenue jusqu'à la rédaction de ce mémoire.

SOMMAIRE

RÉSUMÉ / ABSTRACT	3
REMERCIEMENTS	4
INTRODUCTION	7
PARTIE I : GÉNÉRALITÉS ET CONTEXTE.....	8
I. REGLEMENTATION ET SOLVABILITE.....	8
I.1. Assurance et solvabilité	8
I.2. Solvabilité I.....	8
I.3. Solvabilité II	9
II. CONTEXTE DU MEMOIRE	12
II.1. Le SCR non-vie.....	12
II.2. La formule standard (SCR Non-Vie) provisions	13
II.3. Les modèles internes	14
II.4. Présentation des données	15
PARTIE II : LES METHODES DE PROVISIONNEMENT	17
I. LA METHODE CHAIN LADDER	17
II. LE MODELE DE MACK.....	21
III. LES MODELES FACTORIELS STOCHASTIQUES	22
III.1. La régression Log Normale.....	23
III.2. La modélisation GLM	27
IV. LE BOOTSTRAP.....	30
V. EXTENSIONS ET LIMITES DES MODELES.....	33
V.1. Gestion de la liquidation incomplète.....	33
V.2. Les recours	36
PARTIE III : EVALUATION DU RISQUE A UN AN	38
I. MODELE DE MERZ-WÜTHRICH	38
I.1. Méthodologie.....	39
I.2. Utilisation dans le calcul du SCR non-vie.....	40
II. ADAPTATION DU BOOTSTRAP POUR L'EVALUATION DU RISQUE A UN AN	42
PARTIE IV : PRISE EN COMPTE DE LA DÉPENDANCE ENTRE	
BRANCHE.....	44
I. LA DEPENDANCE	44
I.1. Notion de dépendance et corrélation.....	44
I.2. Mesures de dépendance.....	45
I.3. Types de dépendance	46

II. LA MODELISATION DE LA DEPENDANCE PAR L'UTILISATION DES COPULES	47
II.1. Introduction à la théorie des copules	47
II.2. Copules archimédiennes	48
II.3. Estimation du paramètre et choix de la copule.....	52
II.4. Détermination des Best Estimates corrélés	52
III. APPLICATION AUX DONNEES.....	53
III.1. Recherche de dépendance entre branches	53
III.2. Estimation du paramètre et choix de la copule	54
III.3. Application au calcul du SCR agrégé des branches Dom.Auto et Dom.Biens	54
III.4. Application au calcul du SCR agrégé des branches Dom.Auto et RC Auto	56
CONCLUSION.....	58
ANNEXES	59
BIBLIOGRAPHIE	86

INTRODUCTION

Les provisions pour sinistres à payer (PSAP) correspondent à la valeur estimative des dépenses nécessaires au règlement de l'ensemble des sinistres survenus non payés. Elles représentent la partie la plus importante du passif d'une entreprise d'assurance non-vie. Pour faire face à ses engagements, l'assureur IARD doit donc être en mesure d'évaluer correctement ses provisions mais aussi l'incertitude autour de ce montant.

De nombreuses méthodes, déterministes ou stochastiques, permettent d'estimer ces valeurs. Généralement, elles proposent une vision à long terme, c'est-à-dire l'étude du développement des sinistres jusqu'aux sinistres ultimes. Aujourd'hui, la nouvelle réglementation Solvabilité II impose également une vision à un an.

La compagnie d'assurance doit alors répondre à des exigences de fonds propres requis pour absorber ses pertes dans le cas de scénarios des plus défavorables. Pour calculer ces montants de capitaux, deux possibilités lui sont offertes : un calcul simplifié par une formule standard ou un calcul incorporant le profil de l'entreprise par un modèle interne.

Ce mémoire propose une analyse du risque de provisionnement non-vie dans le cadre de la réforme Solvabilité II. Nous avons donc tout d'abord introduit les principes généraux de la réforme et situé le contexte dans lequel se place le mémoire à savoir l'évaluation du capital de solvabilité requis (SCR).

Les méthodes classiques de calcul de provisions restant incontournables, nous avons effectué dans une seconde partie une application pratique de ces méthodes pour établir leur adéquation par branche d'activité et évaluer les provisions à l'ultime. Ces montants ont été retraités dans le but d'obtenir le Best Estimate Solvabilité II.

Nous nous sommes ensuite focalisés sur l'évaluation du risque à un an. Dans ce but, nous avons adapté les méthodes retenues précédemment et implémenté une procédure utilisant des simulations Monte Carlo. Nous avons ainsi pu comparer nos résultats avec ceux donnés par la formule standard et le modèle de Merz-Wüthrich, exposé auparavant.

Enfin, dans une dernière partie, nous avons cherché à quantifier l'impact de la prise en compte de la dépendance entre branches d'activités dans le calcul du SCR. Les copules étant de plus en plus utilisées pour modéliser la dépendance en assurance non-vie, nous avons choisi cet outil pour représenter l'éventuel lien entre branche.

L'ensemble des résultats présentés dans ce mémoire ont été obtenus à l'aide d'outils développés sous le logiciel R.

PARTIE I : GÉNÉRALITÉS ET CONTEXTE

I. Réglementation et solvabilité

I.1. Assurance et solvabilité

Un contrat d'assurance est un contrat par lequel l'assureur s'engage, en contrepartie du paiement d'une prime ou d'une cotisation par l'assuré, à verser une prestation au profit d'un bénéficiaire en cas de réalisation d'un risque ou sinistre.

L'assureur se doit donc d'être solvable, c'est-à-dire, être capable de faire face aux engagements vis-à-vis des bénéficiaires. Pour protéger les assurés, il doit alors tarifer et provisionner prudemment par mutualisation des risques. Cependant, il existe des risques non mutualisables (erreur de modèles, erreur de paramètres, nombre limité de contrats ...) ce qui rend la solvabilité non évidente. L'assureur a donc besoin d'une marge de solvabilité sous forme de fonds propres, en plus de ses provisions.

I.2. Solvabilité I

Chaque pays a ses propres textes concernant ces exigences financières, même si des Directives cadres existent. Néanmoins, pour que les entreprises puissent opérer dans un autre État membre sans contraintes supplémentaires et pour que les assurés soient toujours bien protégés, une harmonisation au niveau européen est nécessaire. Une première tentative d'harmonisation a été effectuée dans les années 1970 à travers deux directives en assurance non-vie (1973) et en assurance vie (1979). Elle a été suivie d'une mise à jour en 2002, la directive Solvabilité I.

Solvabilité I correspond à trois règles principales :

- Les provisions techniques doivent être évaluées correctement et avoir un montant suffisant.
- En représentation des engagements réglementés, la compagnie d'assurance doit détenir des actifs sûrs, liquides et rentables.
- La valeur des actifs doit toujours être supérieure aux engagements.

Cette directive a pour avantage d'utiliser un calcul d'exigence de marge de solvabilité (EMS) peu coûteux et facilement comparable au niveau national.

Toutefois, certaines critiques ont été faites à l'encontre de Solvabilité I :

- C'est une méthode trop simpliste ayant une approche mécanique du risque.
- Elle pénalise la prudence de provisionnement et de tarification.
- Elle ne tient pas compte du profil de risques de la compagnie.
- L'harmonisation est limitée au niveau international.

Afin de créer un véritable marché unique au niveau européen, il est nécessaire de moderniser et d'harmoniser les règles prudentielles et d'homogénéiser les pratiques des autorités de contrôle à travers une nouvelle directive : Solvabilité II.

I.3. Solvabilité II

Les principaux objectifs de Solvabilité II sont de :

- Renforcer la protection des assurés.
- Renforcer l'intégration du marché européen de l'assurance ;
- Favoriser la compétitivité des assureurs européens.
- « Mieux légiférer »

Pour parvenir à ces objectifs, la directive se base sur des principes forts :

- Trouver un équilibre entre protection des assurés et coût du capital pour les assureurs. L'approche est fondée sur le profil de risque particulier de l'entreprise.
- Promouvoir une bonne gestion des risques et responsabiliser les assureurs en leur donnant plus de libertés mais avec un dialogue prudentiel renforcé en contrepartie.
- Harmoniser les normes et les pratiques prudentielles en Europe et entre secteurs.

Le cadre Solvabilité II est basé sur l'approche Lamfalussy qui est une approche de principes composée de trois niveaux. Au niveau 1 sont adoptés les grands principes et objectifs, les droits et devoirs. Au niveau 2 ce sont les mesures techniques d'exécution, régulièrement mises à jour par la commission européenne. Enfin au niveau 3 sont adoptées les recommandations pratiques par le Committee of European Insurance and Occupational Pensions Supervisors (CEIOPS). Un niveau 4 est prévu pour le renforcement des procédures de suivi.

L'un des principes forts de la directive est le principe de proportionnalité : les exigences énoncées doivent être appliquées de manière proportionnée en fonction de la nature, l'ampleur et la complexité des risques inhérents à l'activité d'une entreprise d'assurance ou de réassurance.

Le régime se veut également plus libéral avec :

- Une vision plus économique du bilan (fair value / juste valeur).
- Une vision globale de la situation financière de l'entreprise.

- La prise en compte de tous les aspects de la vie de l'entreprise (financiers, opérationnels, administratifs...)
- Une reconnaissance des outils de réductions et transferts de risques.
- « Le contenu prévaut sur la forme » : on regarde en priorité le risque.

L'architecture de Solvabilité II se décompose en trois piliers :

- Pilier 1 : exigences quantitatives applicables aux entreprises d'assurance et de réassurance.
- Pilier 2 : exigences qualitatives et surveillance prudentielle
- Pilier 3 : discipline de marché et communication externe

➤ **Détails sur le Pilier 1**

- **Valorisation des actifs et des passifs :**

« Les entreprises d'assurance et de réassurance valorisent leurs actifs et leurs passifs comme suit:

a) les actifs sont valorisés au montant pour lequel ils pourraient être échangés dans le cadre d'une transaction conclue, dans des conditions de concurrence normales, entre des parties informées et consentantes;

b) les passifs sont valorisés au montant pour lequel ils pourraient être transférés ou réglés dans le cadre d'une transaction conclue, dans des conditions de concurrence normales, entre des parties informées et consentantes.

Lors de la valorisation des passifs au titre du point b), aucun ajustement visant à tenir compte de la qualité de crédit propre à l'entreprise d'assurance ou de réassurance n'est effectué. »

- **Provisions techniques :**

La valeur des provisions techniques correspond au montant actuel que les entreprises d'assurance et de réassurance devraient payer si elles transféraient sur le champ leurs engagements à une autre entreprise d'assurance ou de réassurance. Les provisions techniques doivent être calculées de manière prudente, fiable, objective et en cohérence avec le marché.

La valeur des provisions techniques est égale à la somme de la meilleure estimation (Best Estimate) et de la marge pour risque.

Le best estimate (BE): *« La meilleure estimation correspond à la moyenne pondérée par leur probabilité des flux de trésorerie futurs, compte tenu de la valeur temporelle de l'argent (valeur actuelle attendue des flux de trésorerie futurs), estimée sur la base de la courbe des taux sans risque pertinents.*

Le calcul de la meilleure estimation est fondé sur des informations actualisées et crédibles et des hypothèses réalistes et il fait appel à des méthodes actuarielles et statistiques adéquates, applicables et pertinentes.

La projection en matière de flux de trésorerie utilisée dans le calcul de la meilleure estimation tient compte de toutes les entrées et sorties de trésorerie nécessaires pour faire face aux engagements d'assurance et de réassurance pendant toute la durée de ceux-ci. »

La marge pour risque : « les entreprises d'assurance et de réassurance calculent la marge de risque en déterminant le coût que représente la mobilisation d'un montant de fonds propres éligibles égal au capital de solvabilité requis nécessaire pour faire face aux engagements d'assurance et de réassurance pendant toute la durée de ceux-ci.

Le taux utilisé pour déterminer le coût que représente la mobilisation de ce montant de fonds propres éligibles (taux du coût du capital) est le même pour toutes les entreprises d'assurance et de réassurance et il est révisé périodiquement. Il est actuellement égal à 6%. Le taux du coût du capital utilisé est égal au taux supplémentaire, s'ajoutant au taux d'intérêt sans risque pertinent, que supporterait une entreprise d'assurance ou de réassurance détenant un montant de fonds propres éligibles, égal au capital de solvabilité requis qui est nécessaire pour faire face aux engagements d'assurance et de réassurance pendant toute la durée de ceux-ci. »

Les éléments suivants sont également à prendre en compte dans le calcul des provisions techniques :

- L'inflation
- Les participations aux bénéfices futures
- Les frais
- Les options et garanties du contrat

- **Les fonds propres (FP) :**

Les fonds propres sont la somme :

- des fonds propres de base : excédent des actifs sur les passifs, passifs subordonnés.
- des fonds propres auxiliaires : tout engagement de financement/ressource, juridiquement contraignant, qu'un assureur peut appeler pour absorber des pertes (ex : lettres de crédit, rappels de cotisations, capital souscrit non appelé)

Les fonds propres doivent être compris entre deux exigences de capital, le SCR et le MCR.

- **Le SCR :**

Le SCR est le Capital de Solvabilité Requis (Solvency Capital Requirement). Il est égal à la Value at Risk à 99,5% à un an des fonds propres de base, soit la perte maximale possible à horizon un an avec une probabilité de 99,5%.

Le calcul du SCR peut se faire par une formule standard ou par un modèle interne (calculs des FP propres aux sociétés) complet ou partiel.

Il se décompose en plusieurs modules :

- Risque de marché
- Risque de contrepartie
- Risque de souscription en santé
- Risque de souscription en vie
- Risque de souscription en non-vie

- **Le MCR :**

Le MCR est le Capital Minimum Requis (Minimum Capital Requirement). C'est le plancher à ne pas dépasser afin de respecter les engagements envers les assurés, même en situation délicate. Il est réputé être égal à la Value at Risk à 85% à un an des fonds propres de base et doit être compris entre 25% du SCR et 45% du SCR.

- **Les investissements :**

La politique de placement doit être prudente et adaptée à la structure de passif. Les actifs en représentation doivent être de qualité, rentables, liquides.

II. Contexte du mémoire

II.1. Le SCR non-vie

Nous allons nous intéresser principalement au calcul du SCR non-vie. Ce module est divisé en trois sous-modules :

- le risque catastrophe naturelle
- le risque de rachat
- le risque de primes et de provisions

Seul le risque de provisions sera étudié, c'est-à-dire le risque d'avoir des provisions de sinistres insuffisantes. Il peut provenir de trois sources d'incertitudes :

- L'incertitude liée à la valeur des engagements (le montant des sinistres peut dévier du coût attendu)
- L'incertitude concernant le développement des sinistres (qui peut dévier de la cadence attendue)
- L'incertitude liée au nombre de sinistres (qui peut dévier de la fréquence attendue)

L'objectif de la réforme Solvabilité II est d'éviter la ruine à un an des entreprises d'assurance dans 99.5% des cas. Nous allons donc chercher à estimer la volatilité des provisions à horizon un an.

Nous avons vu précédemment que les provisions techniques étaient la somme d'un Best estimate et d'une marge pour risque. Nous supposons par la suite que la marge pour risque est nulle et nous nous concentrerons donc sur le calcul de Best Estimate (BE).

Nous partons d'un bilan à t=0 de la forme suivante :

ACTIF	PASSIF
A ₀	FP ₀ = A ₀ - BE ₀
	BE ₀

Nous supposons que l'actif est investi au taux sans risque r et que les paiements surviennent en milieu d'année.

À t=1, le bilan sera de la forme :

ACTIF	PASSIF
A ₁ = A ₀ (1+r) - P ₁ (1+r) ^{1/2}	FP ₁ = A ₁ - BE ₁
	BE ₁

Les pertes d'une année sur l'autre sont représentées par $FP_0 - \frac{FP_1}{1+r}$, soit :

$$Pertes = (A_0 - BE_0) - \frac{(A_1 - BE_1)}{1+r} \Leftrightarrow Pertes = \frac{BE_1}{1+r} + \frac{P_1}{\sqrt{1+r}} - BE_0$$

Le calcul du (SCR Non-Vie)_{provisions} revient donc à évaluer la $VaR_{99,5\%}^1 \left(\frac{BE_1}{1+r} + \frac{P_1}{\sqrt{1+r}} - BE_0 \right)$.

II.2. La formule standard (SCR Non-Vie) provisions

➤ **Pour une branche d'activité (Lob):**

$SCRNL_R^{lob} = \rho(\sigma) * V_{lob}$
--

Où V_{lob} = best estimate des provisions pour sinistres à payer pour un Lob
 σ = écart-type pour le risque de provisions (spécifiques à chaque branche)

$$\rho = \text{une fonction de } \sigma \text{ définie par } \rho(\sigma) = \frac{\exp(\text{var}_{99,5\%}(\mathcal{N}(0,1)) * \sqrt{\log(\sigma^2+1)})}{\sqrt{\sigma^2+1}} - 1$$

Tableau des écart-types pour les branches d'activité étudiées:

Branches d'activité	Erreur standard pour le risque de provisions (net de réassurance)
Dommmages Auto	10%
Dommmages aux biens	11%
RC Auto	9.5%
RC Divers	11%

➤ **Pour l'ensemble des branches :**

$$SCRNL_R = \rho(\sigma) * V$$

Où $V = \sum_{lob} V_{lob}$

$$\sigma = \sqrt{\frac{1}{V^2} \sum_{i,j} CorrLob_{i,j} \sigma_i \sigma_j V_i V_j}$$

$CorrLob_{i,j}$ = la matrice de corrélation entre Lob

Matrice de corrélation entre Lob :

Corr lob	Dom. Auto	Dom. biens	RC auto	RC Divers
Dom. Auto	1			
Dom. biens	0.25	1		
RC auto	0.5	0.25	1	
RC Divers	0.25	0.25	0.5	1

II.3. Les modèles internes

La directive Solvabilité II tient compte du profil de risque des entreprises dans le calcul des exigences de capital. Toutefois, la formule standard n'en donne qu'une vision imparfaite. Les entreprises ont alors la possibilité d'utiliser des modèles mathématiques s'intégrant dans leur propre gestion des risques, les modèles internes. Ces modèles sont néanmoins soumis à l'accord préalable de l'autorité de contrôle.

II.4. Présentation des données

Par souci de confidentialité, les données utilisées ont été intégralement modifiées de sorte que nul ne puisse retrouver les données initiales. Ces données peuvent donc être qualifiées de fictives.

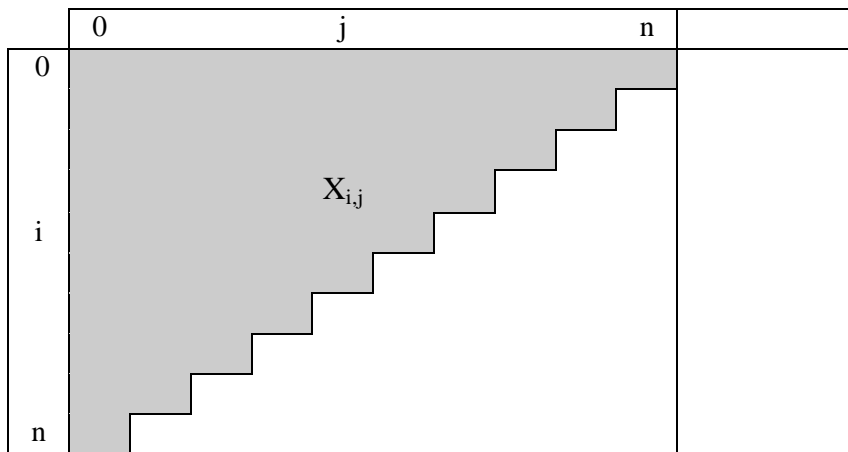
Notre étude se base sur des triangles de règlements d'historiques comptables, issus de quatre branches d'activité, nets de réassurance et en run-off. Ils ont été retraités de l'inflation de manière à manipuler des sinistres en euros constants.

La périodicité des données considérées est annuelle.

Nous utiliserons les notations suivantes :

- i l'année de survenance des sinistres
- j l'année de développement
- $(X_{i,j})$ les paiements incrémentaux
- $(C_{i,j})$ les paiements cumulés : $C_{i,j} = \sum_{k=1}^j X_{i,k}$
- R_i les provisions par année de survenance : $R_i = \sum_{j=n-i+1}^n X_{i,j}$, $i = 1 \dots n$
- R le montant total des provisions pour une branche : $R = \sum_{i=1}^n R_i$

Triangle de liquidation :



La partie grisée représente les données historiques. Nous cherchons à estimer les paiements futurs concernant les sinistres survenus avant l'année n , c'est-à-dire la partie du tableau non grisée. Il existe de nombreuses méthodes permettant d'estimer le triangle inférieur. Nous avons retenus les méthodes classiques que sont la méthode Chain-Ladder et sa version stochastique, le modèle de Mack ; les modèles factoriels stochastiques - la régression Log Normale et les modèles linéaires généralisés (GLM) - et la technique du bootstrap.

Ces méthodes ne tiennent compte ni de l'actualisation, ni de l'inflation future. Notre étude ayant pour but les calculs de Best Estimate et de SCR, nous devons prendre en compte ces paramètres. Nous avons utilisé des zéro-coupons pour l'actualisation et supposé une inflation future de 2%.

Pour estimer la partie rectangulaire droite du tableau (nommée queue du triangle), nous avons mis en œuvre deux méthodes basées sur la méthode Chain-Ladder et les modèles GLM.

Dans la partie suivante, nous présentons ces méthodes et les appliquons aux triangles de règlements (donnés en annexe) en vue de déterminer laquelle ou lesquelles sont les plus appropriées aux données. Nous procédons ensuite aux calculs de Best Estimates.

PARTIE II : LES METHODES DE PROVISIONNEMENT

I. La méthode Chain Ladder

La méthode Chain Ladder est une méthode déterministe fréquemment utilisée car facile à mettre en œuvre. Elle s'applique à des triangles de paiements cumulés ou des triangles de charges.

Soit le facteur de développement individuel $f_{i,j} = \frac{C_{i,j+1}}{C_{i,j}}$ pour $i=1 \dots n, j=1 \dots n$.

La méthode de Chain Ladder repose sur l'hypothèse forte suivante :

Pour j allant de 1 à n , les facteurs de développement $f_{i,j}$ sont indépendants de l'année de survenance i .

Nous allons alors considérer des coefficients de passage, d'une année à l'autre, commun pour les années de survenance, et dont l'estimation est donnée par :

$$\hat{f}_j = \frac{\sum_{i=0}^{n-j+1} C_{i,j+1}}{\sum_{i=0}^{n-j+1} C_{i,j}}, j = 0, \dots, n$$

Grâce à ces facteurs, nous pouvons estimer :

- les charges ultimes par exercice de survenance $\hat{C}_{i,n} = C_{i,n-i} * \prod_{j=n-i}^{n-1} \hat{f}_j$
- les provisions par exercice de survenance $\hat{R}_i = \hat{C}_{i,n} - \hat{C}_{i,n-i}$
- les provisions totales $\hat{R} = \sum_{i=1}^n \hat{R}_i$

Application au triangle Dommages Auto :

	dvp 0	dvp 1	dvp 2	dvp 3	dvp 4	dvp 5	dvp 6	dvp 7	dvp 8	dvp 9	\hat{R}_i
surv 0	89037	103299	104129	104587	104898	105219	105266	105290	105300	105302	0
surv 1	87077	102891	103786	104308	104657	105047	105147	105176	105188	105190	2
surv 2	83689	107465	108355	108557	108571	108826	108884	108914	108925	108927	13
surv 3	82186	101359	102323	102685	103050	103055	103067	103094	103105	103107	40
surv 4	85840	108156	108909	109294	109721	109867	109923	109952	109964	109966	99
surv 5	91965	109850	110996	111529	111777	112012	112070	112099	112111	112113	336
surv 6	84282	99912	101132	101768	102040	102255	102307	102334	102345	102347	579
surv 7	78270	94045	95583	95983	96240	96443	96492	96517	96527	96529	946
surv 8	71708	87785	88659	89031	89269	89457	89503	89526	89535	89537	1752
surv 9	76604	92930	93856	94249	94501	94700	94748	94773	94783	94785	18181
Total											21948

Coefficients de passage :

j	0	1	2	3	4	5	6	7	8
\hat{f}_j	1,2131	1,01	1,0042	1,0027	1,0021	1,0005	1,0003	1,0001	1,000

Nous rappelons qu'il existe une correspondance entre les coefficients de passage et les cadences de règlements des sinistres : $cadence_j = \frac{1}{f_j * \dots * f_{n-1}}$.

Ainsi, nous constatons que pour le triangle de la branche Dommages Auto, 80% des sinistres sont réglés dès la première année et 99% au bout de la troisième.

La branche Dommages aux biens a également une liquidation courte, 95% des paiements étant effectués dès la deuxième année.

En revanche, comme nous pouvions nous y attendre, les deux branches de Responsabilité Civile ont une liquidation beaucoup plus longue : il faut attendre la cinquième année pour atteindre 95% des paiements en RC Auto et la huitième pour la RC Divers.

Tableau des provisions pour l'ensemble des branches :

i	Dom. Auto	Dom. Biens	RC Auto	RC Divers
0	0	0	0	0
1	2	5	784	1983
2	13	14	1140	2850
3	40	67	1642	6291
4	99	182	2611	6986
5	336	317	4316	9787
6	579	620	5481	12201
7	946	962	8536	15169
8	1752	1711	13428	19586
9	18181	14558	31057	25259
Total	21948	18436	68995	100112

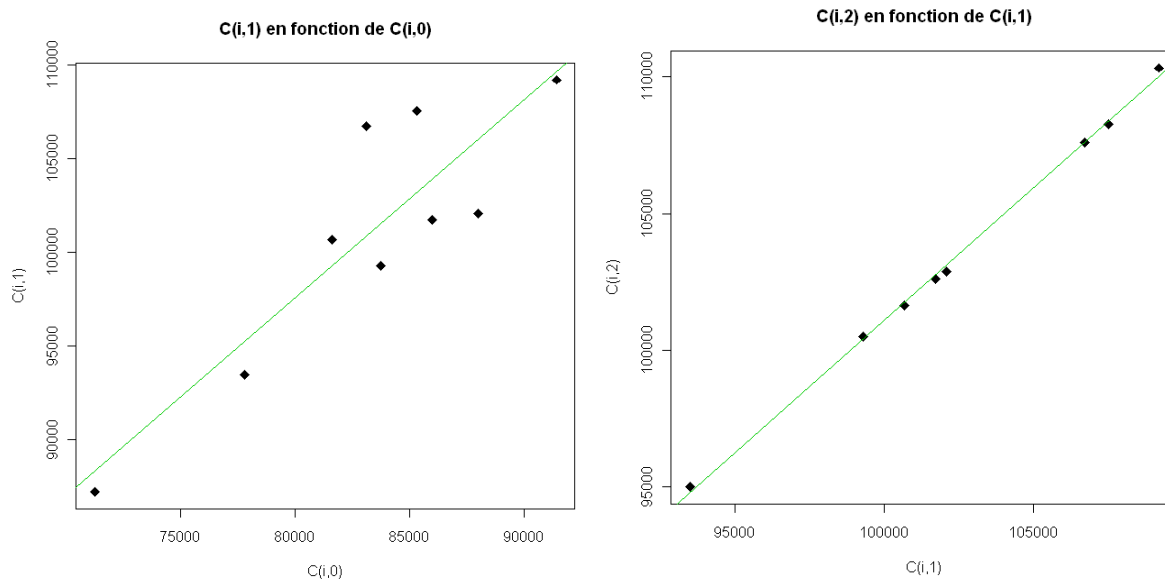
Malgré sa simplicité d'utilisation, il est important d'effectuer certaines vérifications pour valider cette méthode.

Nous allons retenir deux tests graphiques :

➤ **L'alignement des couples $(C_{i,j}, C_{i,j+1})_{i=0,\dots,n-j-1}$:**

Pour j fixé, nous avons supposé l'existence d'un coefficient f_j tel que $C_{i,j+1} = f_{i,j}C_{i,j}$. Les couples $(C_{i,j}, C_{i,j+1})_{i=0,\dots,n-j-1}$ doivent donc être sensiblement alignés par une droite passant par l'origine.

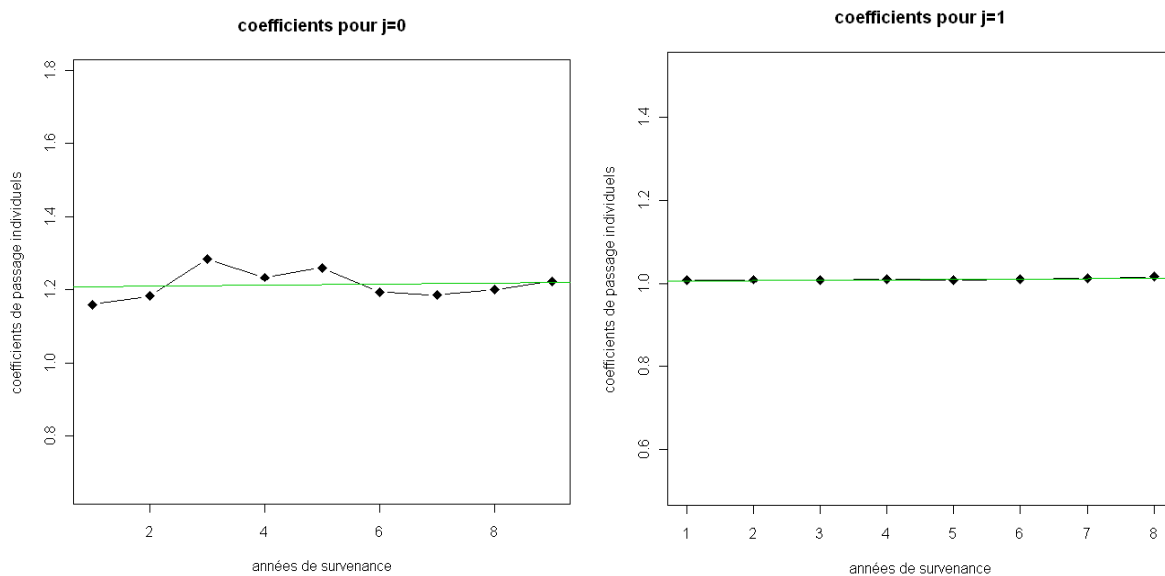
Graphiques pour le triangle Dom.Auto :



Il semblerait que pour l'année de développement 0, la condition ne soit pas vraiment vérifiée. Toutefois, à partir de l'année de développement 1, les couples sont sensiblement alignés sur une droite.

➤ **L'examen du triangle de développement :**

Le triangle de développement est défini par les facteurs de développement individuels $f_{i,j} = \frac{C_{i,j+1}}{C_{i,j}}$. L'hypothèse fondamentale est vérifiée si, pour $j=0, \dots, n-1$, les $(f_{i,j})$ sont sensiblement constants.



Les coefficients de développement sont également sensiblement constants à partir de la première année de développement.

Triangle de développement :

	dvp 0	dvp 1	dvp 2	dvp 3	dvp 4	dvp 5	dvp 6	dvp 7	dvp 8
surv 0	1,160	1,008	1,004	1,003	1,003	1,000	1,000	1,000	1,000
surv 1	1,182	1,009	1,005	1,003	1,004	1,001	1,000	1,000	
surv 2	1,284	1,008	1,002	1,000	1,002	1,001	1,000		
surv 3	1,233	1,010	1,004	1,004	1,000	1,000			
surv 4	1,260	1,007	1,004	1,004	1,001				
surv 5	1,194	1,010	1,005	1,002					
surv 6	1,185	1,012	1,006						
surv 7	1,202	1,016							
surv 8	1,224								

Les graphiques concernant les trois autres branches sont exposés en annexe.

Il ressort que pour le triangle de Dommages aux biens les conditions sont vérifiées de la même manière pour le triangle de Dommages Auto.

En revanche, la validation des conditions de Chain-Ladder est plus contestable pour les triangles de Responsabilité Civile, en particulier pour la branche RC Divers. En effet, pour ce triangle, les couples $(C_{i,j}, C_{i,j+1})$ ne sont pas alignés et les coefficients de développement ne sont pas sensiblement constants quelque soit l'année de développement.

II. Le modèle de Mack

Le modèle de Mack est la version stochastique de la méthode de Chain Ladder ; en effet, le montant des provisions estimé est identique. Il a cependant l'avantage d'estimer une erreur de prédiction des provisions.

Ce modèle non paramétrique repose sur trois hypothèses :

H1 - L'indépendance des années d'origine :

$\{C_{i,0}, \dots, C_{i,n}\}$ et $\{C_{k,0}, \dots, C_{k,n}\}$ sont indépendants pour $i \neq k$

H2 - Il existe f_j tel que $E(C_{i,j+1} | C_{i,1}, \dots, C_{i,n}) = f_j C_{i,j}$ pour $0 \leq i \leq n, 0 \leq j \leq n$

H3 - Il existe σ_j tel que $Var(C_{i,j+1} | C_{i,1}, \dots, C_{i,n}) = \sigma_j^2 C_{i,j}$ pour $0 \leq i \leq n, 0 \leq j \leq n$

Soit $D = \{C_{i,j} / i + j \leq n\}$ l'ensemble des variables observables.

Sous les hypothèses ci-dessus, $E(C_{i,n} | D) = C_{i,n-i} f_{n-i} \dots f_{n-1}$.

Dans ce modèle, les facteurs (f_j) sont estimés par les facteurs de développement de Chain Ladder (\hat{f}_j) qui sont sans biais et non corrélés.

Les (σ_j^2) sont estimés par
$$\left\{ \begin{array}{l} \hat{\sigma}_j^2 = \frac{1}{n-j-1} \sum_{i=0}^{n-j-1} C_{i,j} \left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j \right)^2, \quad 0 \leq j \leq n-2 \\ \hat{\sigma}_{n-1}^2 = \min(\hat{\sigma}_{n-2}^4 / \hat{\sigma}_{n-3}^2, \min(\hat{\sigma}_{n-3}^2, \hat{\sigma}_{n-2}^2)) \end{array} \right.$$

Cet estimateur est également sans biais.

Nous pouvons alors calculer l'écart quadratique moyen (mean squared error of prediction) qui mesure l'incertitude de prédiction. Pour les provisions par années d'exercice, il est défini par $MSEP(\hat{R}_i) = E[(\hat{R}_i - R_i)^2 | D]$.

Sous les hypothèses énoncées et si $\hat{C}_{i,n-i} = C_{i,n-i}$, on peut l'estimer par :

$$\widehat{MSEP}(\hat{R}_i) = \hat{C}_{i,n}^2 \sum_{j=n-i}^{n-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j} \left(\frac{1}{\hat{C}_{i,j}} + \frac{1}{\sum_{k=1}^{n-j} C_{i,k}} \right), i = 1, \dots, n$$

L'estimation du MSEP de la provision totale est alors donnée par:

$$\widehat{MSEP}(\hat{R}) = \sum_{i=1}^n \left\{ \widehat{MSEP}(\hat{R}_i) + \hat{C}_{i,n} \left(\sum_{k=i+1}^n \hat{C}_{k,n} \right) \sum_{j=n-i}^{n-1} \frac{2\hat{\sigma}_j^2}{\hat{f}_j^2 \sum_{u=0}^{n-j} C_{u,j}} \right\}$$

On peut aussi donner l'erreur standard relative i.e. l'écart-type rapporté au montant de charge

ultime $\frac{\sqrt{\widehat{MSEP}(\hat{R})}}{\hat{R}}$.

Tableau des erreurs standards relatives pour chaque branche :

	Dom. Auto	Dom. Biens	RC Auto	RC Divers
$\frac{\sqrt{\widehat{MSEP}(\hat{R})}}{\hat{R}}$	16,21%	16,83%	8,80%	21,88%

Nous constatons que l'erreur la plus importante concerne également la branche RC Divers.

III. Les modèles factoriels stochastiques

Dans le cadre de ces modèles, nous utilisons les triangles de paiements décumulés. L'hypothèse fondamentale sous-jacente est l'indépendance des incréments :

H : les variables $(X_{i,j})$ sont indépendantes pour $i, j=0 \dots n$

Nous cherchons à expliquer les paiements observés $(x_{i,j})$ d'un triangle en fonction de trois types de variables explicatives :

- Les variables années d'origine : (α_i) , $i=0 \dots n$ avec $\alpha_0=0$
- Les variables années de développement : (β_j) , $j=0 \dots n$ avec $\beta_0=0$
- Les variables années calendaires : (μ_{i+j}) . Pour un triangle déflaté, elles seront représentées par un unique paramètre μ .

Dans ce mémoire, deux types de modèles ont été retenus : la régression Log Normale et les modèles linéaires généralisés (GLM).

III.1. La régression Log Normale

➤ Le modèle

Nous supposons que les incréments $(X_{i,j})$ suivent chacun une loi $\mathcal{LN}(m_{i,j}, \sigma^2)$ où

$$m_{i,j} = \alpha_i + \beta_j + \mu$$

Soit la variable $Y_{i,j} = \ln(X_{i,j})$ (appelée log-incrément). Cette variable suit donc une loi $\mathcal{N}(m_{i,j}, \sigma^2)$ avec :

$$m_{i,j} = E(Y_{i,j}) \quad \text{et} \quad \text{var}(Y_{i,j}) = \sigma^2.$$

$$\text{Ainsi : } E(X_{i,j}) = \mu_{i,j} = e^{m_{i,j} + \frac{\sigma^2}{2}} \quad \text{et} \quad \text{var}(X_{i,j}) = e^{2m_{i,j} + \sigma^2} (e^{\sigma^2} - 1).$$

La régression Log Normale est donc un modèle linéaire Normal appliqué aux log-incréments du triangle : $Y = M\zeta + \varepsilon$ où :

- Y est le vecteur des log-incréments du triangle pris ligne à ligne (de taille N)
- M est la matrice de régression. Elle correspond à la matrice Jacobienne de la transformation $m: \zeta \rightarrow m(\zeta) = m_{ij}$ définie par

$$\begin{cases} \frac{\partial m_{ij}}{\partial \mu} = 1 \\ \frac{\partial m_{ij}}{\partial \alpha_k} = 1 \text{ si } k = i, 0 \text{ sinon} \\ \frac{\partial m_{ij}}{\partial \beta_l} = 1 \text{ si } l = j, 0 \text{ sinon} \end{cases}$$
- $\zeta = (\mu, \alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n)$ le vecteur des paramètres de régression (de taille p)
- ε le vecteur des erreurs, distribué selon une loi $\mathcal{N}(0, \sigma^2)$.

L'estimateur sans biais du maximum de vraisemblance de ζ est $\hat{\zeta} = (M'M)^{-1}Y$.

Notons (e_i) le résidu, estimation de l'erreur (ε_i) ; $e_i = Y_i - \hat{Y}_i$ où \hat{Y}_i est la $i^{\text{ème}}$ valeur prévue par le modèle.

L'estimateur sans biais de σ^2 est $S^2 = \frac{1}{n-p} \sum_{i=1}^n e_i^2$.

L'estimation des incréments est alors donnée par $\hat{\mu}_{ij} = e^{\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + S^2/2}$ et donc celle des provisions est $\hat{R} = \sum_{i+j > n} \hat{\mu}_{ij}$.

➤ Risque d'estimation et de prédiction

La matrice de variance-covariance des provisions $\hat{\Sigma}_{\hat{R}}$ est estimée à l'aide de la méthode Delta¹.

L'erreur d'estimation des provisions, erreur commise dans l'estimation des paramètres du

modèle, est donnée par : $\hat{\sigma}_{\hat{R}} = \sqrt{\sum_{i,j}(\hat{\Sigma}_{\hat{R}})_{ij}}$.

L'erreur de processus, liée à la variabilité des variables aléatoires qui composent le modèle,

est donnée par : $\sqrt{V(\hat{R})} = \sqrt{\sum_{i+j>n} V(\hat{X}_{ij})}$.

Enfin, nous obtenons l'erreur de prédiction: $\widehat{sep}_{\hat{R}} = \sqrt{V(\hat{R}) + \hat{\sigma}_{\hat{R}}^2}$.

Tableau des provisions, erreurs relatives et paramètres de dispersion :

	Dom. Auto	Dom. Biens	RC Auto	RC Divers
provisions	28449	21156	74745	113251
sep	99,4%	65,4%	29,3%	30,1%
S ²	0,594	0,316	0,211	0,345

Nous remarquons que l'erreur standard relative est très élevée pour les branches à développement court, en particulier pour la branche Auto dont le paramètre de dispersion est également important.

Par ailleurs, il semblerait que cette méthode surestime de manière générale le montant des provisions.

➤ Validation du modèle par examen des résidus

Les résidus sont supposés tous suivre une loi normale mais de paramètres différents. Il convient donc de les normaliser.

Nous utiliserons les résidus studentisés définis comme suit :

$$e_i^* = \frac{e_i}{s\sqrt{1-h_i}}$$

¹ L'application de la méthode Delta dans le cadre des modèles factoriels stochastiques est exposée dans PARTRAT C. et alii (2007), *Provisionnement Technique En Assurance Non-Vie. Economica*

où :

- s est l'estimation de σ l'erreur standard du modèle
- h_i est le $i^{\text{ème}}$ élément de la diagonale de la matrice de Hat, matrice de projection des Y_i sur les \hat{Y}_i : $H = M(M'M)^{-1}M'$

Leur étude permet en particulier de détecter les valeurs « atypiques », observations pour lesquelles $|e_i^*| > 2$.

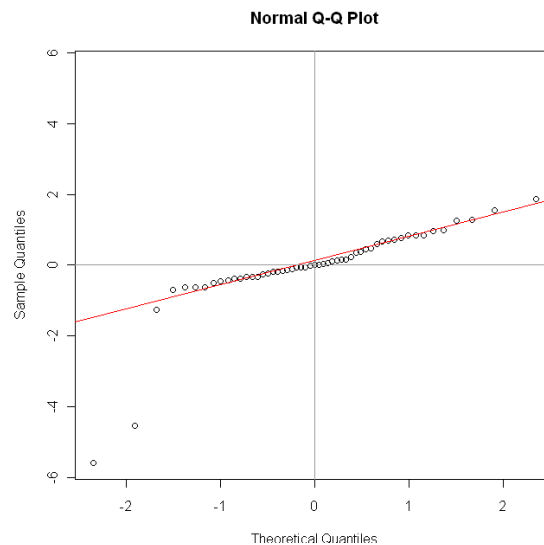
Après s'être assurés du caractère « indépendants et identiquement distribués » (iid) des résidus, nous devons également vérifier l'hypothèse de normalité. Nous pouvons dans un premier temps effectuer un graphique de type QQ-plot et contrôler l'alignement des points sur une droite de pente positive passant par 0. Pour compléter cet examen qualitatif, nous pouvons effectuer un test de Kolmogorov-Smirnov.

Enfin, l'examen des graphiques des résidus en fonction des valeurs prévues et des différentes variables explicatives ne doit faire apparaître aucune structure non aléatoire.

Exemple pour la branche Dommages Auto :

Le test d'indépendance des résidus donne une p-value de 0,637. Nous acceptons donc l'hypothèse que les résidus sont indépendants.

Le test de normalité de Kolmogorov-Smirnov donne quant à lui une p-value de 0,0415. Nous rejetons l'hypothèse de normalité des résidus pour un seuil de niveau 5%, ce que semble confirmer le graphique QQ-plot sur les résidus.

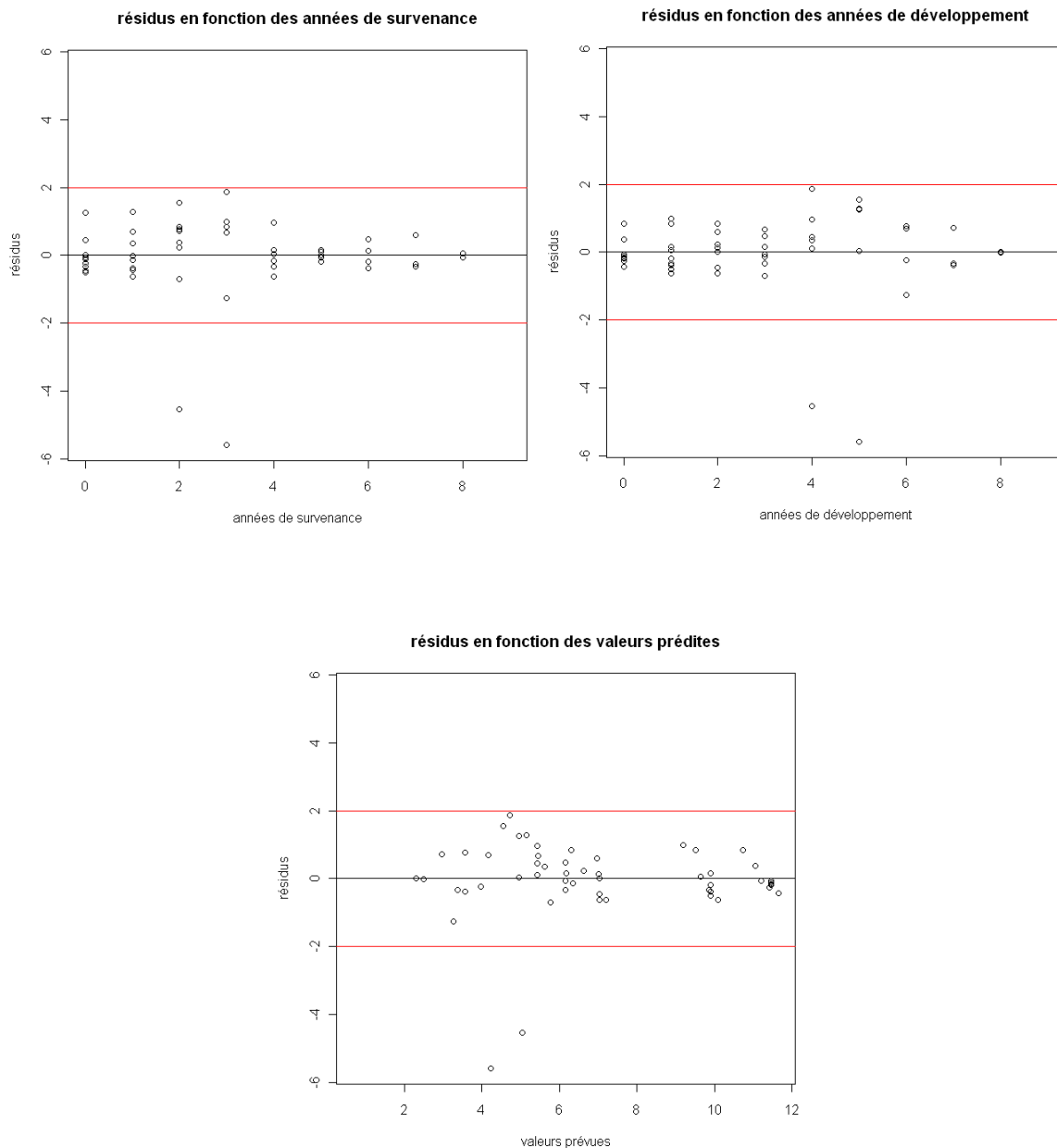


Enfin, les trois graphiques suivants ne laissent apparaître aucune structure non aléatoire.

Nous observons par ailleurs deux valeurs atypiques :

- année de survénance 2 et année de développement 4
- année de survénance 3 et année de développement 5

En effet, ces cellules ont des valeurs relativement faibles par rapport aux cellules d'années de développement similaires.



Les résultats détaillés des autres branches sont exposés en annexe. Nous pouvons noter que l'indépendance des résidus est toujours acceptée. L'hypothèse de normalité n'est quant à elle qu'approximativement respectée, en particulier pour la branche RC Auto.

III.2. La modélisation GLM

Les modèles linéaires généralisés ont été introduits en 1972 par J. Nelder et R. Wedderburn. Ils sont une généralisation du modèle linéaire normal et sont formés de trois composantes : la composante aléatoire, la composante systématique et la fonction de lien.

➤ Les modèles GLM

- **La composante aléatoire**

Nous cherchons toujours à expliquer les variables réponses ($X_{i,j}$). Nous supposons maintenant qu'elles suivent une loi de probabilité de type exponentielle. Leur densité est définie par la formule suivante :

$$f(x_{i,j}; \theta_{i,j}, \phi) = \exp \left\{ \frac{[\theta_{i,j}x_{i,j} - b(\theta_{i,j})]w_{i,j}}{\phi} + c(x_{i,j}, \phi) \right\}$$

où : $\theta_{i,j}$ est un paramètre réel, nommé paramètre naturel

ϕ est un paramètre de dispersion strictement positif

$w_{i,j}$ est une pondération (=1 par la suite)

b et c sont des fonctions caractéristiques du modèle, b étant deux fois dérivables à valeurs dans R et c à valeurs dans R².

- **La composante systématique :**

Soit M la matrice de régression et ζ le vecteur des paramètres. La composante systématique est notée η et est définie par $\eta = M\zeta$.

- **La fonction de lien :**

Notée g, c'est la fonction qui fait le lien entre la composante aléatoire et la composante systématique.

Une fois ces trois composantes définies, nous avons alors :

$$\begin{cases} \mu_{i,j} = g^{-1}(\eta_{i,j}) \\ E(X_{i,j}) = \mu_{i,j} \\ V(X_{i,j}) = \phi V(\mu_{i,j}) \end{cases}$$

Nous nous sommes particulièrement intéressés aux familles Poisson surdispersé et Gamma à lien log ce qui implique : $\mu_{i,j} = e^{\mu + \alpha_i + \beta_j}$.

Rappel : $X \sim \mathcal{P}_{surd}(\lambda, \phi)$ ssi $\frac{X}{\phi} \sim \mathcal{P}(\frac{\lambda}{\phi})$ ce qui implique $E(X) = \lambda$ et $V(X) = \phi\lambda$.

Famille	Fonction de répartition/Densité	Variance
Poisson $\mathcal{P}(\lambda)$	$P(X = x) = e^{x \ln(\lambda) - \lambda + c(x)}$	$V(\mu) = \mu$
Gamma $\Gamma(v, \frac{v}{\mu})$	$f(x) = \exp \left\{ \left(-\frac{x}{\mu} - \ln(\mu) \right) v + c(x, v) \right\}$	$V(\mu) = \mu^2$

L'estimation des incréments est alors donnée par $\hat{\mu}_{ij} = e^{\hat{\alpha}_i + \hat{\beta}_j}$ et donc celle des provisions est $\hat{R} = \sum_{i+j > n} \hat{\mu}_{ij}$.

➤ Risque d'estimation et de prédiction

La matrice de variance-covariance des provisions $\hat{\Sigma}_{\hat{R}}$ est toujours estimée à l'aide de la méthode Delta.

L'erreur d'estimation des provisions est donc donnée par : $\hat{\sigma}_{\hat{R}} = \sqrt{\sum_{i,j} (\hat{\Sigma}_{\hat{R}})_{ij}}$.

$$V(\hat{R}) = \hat{\phi} \sum_{i+j > n} V(\hat{\mu}_{ij}), \text{ avec } \hat{\phi} = \frac{1}{N-p} \sum_{i+j \leq n} \frac{(x_{ij} - \hat{\mu}_{ij})^2}{V(\hat{\mu}_{ij})}$$

(N étant le nombre d'éléments du triangle et p le nombre de paramètres).

L'erreur de prédiction est alors donnée par $s\hat{e}p_{\hat{R}} = \sqrt{V(\hat{R}) + \hat{\sigma}_{\hat{R}}^2}$.

Tableau des provisions, erreurs relatives et paramètres de dispersion :

	modèle Log Poisson surdispersé				modèle Log Gamma			
	Dom. Auto	Dom. Biens	RC Auto	RC Divers	Dom. Auto	Dom. Biens	RC Auto	RC Divers
provisions	21948	18436	68995	100112	22439	18598	70240	100179
sep	10,5%	11,0%	11,0%	22,4%	46,7%	48,6%	25,8%	26,1%
$\hat{\phi}$	166	118	247	1011	0,16	0,2	0,18	0,3

Le modèle Log Poisson nous donne bien les mêmes montants de provisions que la méthode Chain-Ladder ; en revanche, l'erreur de prédiction est relativement inférieure dans le cadre de ce modèle, hormis pour la branche RC Divers où l'ordre de grandeur est identique. En ce qui concerne le modèle Log Gamma, les montants sont sensiblement proches des résultats précédents. Nous pouvons néanmoins remarquer que l'erreur de prédiction est plus importante pour les branches à liquidation courte, comme pour la régression Log Normale.

➤ Validation du modèle par examen de résidus

Comme pour la régression Log Normale, il est nécessaire de passer par l'examen des résidus pour mesurer la qualité d'ajustement du modèle. Cependant, dans le cadre des modèles GLM, les résidus bruts ne sont pas pertinents. Il est préférable d'utiliser des résidus normalisés comme ceux de Pearson.

$$\text{Résidus de Pearson : } r_{ij}^p = \frac{x_{ij} - \hat{\mu}_{ij}}{\sqrt{V(\hat{\mu}_{ij})}}$$

Comme précédemment, les analyses graphiques et les tests de normalité sont effectués sur les résidus studentisés: $r_i^* = \frac{r_i^p}{\hat{\phi}\sqrt{1-h_i}}$

Application : (les graphiques sont exposés en annexe)

Pour le modèle Log Poisson surdispersé, l'analyse des résidus révèle un bon ajustement du modèle pour l'ensemble des branches.

Le modèle Log Gamma semble peu adapté à la branche RC Auto, comme la régression Log Normale. Il fait d'ailleurs apparaître les mêmes valeurs atypiques. Pour les autres branches, on observe une bonne qualité d'ajustement.

Tableau récapitulatif des provisions et des erreurs de prédiction associées :

	Chain-Ladder/Mack		Reg. Log Normale		Log Poisson Surd.		Gamma	
	Provisions	sep	Provisions	sep	Provisions	sep	Provisions	sep
Dom.Auto	21948	16,21%	28449	99,40%	21948	10,50%	22439	46,70%
Dom.biens	18436	16,83%	21156	65,40%	18436	11,00%	18598	48,60%
RC Auto	68995	8,80%	74745	29,30%	68995	11,00%	70240	25,80%
RC Divers	100112	21,88%	113251	30,10%	100112	22,40%	100179	26,10%

Le modèle Log Poisson, qui donne les mêmes montants de provisions que la méthode Chain-Ladder, avec toutefois une erreur de prédiction plus faible, semble être le modèle qui s'adapte le mieux aux données. Le modèle Log Gamma estime des montants du même ordre, et confirme donc les résultats précédents. En revanche, la régression Log Normale surestime l'ensemble des montants. L'erreur de prédiction associée est d'ailleurs supérieure à celle des autres méthodes, quelque soit la branche considérée, plus particulièrement pour les branches à liquidation courte.

Par la suite, nous allons donc principalement retenir le modèle Log Poisson Surdispersé et le modèle Log Gamma à titre de comparaison.

Comme nous l'avons dit précédemment, ces méthodes ne donnent qu'une estimation des provisions. Après avoir pris en compte l'inflation future et actualisé les paiements, nous obtenons les Best Estimates suivants :

	Chain-Ladder	Reg. Log Normale	Log Poisson Surd.	Gamma
Dom.Auto	21990	28518	21990	22489
Dom.biens	18455	21189	18455	18619
RC Auto	68552	74241	68552	69763
RC Divers	98711	111448	98711	98692

IV. Le bootstrap

Le bootstrap est une méthode de rééchantillonnage (avec remise) utilisant des simulations Monte-Carlo. Inventée par Efron en 1979, elle permet en particulier d'estimer la variabilité d'un paramètre. Simple à mettre en œuvre, cette méthode est de plus en plus utilisée en assurance non-vie. En effet, en la couplant avec les modèles GLM, nous pouvons ainsi avoir une estimation de la distribution des provisions.

L'utilisation du bootstrap suppose que les éléments de l'échantillon de départ soient indépendants et identiquement distribués (iid). Les variables $(X_{i,j})$ ne sont en général pas identiquement distribuées. Il est donc préférable d'avoir recours aux résidus du modèle, en particulier les résidus de Pearson car plus simples à calculer.

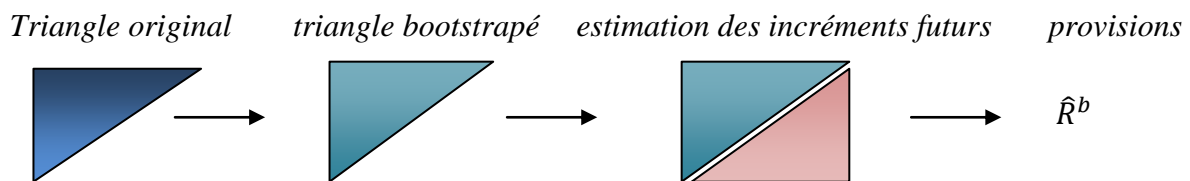
Remarques :

- pour corriger le biais dans la comparaison des estimations analytiques et bootstrap d'erreur de prédiction, il est conseillé d'ajuster les résidus en intégrant le nombre de paramètres de régression dans l'erreur bootstrap de prédiction. On définit les résidus ajustés par $r_{ij}^{aj} = \sqrt{\frac{N}{N-p}} r_{ij}^p$ où N est le nombre d'éléments de l'échantillon et p le nombre de paramètres de régression.
- Le modèle Log Poisson donnant les mêmes résultats que la méthode Chain-Ladder, l'utilisation de cette dernière est une solution pour diminuer le temps de calcul.

Description de la méthode du bootstrap dans le cadre de la modélisation GLM:

- (i) Ajustement un modèle linéaire et estimation des valeurs prévues ($\hat{\mu}_{ij}$).
- (ii) Calcul des résidus de Pearson du modèle. : $r_{ij}^p = \frac{x_{ij} - \hat{\mu}_{ij}}{\sqrt{V(\hat{\mu}_{ij})}}$ (*)
- (iii) Pour $b=1 \dots B$ ($B=5000$ ou 10000) :
- Rééchantillonnage avec remise des résidus.
Ces résidus bootstrapés sont notés $(r_{ij}^p)^b$
 - En inversant la relation (*), obtention d'un nouveau triangle, appelé « pseudo-triangle », composé des valeurs $(x_{ij})^b = \hat{\mu}_{ij} + (r_{ij}^p)^b \sqrt{V(\hat{\mu}_{ij})}$.
 - À l'aide d'une nouvelle modélisation GLM appliquée au pseudo-triangle, détermination d'une nouvelle estimation des provisions notée \hat{R}^b .

Schéma de la procédure à effectuer B fois :



À partir du B-échantillon des provisions (\hat{R}^b), nous pouvons alors estimer la moyenne, la variance et l'erreur de prédiction du montant de provisions.

$$E^B(\hat{R}) = \frac{1}{B} \sum_{b=1}^B \hat{R}^b$$

$$\text{Risque d'estimation : } Var^B(\hat{R}) = \frac{1}{B-1} \sum_{b=1}^B [\hat{R}^b - E^B(\hat{R})]^2$$

$$\text{Risque de processus : } \hat{V}(R) = \hat{\phi} \sum_{i+j > n} V(\hat{\mu}_{ij})$$

$$\text{Erreur de prédiction : } sep = \sqrt{\hat{V}(R) + Var^B(\hat{R})}$$

Nous avons utilisé cette méthode pour 10 000 simulations avec les modèles Log Poisson surdispersé et Log Gamma.

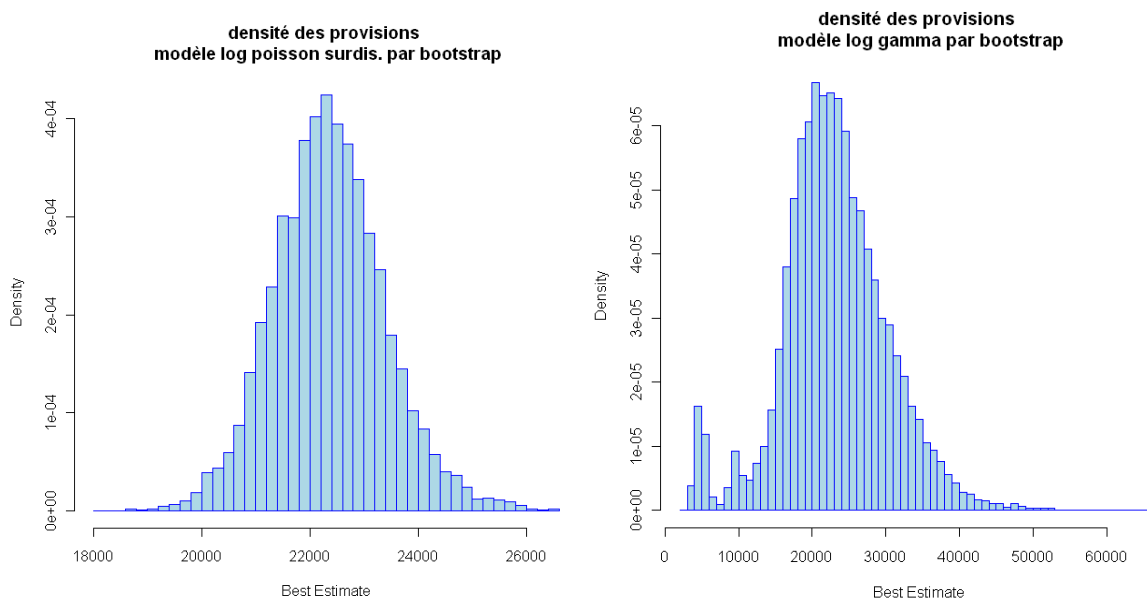
Tableau récapitulatif des résultats donnés par le bootstrap :

Branches	Modèles	\hat{R}	$E^B(\hat{R})$	$Var^B(\hat{R})$	$\hat{V}(R)$	sep/R
Dom.Auto	Poisson	22081	22366	1070766	3641532	9,83%
	Gamma	22577	23063	52525133	44238762	43,57%
Dom.Biens	Poisson	21271	18823	1285491	2177175	8,75%
	Gamma	18705	19092	38411983	31414489	44,67%
RC auto	Poisson	73184	72930	39625465	17012415	10,28%
	Gamma	74898	75704	195673134	75421929	21,98%
RC Divers	Poisson	132562	130657	993982369	101221821	24,96%
	Gamma	138799	140802	1211863474	95656373	26,05%

NB : Ces résultats tiennent compte de l'inflation future et de l'actualisation. Ils sont donc comparables à des Best Estimates. Nous avons également « terminé » les triangles à l'aide de facteurs de queue, dont la méthodologie est exposée dans la partie suivante.

Cette méthode nous donne des résultats sensiblement proches de ceux trouvés par les méthodes classiques, avec une erreur de prédiction légèrement plus faible. En outre, elle permet d'estimer la distribution des provisions et les quantiles associés.

Graphiques des provisions pour la branche Auto :



VaR à 95% et VaR à 99,5% des provisions :

		VaR à 95%	VaR à 99,5%
Dom.Auto	Poisson	24121	25542
	Gamma	35161	44694
Dom.Biens	Poisson	20767	22040
	Gamma	29034	43395
RC Auto	Poisson	84063	90991
	Gamma	101722	124436
RC Divers	Poisson	186383	244708
	Gamma	203266	268317

De manière générale, le modèle Log Gamma donne des distributions à queue plus épaisse que le modèle Log Poisson Surdispersé.

V. Extensions et limites des modèles

V.1. Gestion de la liquidation incomplète

En général, nous ne disposons pas du développement complet de la sinistralité. Il est donc nécessaire d'estimer une queue de développement du triangle et le risque associé.

Nous avons étudié deux méthodes d'estimation de cette queue de développement. La première est basée sur les coefficients de Chain-Ladder, la seconde sur les paramètres de régression des modèles GLM.

➤ Calcul du facteur de queue par extrapolation des coefficients de Chain-Ladder

Dans la première partie de ce chapitre, nous avons estimé des coefficients de passage \hat{f}_j . Nous allons maintenant chercher à ajuster une fonction $y = f(t)$ régulière et vérifiant $f(t) \geq 1$ sur ces points afin de les extrapoler. Nous avons choisi une fonction du type fonction puissance inverse $f(j) = 1 + a(1 + j)^{-b}$. La détermination des paramètres de cette fonction s'effectue par minimisation des carrés des écarts aux facteurs empiriques :

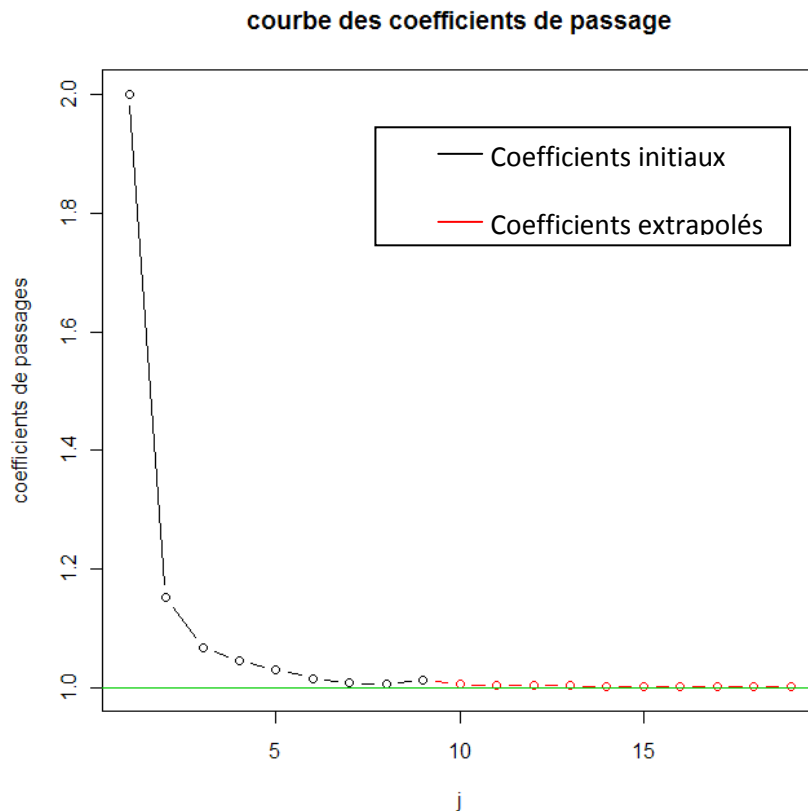
$$\min \sum_{j=0}^{n-1} [(\hat{f}_j - 1) - a(j + 1)^{-b}]^2$$

Le passage au Log permet de résoudre ce problème par les moindres carrés ordinaires :

$$\min \sum_{j=0}^{n-1} [\ln(\hat{f}_j - 1) - [\ln a - b * \ln(1 + j)]]^2$$

Exemple pour la branche RC Auto :

$a = 0.844$ et $b = 2.19$

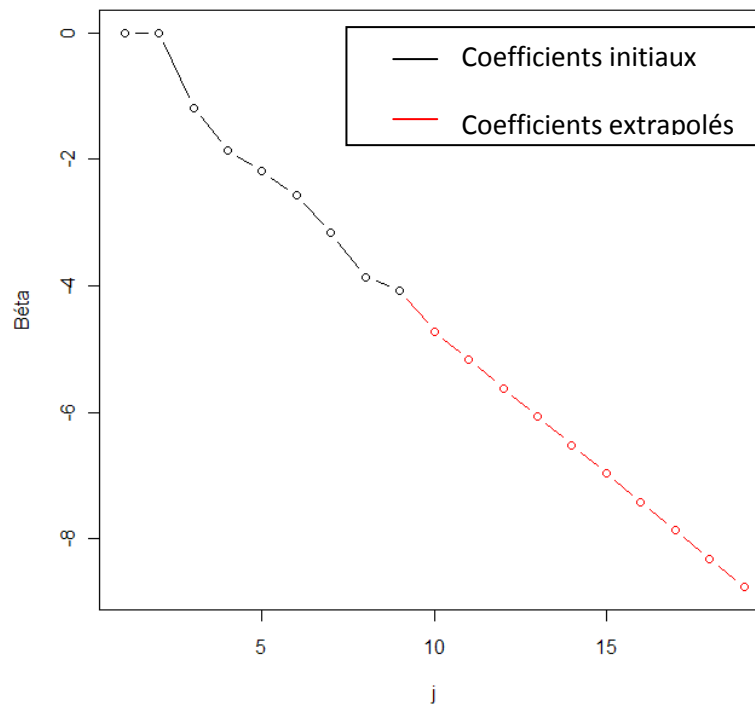


Théoriquement, si un triangle de liquidation est fini, le dernier coefficient de passage est égal à 1. Pour mettre en pratique cette méthode, nous avons fait l'hypothèse forte que le dernier coefficient de passage serait égal à 1,0001. Le facteur de queue sera alors égal au produit des coefficients (partant de la dernière année de survenance, jusqu'au dernier coefficient)

➤ Calcul du facteur de queue par extrapolation des paramètres (β_j)

Dans le cadre des modèles GLM, le passage d'une année de développement à la suivante est représenté par les paramètres (β_j). Ce sont donc ces coefficients que nous allons chercher à extrapoler. Une régression linéaire simple de type $y = ax + b$ semble bien s'adapter à l'ajustement des (β_j) ; ils seront donc estimés par cette méthode pour $j > n$.

régression des β_j



Se pose encore la question : à quel moment arrêter l'extrapolation ?

Nous avons de nouveau dû faire une hypothèse forte : continuer l'estimation des (β_j) tant que l'espérance de l'incrément associé est supérieur à 1.

Tableau des B.E. avec facteur de queue :

	Chain-Ladder	Reg. Log Normale	Log Poisson Surd.	Gamma
Dom.Auto	21993	28619	22081	22577
Dom.biens	18455	21271	18545	18705
RC Auto	84441	79170	73184	74898
RC Divers	242869	154408	132562	138799

Les BE avec facteur de queue des branches Dommages sont très proches des BE sans facteur de queue. En effet, ces branches sont à liquidation courte et arrivent donc en fin de développement après neuf années.

Les branches RC sont quant à elles à liquidation longue. La prise en compte du facteur de queue est donc beaucoup plus importante. Il semble par ailleurs que la méthode basée sur les coefficients de Chain-Ladder soit peu appropriée pour ce type de branches.

V.2. Les recours

Une des limites des modèles présentés est qu'ils ne sont applicables qu'à des triangles bruts de recours. En effet, leur utilisation n'est possible que si les charges incrémentales sont positives. Or, dans certaines branches telles que les branches de responsabilité civile, les recours ont un développement plus long que celui des paiements. Les triangles nets de recours peuvent donc avoir des valeurs négatives dans les dernières années.

Dans la réforme Solvabilité II, les Best Estimates sont calculés sur les triangles nets de recours. Il convient donc d'adapter nos méthodes pour pouvoir prendre en compte les recours.

Nous considérons un triangle de règlement brut de recours et un triangle de recours. Par l'une des méthodes citées précédemment, nous estimons les charges à l'ultime du triangle de paiements.

Nous ne pouvons pas appliquer directement l'une de ces méthodes au triangle de recours pour trouver les charges ultimes associées. En effet, le développement des recours dépend de son historique, mais aussi de la liquidation du triangle de paiements. Nous supposons donc que les montants de recours peuvent s'exprimer à l'ultime en fonction des sinistres bruts à l'ultime.

$$recours_{ultimes} = (1 - Y)bruts_{ultimes}$$

Méthodologie pour estimer Y :

- Par la méthode Chain-Ladder, estimation des charges ultimes du triangle de recours et les coefficients de développement F_j .
- Des charges de sinistres bruts, calcul de $Y^* = 1 - \frac{recours_{ultimes}}{bruts_{ultimes}}$.
- Estimation de Y en supposant que $Y \sim \mathcal{LN}(\mu, \sigma^2)$ avec :
$$\mu = \ln E(Y^*) - \frac{1}{2} \ln \left(1 + \frac{V(Y^*)}{E(Y^*)^2} \right) \text{ et } \sigma^2 = \ln \left(\frac{V(Y^*)}{E(Y^*)^2} + 1 \right)$$

Ainsi, par une procédure du type bootstrap, pour chaque simulation de charges ultimes de sinistres bruts, nous simulons Y pour en déduire une charges ultimes de recours. Par application rétrospective des coefficients de passage F_j , nous obtenons le triangle inférieur des recours et en actualisant les flux, le Best Estimate des recours.

Enfin le Best Estimate des triangles nets de recours est donné par :

$$BE_{net\ recours} = BE_{bruts} - BE_{recours}$$

➤ Exemple d'application sur le triangle de RC Auto :

Triangle historique des recours :

	dvp 1	dvp 2	dvp 3	dvp 4	dvp 5	dvp 6	dvp 7	dvp 8	dvp 9	dvp 10	Y*
surv 1	1697	2885	847	707	323	538	40	17	76	142	0,8793
surv 2	1821	3434	865	582	333	146	86	150	46		0,8705
surv 3	1394	3458	1216	455	327	450	133	54			0,8668
surv 4	1872	4642	1110	620	311	79	370				0,8433
surv 5	1434	3202	1156	633	638	153					0,8732
surv 6	1932	2700	1389	657	582						0,8671
surv 7	1506	2294	807	322							0,8805
surv 8	1474	2426	1074								0,8718
surv 9	1349	3138									0,8506
surv 10	1585										0,8450

La variable Y va donc suivre une loi $\mathcal{LN}(-0.145, 0.025\%)$.

Best Estimates (10 000 simulations) :

	sinistres bruts	recours	sinistres nets de recours
BE	72947	14154	58793
VaR 99,5%	92186	16639	76888

La prise en compte des recours dans les branches à développement long diminue de manière significative le montant du Best Estimate et le quantile associé. Néanmoins, l'évaluation du risque à un an a été effectuée sur les triangles bruts de recours.

PARTIE III : EVALUATION DU RISQUE A UN AN

Comme nous l'avons déjà vu, la réglementation Solvabilité II détermine le niveau de provisions requis à l'aide du critère de Value-at-Risk (VaR). La VaR est une mesure de risque définie comme suit :

Étant donné un risque X et un niveau de probabilité $\alpha \in (0,1)$, la VaR correspondante est le quantile d'ordre α de X , soit $VaR(X, \alpha) = F_X^{-1}(\alpha)$.

Elle correspond au montant de perte maximale pour un niveau de risque et un horizon temporel donnés. Pour le calcul du SCR, le niveau de risque doit être inférieur à 0,05% (soit un seuil de confiance de 99,5%) et l'horizon temporel est d'un an.

La VaR vérifie les propriétés suivantes :

- Invariance par translation : $VaR(X + c, \alpha) = VaR(X, \alpha) + c, c \in \mathbb{R}$
- Homogénéité : $VaR(\lambda X, \alpha) = \lambda VaR(X, \alpha), \lambda \geq 0$
- Monotonie : $X \leq Y \Rightarrow VaR(X, \alpha) \leq VaR(Y, \alpha), \alpha \in (0,1)$

Dans cette partie, nous allons exposer deux méthodes permettant de calculer le SCR non-vie. La première est une application du modèle d'évaluation du risque à un an de Merz-Wüthrich. La seconde est une adaptation de la technique du bootstrap exposée précédemment.

I. Modèle de Merz-Wüthrich

Le modèle de Merz-Wüthrich est une adaptation du modèle de Mack (vision à l'ultime) pour estimer le risque à un an. Nous nous plaçons donc sous les hypothèses de Chain Ladder et considérons des paiements cumulés.

Pour estimer le risque à un an, nous allons chercher à analyser l'incertitude liée aux Claims Development Results (CDR) i.e. liée à la différence entre deux estimations successives du montant des provisions.

Il existe deux facteurs d'incertitude liés aux CDR :

- un facteur de variance dû à la prédiction des paiements futurs (par simulations aléatoires)
- un facteur d'estimation lié au modèle choisi

I.1. Méthodologie

Nous nous plaçons à l'année I et considérons l'année suivante I+1.

Le CDR réel pour l'année comptable (I, I+1] est défini par :

$$\begin{aligned} CDR_i(I+1) &= E(R_i^I | D_I) - (X_{i,I-i+1} + E(R_i^{I+1} | D_{I+1})) \\ &= E(C_{i,J} | D_I) - E(C_{i,J} | D_{I+1}) \end{aligned}$$

Où : $R_i^I = C_{i,J} - C_{i,I-i}$

$R_i^{I+1} = C_{i,J} - C_{i,I-i+1}$

$X_{i,I-i+1} = C_{i,I-i+1} - C_{i,I-i}$

$D_I = \{C_{i,j} ; i+j \leq I \text{ et } i \leq I\}$, le triangle de développement connu à la date I

$D_{I+1} = \{C_{i,j} ; i+j \leq I+1 \text{ et } i \leq I\}$, le triangle de développement à la date I+1

Le CDR réel agrégé est défini par : $CDR_{agg} = \sum_{i=1}^I CDR_i(I+1)$.

Ces deux valeurs vont être estimées respectivement par :

- le CDR observable $\widehat{CDR}_i(I+1) = \hat{C}_{i,J}^I - \hat{C}_{i,J}^{I+1}$
- le CDR observable agrégé $\widehat{CDR}_{agg} = \sum_{i=1}^I \widehat{CDR}_i(I+1)$

L'objectif suivant est de quantifier le degré de variabilité du CDR observable agrégé à la date I+1.

La variabilité entre le CDR réel et le CDR observable est donnée par l'écart quadratique moyen, notée MSEP (mean square error of prediction) :

$$MSEP_{CDR_i|D_I}(\widehat{CDR}_i(I+1)) = E \left[\left(CDR_i(I+1) - \widehat{CDR}_i(I+1) \right)^2 | D_I \right]$$

De même la variabilité du CDR observable autour de 0 est donnée par :

$$MSEP_{\widehat{CDR}_i(I+1)|D_I}(0) = E \left[\left(\widehat{CDR}_i(I+1) - 0 \right)^2 | D_I \right]$$

et est estimée par : $\widehat{MSEP}_{\widehat{CDR}_i(I+1)|D_I}(0) =$

$$(\hat{C}_{i,J}^I)^2 \left[\frac{\hat{\sigma}_{I-i}^2 / (\hat{f}_{I-i}^I)^2}{C_{i,I-i}} + \frac{\hat{\sigma}_{I-i}^2 / (\hat{f}_{I-i}^I)^2}{\sum_{k=0}^{I-1} C_{k,I-i}} + \sum_{j=I-i+1}^{J-1} \frac{C_{I-j,j}}{\sum_{k=0}^{I-j} C_{k,j}} * \frac{\hat{\sigma}_j^2 / (\hat{f}_j^I)^2}{\sum_{k=0}^{I-j-1} C_{k,j}} \right]$$

$$\text{où } \hat{f}_j^I = \frac{\sum_{i=0}^{I-j-1} C_{i,j+1}}{\sum_{i=0}^{I-j-1} C_{i,j}}$$

Nous pouvons alors estimer l'erreur de prédiction du CDR observable agrégé :

$$\widehat{MSEP}_{\Sigma_i \widehat{CDR}_i(I+1)|D_I}(0) = \sum_i \widehat{MSEP}_{\widehat{CDR}_i(I+1)|D_I}(0) + 2 \sum_{i < l} \hat{C}_{i,J}^I \hat{C}_{l,J}^I \left[\frac{\hat{\sigma}_{I-i}^2 / (\hat{f}_{I-i}^I)^2}{\sum_{k=0}^{i-1} C_{k,I-i}} + \sum_{j=I-l+1}^{J-1} \frac{C_{I-j,j}}{\sum_{k=0}^{I-j} C_{k,j}} * \frac{\hat{\sigma}_j^2 / (\hat{f}_j^I)^2}{\sum_{k=0}^{I-j-1} C_{k,j}} \right]$$

C'est cette dernière erreur de prédiction qui nous donne la variabilité à un an des provisions.

I.2. Utilisation dans le calcul du SCR non-vie

➤ Formule standard

La cinquième étude quantitative d'impacts (QIS 5) propose différentes méthodes permettant d'estimer les paramètres utilisés dans la formule standard. L'une d'elle utilise l'erreur de prédiction estimée par le modèle de Merz-Wüthrich pour calculer l'écart-type du risque de provisions.

Nous avons $\sigma^*_{lob} = \frac{\sqrt{MSEP}}{BE_CL_{lob}}$ où $MSEP$ est donnée par l'estimation ci-dessus et BE_CL_{lob} est le Best Estimate du montant des provisions estimé par la méthode de Chain-Ladder.

Résultats des SCR donnés par la formule standard avec différents paramètres :

	Paramètres donnés dans le QIS 5 (Partie I)			Paramètres donnés par le modèle de Merz- Wüthrich		
	σ_{lob}	SCR		σ^*_{lob}	SCR	
		poisson	gamma		poisson	gamma
Dom.Auto	10%	6327	6469	16%	10757	10999
Dom.Biens	11%	5906	5957	16,5%	9363	9444
RC Auto	9,5%	19819	20283	6%	12066	12348
RC Divers	11%	42218	44204	13%	50921	53317

Dans notre cas, il semblerait que les écart-types du risque de provisions donnés dans le QIS 5 sous-estime le montant du SCR.

➤ Simulations de lois Log Normales

Nous supposons que les provisions suivent une loi Log Normale de paramètres (μ, σ^2) avec :

$$\mu = \ln(BE_CL_{lob}) - \frac{1}{2} \ln \left(1 + \frac{MSEP}{BE_CL_{lob}^2} \right)$$

$$\sigma^2 = \ln \left(\frac{MSEP}{BE_CL_{lob}^2} + 1 \right)$$

En effectuant 10 000 simulations de cette loi, nous obtenons un quantile à 99,5% des provisions.

Toutefois, il est important de remarquer que le modèle de Merz-Wüthrich évalue le risque à un an des provisions et non du Best Estimate. Nous allons donc faire l'hypothèse suivante :

Quelque soit le triangle à l'ultime, $BE = \alpha\%$ Provisions.

À partir des résultats de Merz-Wüthrich, nous estimons la VaR à 99,5% à un an des provisions et en déduisons la VaR à 99,5% du BE à un an :

$$VaR_{99,5\%}^1 \left(\frac{BE_1}{1+r} + \frac{P_1}{\sqrt{1+r}} - BE_0 \right) = \alpha\% VaR_{99,5\%}^1 (provisions_1 + P_1 - provisions_0)$$

Application pour 10 000 simulations :

Branches	SCR provisions	alpha	SCR BE
Dom.Auto	10804	1,0055	10864
Dom.Biens	9618	1,0041	9658
RC auto	11651	0,9891	11524
RC Divers	38892	0,9646	37517

Pour les branches Dommages Auto et Dommages aux Biens le paramètre alpha est supérieur à 1. Ceci est dû au fait que ce sont des branches à liquidation courtes : l'actualisation n'a que peu d'impact sur les flux futurs (à faibles montants).

Par ailleurs, cette méthode semble sous-estimer le montant du SCR pour les branches de Responsabilité Civile.

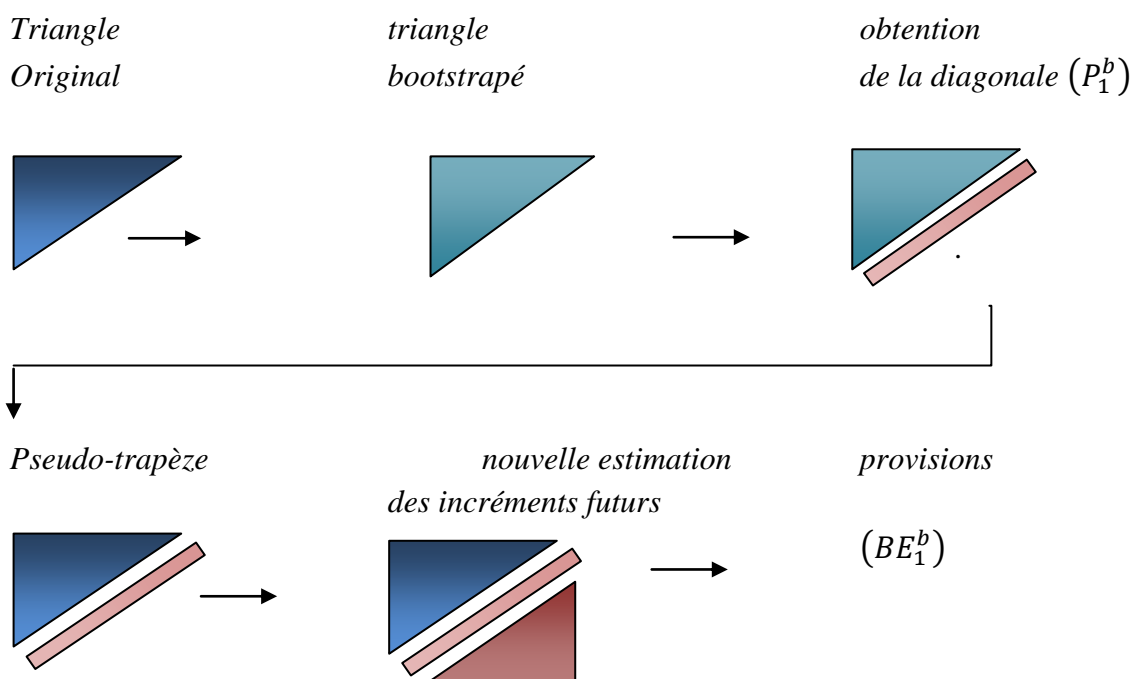
II. Adaptation du bootstrap pour l'évaluation du risque à un an

Une méthode alternative est l'adaptation de la méthode du Bootstrap. Nous cherchons à simuler un grand nombre de nouvelles « diagonales » d'incrémentes i.e. les paiements survenus entre n et $n+1$.

Les premières étapes sont similaires à la méthode traditionnelle : rééchantillonnage des résidus et obtention du pseudo-triangle. Ensuite :

- par modélisation GLM, tirage de la première diagonale notée (P_1^b) .
- intégration au triangle initial et obtention d'un « trapèze ».
- application d'une nouvelle modélisation GLM et estimation des provisions à un an, (BE_1^b) .
- on réitère ce procédé B fois.

Schéma de la procédure à effectuer B fois :



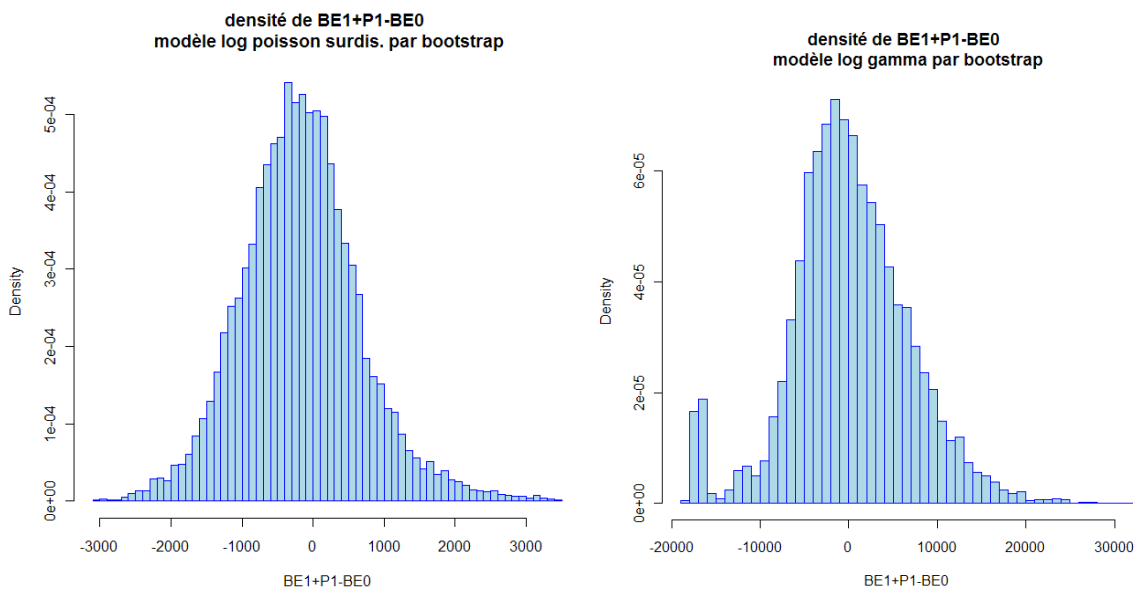
À l'aide de ces simulations, nous pouvons calculer le SCR : $VaR_{99.5\%} \left(\frac{BE_1}{1+r_1} + \frac{P_1}{\sqrt{1+r_1}} - BE_0 \right)$
 où r_1 est le taux d'actualisation à un an et BE_0 le best estimate du montant de provisions à provisionner calculer en $t=0$ par l'une des méthodes exposées dans la partie précédente.

Résultats du bootstrap à un an :

Branches	Modèles	moyenne	SCR
Dom.Auto	Poisson	-170	4607
	Gamma	115	19370
Dom.Biens	Poisson	-142	3653
	Gamma	4,92	21648
RC auto	Poisson	-1065	15485
	Gamma	-630	37173
RC Divers	Poisson	-8914	50032
	Gamma	-1751	54249

Il semblerait que cette méthode sous-estime le SCR des branches à développement court dans le cas Log Poisson. En revanche les résultats pour les branches RC sont relativement semblables à ceux trouvés par la formule standard ou à l'aide du modèle de Merz-Wüthrich, mis à part le SCR obtenu par modélisation Gamma sur la RC Auto (nous avons remarqué précédemment que ce modèle n'était pas adapté au triangle RC Auto).

Distribution de $\frac{BE_1}{1+r_1} + \frac{P_1}{\sqrt{1+r_1}} - BE_0$ pour la branche Dommage Auto :



PARTIE IV : PRISE EN COMPTE DE LA DÉPENDANCE ENTRE BRANCHE

Jusqu'à présent, nous nous sommes uniquement intéressés au besoin de fonds propres par branches d'activités, ce qui supposait que ces branches étaient indépendantes. En pratique, cette hypothèse est rarement vérifiée. En effet, on observe, par exemple, une dépendance entre la branche Dommages Auto et la branche Dommages aux Biens dans le cas d'une tempête. Ou encore, un sinistre en assurance Auto peut engendrer à la fois des dommages au véhicule et des dommages corporels, mettant donc en jeu l'assurance Dommages Auto et la Responsabilité Civile Auto. Il convient donc de prendre en compte cette dépendance dans le calcul du SCR et d'en mesurer l'impact.

La copule est un outil flexible représentant les structures de dépendance entre deux variables. Nous allons donc l'utiliser pour modéliser la dépendance entre branches.

Tout d'abord, nous allons décrire différents types et mesures de dépendances, puis nous présenterons une introduction à la théorie des copules. Enfin un cas pratique illustrera la modélisation de la dépendance entre branche par l'utilisation des copules.

I. La dépendance

I.1. Notion de dépendance et corrélation

Les notions de dépendance et de corrélation sont souvent confondues ; ce sont pourtant deux concepts différents mais qui sont toutefois associés : si X et Y sont indépendantes, alors X et Y sont non corrélées (linéairement). La réciproque est fautive.

La dépendance englobe tout type de liaisons, linéaires et non linéaires. Le coefficient de corrélation linéaire n'est qu'une mesure de dépendance particulière qui perd de sa pertinence en dehors de l'univers gaussien. Nous nous intéresserons donc à d'autres mesures de dépendance, en particulier aux coefficients de corrélation de rang.

I.2. Mesures de dépendance

➤ Coefficient de Pearson (corrélation linéaire) :

Il est défini par $\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)} \in [-1; 1]$ où X et Y sont deux variables aléatoires.

X et Y sont linéairement liées si et seulement si $\rho(X, Y) \neq 0$.

X et Y sont non linéairement liées si et seulement si $\rho(X, Y) = 0$.

Comme nous l'avons dit précédemment, ce coefficient s'utilise si la loi du couple (X, Y) est gaussienne, ce qui est rarement le cas en assurance. Par ailleurs, il est très sensible aux valeurs extrêmes. Dans notre cas, il convient donc de s'intéresser aux coefficients de corrélation de rang comme le « tau » de Kendall ou le « rhô » de Spearman. Basés sur les notions de concordance et de discordance, ils sont robustes car non-sensibles aux valeurs extrêmes.

Notion de concordance-discordance :

Soient (X_1, X_2) et (X'_1, X'_2) deux couples indépendants et identiquement distribués.

La probabilité de concordance est la probabilité d'avoir simultanément $X_1 < X'_1$ et $X_2 < X'_2$ (ou $X_1 > X'_1$ et $X_2 > X'_2$).

La probabilité de discordance est la probabilité d'avoir simultanément $X_1 < X'_1$ et $X_2 > X'_2$ (ou $X_1 > X'_1$ et $X_2 < X'_2$).

➤ « Tau » de Kendall :

Le tau de Kendall associé au couple de variables aléatoires (X_1, X_2) est défini par

$$\tau(X_1, X_2) = \mathbb{P}((X_1 - X'_1)(X_2 - X'_2) > 0) - \mathbb{P}((X_1 - X'_1)(X_2 - X'_2) < 0)$$

avec (X_1, X_2) et (X'_1, X'_2) iid.

➤ « Rhô » de Spearman :

Soit (X_1, X_2) un couple de variables aléatoires et (X'_1, X'_2) et (X_1^*, X_2^*) deux couples identiques à (X_1, X_2) . Le rho de Spearman est alors défini par :

$$\rho_S(X_1, X_2) = 3[\mathbb{P}((X_1 - X'_1)(X_2 - X_2^*) > 0) - \mathbb{P}((X_1 - X'_1)(X_2 - X_2^*) < 0)]$$

Ces deux coefficients vérifient les propriétés suivantes :

- $\delta(X, Y) = \delta(Y, X), \forall (X, Y)$
- $-1 \leq \delta(X, Y) \leq 1, \forall (X, Y)$
- $\delta(X, Y) = 1$ ssi X et Y sont comonotones
- $\delta(X, Y) = -1$ ssi X et Y sont antimonotones
- Si g est une fonction strictement croissante alors $\delta(g(X), g(Y)) = \delta(X, Y)$

où δ est une mesure de dépendance.

Dans la suite du mémoire, nous privilégierons l'utilisation du τ de Kendall car nous pourrions exprimer les paramètres des copules utilisées en fonction de ce dernier.

I.3. Types de dépendance

➤ Dépendance parfaite

On parle de dépendance parfaite lorsque deux risques peuvent s'écrire comme des fonctions croissantes ou décroissantes d'une même variable aléatoire sous-jacente.

Le couple (X_1, X_2) est dit **comonotone** s'il existe des fonctions non-décroissantes g_1 et g_2 et une variable aléatoire Z telles que $(X_1, X_2) \stackrel{=}{\text{loi}} (g_1(Z), g_2(Z))$. Dans ce cas on parle de dépendance totale positive.

Le couple (X_1, X_2) est dit **antimonotone** s'il existe une fonction non-décroissante g_1 , une fonction non-croissante g_2 et une variable aléatoire Z telles que $(X_1, X_2) \stackrel{=}{\text{loi}} (g_1(Z), g_2(Z))$. On parle alors de dépendance totale négative.

➤ Dépendance de queue

Elle fournit une description de la dépendance au niveau des queues de distribution, soit l'étude de la dépendance des valeurs extrêmes.

Soit deux variables aléatoires X_1, X_2 de fonctions de répartition respectives F_1, F_2 .

Le **coefficient de dépendance de queue inférieure** (« lower tail dependence coefficient ») est défini par : $\lambda_L(X_1, X_2) = \lim_{\alpha \rightarrow 0^+} \mathbb{P} (X_1 \leq F_1^{-1}(\alpha) | X_2 \leq F_2^{-1}(\alpha))$.

De façon simplifiée on mesure la probabilité que X_1 soit petit sachant que X_2 est petit.

Le **coefficient de dépendance de queue supérieure** (« upper tail dependence coefficient ») est défini par : $\lambda_U(X_1, X_2) = \lim_{\alpha \rightarrow 1^-} \mathbb{P} (X_1 \geq F_1^{-1}(\alpha) | X_2 \geq F_2^{-1}(\alpha))$.

On mesure la probabilité que X_1 soit grand sachant que X_2 est grand.

Remarque : F_i^{-1} est la fonction quantile définie par : $F_i^{-1}(t) = \inf\{x \in \mathbb{R}, F_i(x) \geq t\} \forall t \in [0, 1]$.

II. La modélisation de la dépendance par l'utilisation des copules

II.1. Introduction à la théorie des copules

- Définition :

Une copule à deux dimensions est une application C de $[0,1] \times [0,1]$ dans $[0,1]$ satisfaisant :

- (i) $C(u, 0) = C(0, u) = 0$ et $C(u, 1) = C(1, u) = u \quad \forall 0 \leq u \leq 1$
- (ii) C est supermodulaire i.e. $\forall (u_1, u_2)$ et (v_1, v_2) dans $[0,1]^2$ on a:
$$C(v_1, v_2) - C(u_1, v_2) - C(v_1, u_2) + C(u_1, u_2) \geq 0$$

Une copule n'est donc qu'une fonction de répartition bivariée dont les lois marginales sont uniformes.

- Théorème de Sklar :

Soit un couple X dont la fonction de répartition F_X fait partie de $\mathcal{F}(F_1, F_2)$. Il existe alors une copule C telle que pour tout $x \in \mathbb{R}^2 : F_X(x_1, x_2) = C(F_1(x_1), F_2(x_2))$.

De plus, si F_1 et F_2 sont continues, C est unique.

La copule permet de séparer l'étude des lois marginales de la structure de dépendance.

- Densité d'une copule :

La densité d'une copule c si elle existe est définie comme suit :

$$c(u_1, u_2) = \frac{\partial^2}{\partial u_1 \partial u_2} C(u_1, u_2), (u_1, u_2) \in [0,1]^2$$

- Lien avec le tau de Kendall :

Le tau de Kendall peut s'exprimer en fonction d'une copule comme suit :

$$\tau(X_1, X_2) = 4 \iint_0^1 C(u_1, u_2) dC(u_1, u_2) - 1$$

- Lien avec les coefficients de queue :

Soit X_1, X_2 deux variables aléatoires.

$$\lambda_L(X_1, X_2) = \lim_{u \rightarrow 0^+} \frac{C(u, u)}{u} \quad \text{et} \quad \lambda_U(X_1, X_2) = \lim_{u \rightarrow 1^-} \frac{C(u, u) - 2u + 1}{1 - u}$$

II.2. Copules archimédiennes

Nous nous sommes essentiellement intéressés à la famille des copules archimédiennes. Elles ont l'avantage de décrire des structures de dépendance très diverses, notamment les dépendances asymétriques (où les coefficients de queue inférieure et de queue supérieure diffèrent). Elles sont donc particulièrement adaptées en assurance.

Définition :

Soit φ une fonction convexe, continue, strictement décroissante de $[0,1]$ dans $[0,+\infty]$ telle que $\varphi(1) = 0$ et $\varphi(0) = \infty$. Alors $C(u, v) = \varphi^{-1}(\varphi(u) + \varphi(v))$, $(u, v) \in [0,1]^2$, est une copule archimédienne stricte et φ est appelé générateur strict de C .

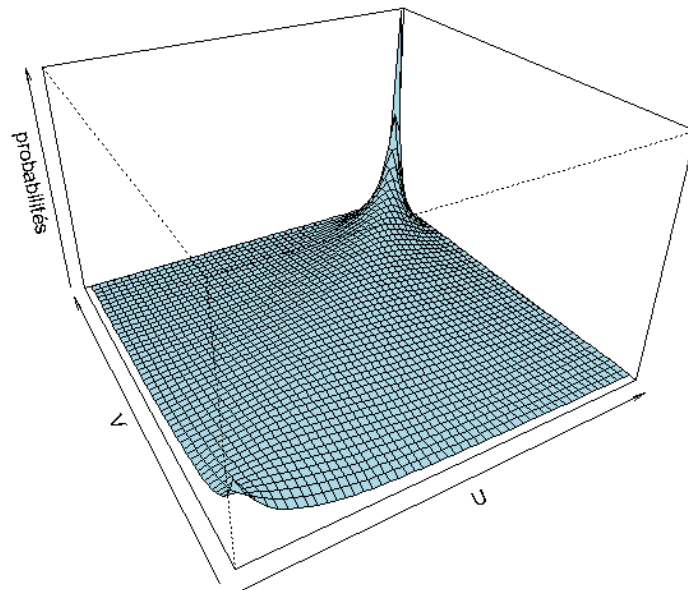
Nous avons retenu trois copules archimédiennes principales : les copules de Gumbel, de Frank et de Clayton. Leurs caractéristiques sont présentées dans le tableau suivant :

	Gumbel	Frank	Clayton
$\varphi(u)$	$(-\ln u)^\theta, \theta \geq 1$	$-\ln\left(\frac{e^{-\theta u} - 1}{e^{-\theta} - 1}\right), \theta \neq 0$	$\frac{u^{-\theta} - 1}{\theta}, \theta > 0$
$C(u, v)$	$\exp\left(-[(-\ln u)^\theta + (-\ln v)^\theta]^{\frac{1}{\theta}}\right)$	$-\frac{1}{\theta} \ln\left[1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}\right]$	$(u^{-\theta} + v^{-\theta} - 1)^{\frac{1}{\theta}}$
λ_L	0	0	$2^{-1/\theta}$
λ_U	$2 - 2^{1/\theta}$	0	0
τ_θ	$1 - 1/\theta$	$1 - \frac{4(1 - D_1(\theta))}{\theta}$	$\frac{\theta}{\theta + 2}$

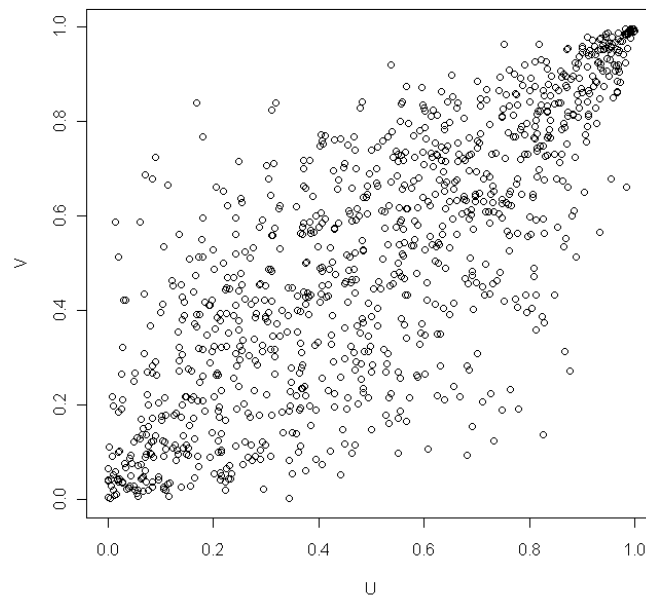
La **copule de Gumbel** n'appréhende que des dépendances positives, accentuées sur la queue supérieure. Elle permet donc d'étudier l'impact de la survenance d'évènements de forte intensité sur la dépendance entre branches.

Densité et dépendogramme pour une copule de Gumbel de paramètre $\theta = 2.5$, $\tau = 0.6$

Densité d'une Copule de Gumbel



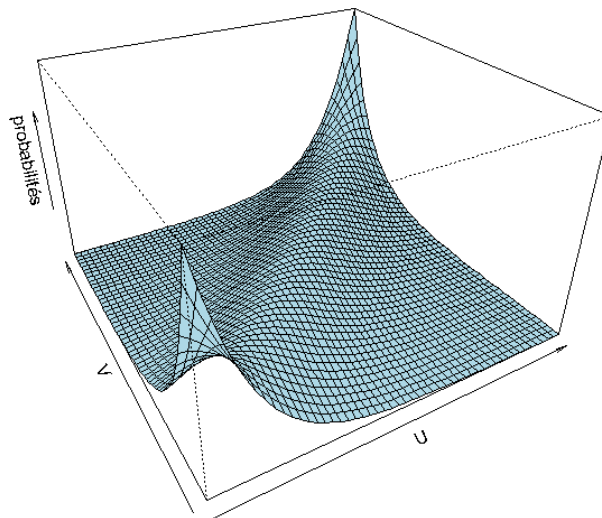
Dépendogramme, Copule de Gumbel



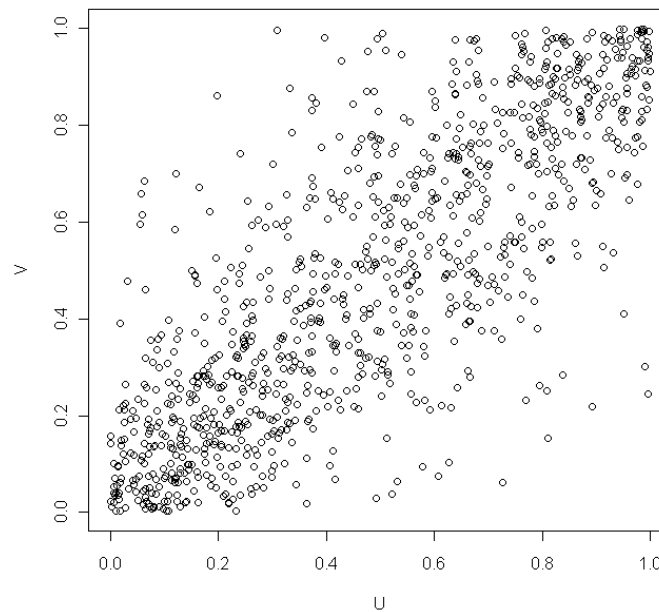
La **copule de Frank** permet de modéliser les dépendances positives et négatives. La densité est plus élevée aux extrêmes sans accentuation sur l'une des queues. Le tau de Kendall s'exprime en fonction de θ à l'aide de la fonction Debye : $D_k(x) = \frac{k}{x^k} \int_0^x \frac{t^k}{e^t - 1} dt$.

Densité et dépendogramme pour une copule de Frank de paramètre $\theta = 7.93$, $\tau = 0.6$

Densité d'une Copule de Frank



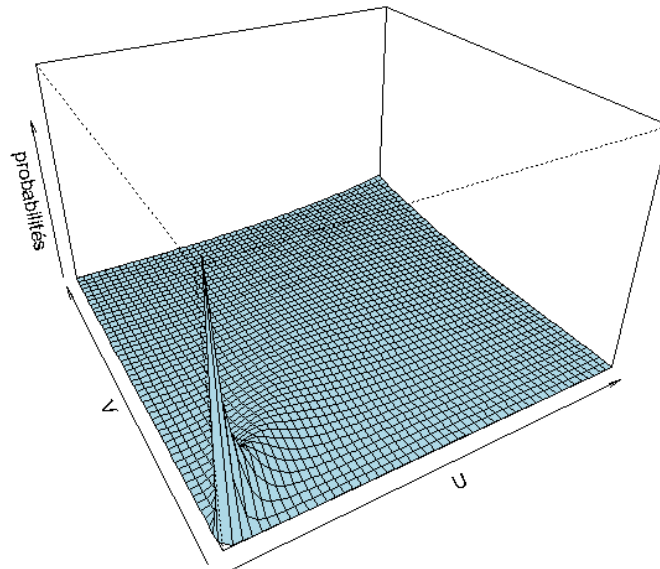
Dépendogramme, Copule de Frank



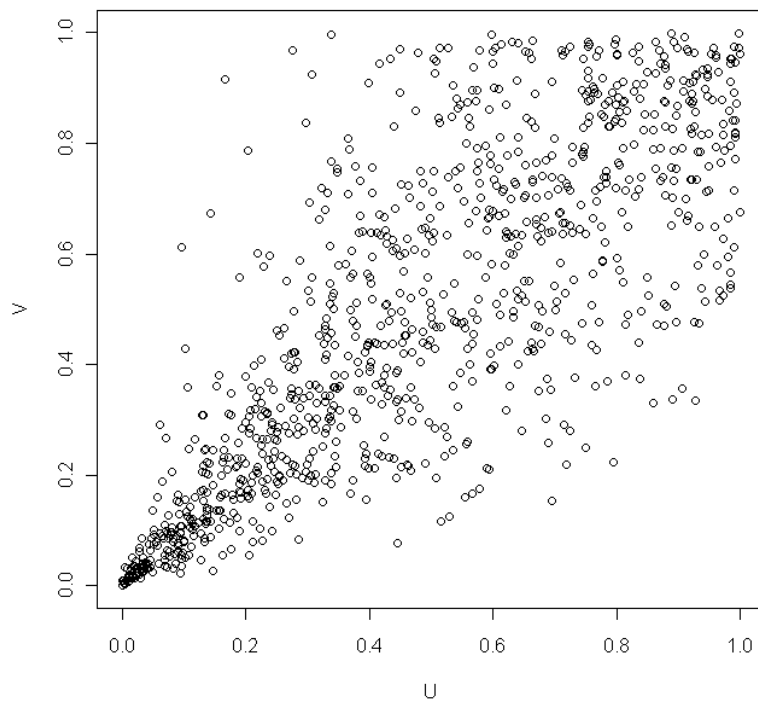
La **copule de Clayton** ne permet de modéliser que les dépendances positives, accentuées sur la queue inférieure, contrairement à la copule de Gumbel.

Densité et dépendogramme pour une copule de Clayton de paramètre $\theta = 3$, $\tau = 0.6$

Densité d'une Copule de Clayton



Dépendogramme, Copule de Clayton



II.3. Estimation du paramètre et choix de la copule

Il existe plusieurs méthodes permettant d'estimer le(s) paramètre(s) des copules, la plupart étant basées sur la méthode du maximum de vraisemblance. Nous avons privilégié la méthode CML (Canonical Maximum Likelihood) car elle a l'avantage de déterminer le paramètre de la copule indépendamment de la forme paramétrique des lois marginales.

En effet, les observations $\{x_1^t, \dots, x_n^t\}_{t=1}^T$ sont transformées en uniformes empiriques $\{u_1^t, \dots, u_n^t\}_{t=1}^T$ où $u_i^t = \frac{\text{rang}(x_i^t)}{T+1}$, pour $t=1, \dots, T$ et pour tout i allant de 1 à n . Avec les copules, nous ne travaillons donc plus en montants mais en rangs.

Le paramètre est alors exprimé comme suit : $\hat{\theta} = \underset{\theta}{\text{argmax}} \sum_{t=1}^T \ln c(u_1^t, \dots, u_n^t; \theta)$.

Cette méthode a également l'avantage d'engendrer des temps de calcul limités. Pour accélérer la maximisation du programme, nous pouvons donner une valeur initiale à θ à l'aide de la relation liant le paramètre de la copule au « tau » de Kendall.

La copule est ensuite choisie en maximisant la vraisemblance.

II.4. Détermination des Best Estimates corrélés

- Simulation d'une variable aléatoire uniforme sur $[0,1]$ = u et d'une variable aléatoire uniforme sur $[0,1]$, v , indépendante de u
- Simulation d'une variable aléatoire uniforme sur $[0,1]$, p , indépendante de u et v
- Détermination de la valeur v_d telle que $C(v_d|U = u) = \frac{\partial C(u, v_d)}{\partial u} = p$
- Obtention d'une simulation $P_i^{1,k}$, élément de la diagonale du premier triangle en résolvant $u = F_{1,i}(P_i^{1,k})$, $F_{1,i}$ étant la loi de distribution suivie par l'élément P_i^1 .
- Obtention d'une simulation de $P_i^{2,k}$, élément de la diagonale du deuxième triangle en résolvant $v_d = F_{2,i}(P_i^{2,k})$, $F_{2,i}$ étant la loi de distribution suivie par l'élément P_i^2 .

On obtient donc 10 000 simulations pour chaque élément de chaque triangle liées par la copule choisie.

- Pour k allant de 1 à 10 000, intégration des diagonales $P^{1,k}$ et $P^{2,k}$ dans leur triangle respectif. Obtention de 10 000 nouveaux « pseudo-trapèzes » corrélés.
- Application d'une modélisation GLM sur chaque pseudo-trapèze et calcul des provisions. On obtient alors 10 000 couples de provisions (BE_1^1, BE_1^2) corrélés.

III. Application aux données

III.1. Recherche de dépendance entre branches

Nous cherchons à estimer le degré de dépendance entre branches. Comme il existe un lien entre le paramètre des copules et le tau de Kendall, nous allons commencer par estimer les coefficients de corrélation de Kendall sur les triangles de règlements.

Matrice de Kendall sur triangles de règlements :

	Dom.Auto	Dom.biens	RC Auto	RC Divers
Dom.Auto	1	0,825	0,754	0,610
Dom.biens	0,825	1	0,725	0,630
RC Auto	0,754	0,725	1	0,667
RC Divers	0,610	0,630	0,667	1

Ces corrélations mettent toutes en évidence un fort lien positif entre les différentes branches, particulièrement entre les branches Dommages Auto et Dommages aux Biens. Toutefois, l'importance des valeurs des coefficients de corrélation vient de la liquidation. En effet, quelque soit la branche considérée, les montants de la première année de développement sont toujours plus élevés que ceux de la sixième. Ces valeurs ne nous renseignent donc pas sur le lien existant entre deux montants de la même année de branches différentes.

Pour supprimer l'effet liquidation, nous allons calculer les coefficients de corrélation sur les résidus des modèles Log Poisson surdispersé et Log Gamma.

Matrice de Kendall sur les résidus du modèle Log Poisson surdispersé:

	Dom.Auto	Dom.biens	RC Auto	RC Divers
Dom.Auto	1	0,415	0,144	-0,115
Dom.biens	0,415	1	0,073	0,093
RC Auto	0,144	0,073	1	0,089
RC Divers	-0,115	0,093	0,089	1

Les coefficients de corrélation sont alors beaucoup plus faibles. Il apparaît que seule la dépendance entre Dommages Auto et Dommages aux Biens est significative. La tempête de 1999 pourrait être la cause de cette corrélation.

Matrice de Kendall sur les résidus du modèle Log Gamma:

	Dom.Auto	Dom.biens	RC Auto	RC Divers
Dom.Auto	1,000	0,089	-0,006	-0,273
Dom.biens	0,089	1,000	-0,119	-0,055
RC Auto	-0,006	-0,119	1,000	0,091
RC Divers	-0,273	-0,055	0,091	1,000

Ici encore les valeurs des coefficients de corrélation sont très faibles, et laissent apparaître des dépendances négatives.

Il semblerait que le nombre d'observations ne soit pas suffisant pour obtenir des résultats acceptables. Néanmoins, nous allons étudier le cas de la dépendance entre Dommages Auto et Dommages aux Biens par le modèle Log Poisson surdispersé.

III.2. Estimation du paramètre et choix de la copule

copule	Clayton	Gumbel	Frank
paramètre	1,055	1,775	4,25
vraisemblance	8,969	12,95	10,01

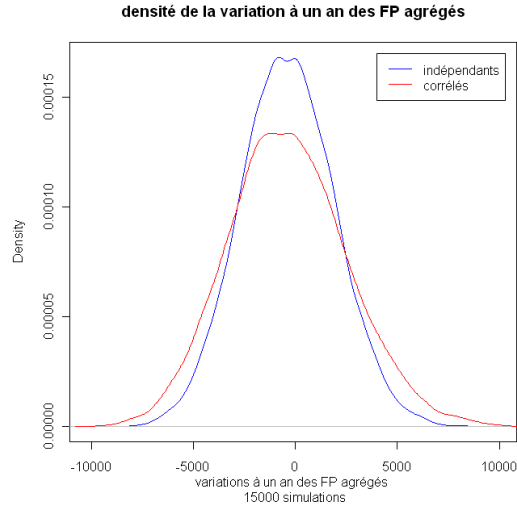
La copule de vraisemblance maximale est la copule de Gumbel de paramètre 1,775. Nous allons donc utiliser cette copule pour modéliser la dépendance entre les deux branches.

III.3. Application au calcul du SCR agrégé des branches Dom.Auto et Dom.Biens

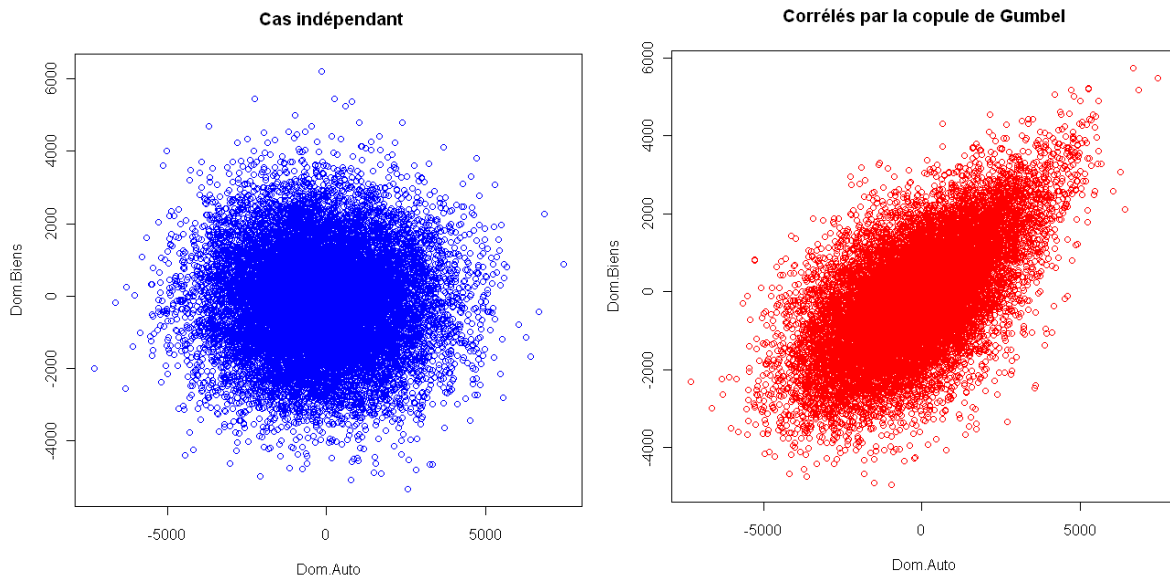
(pour 15000 simulations)

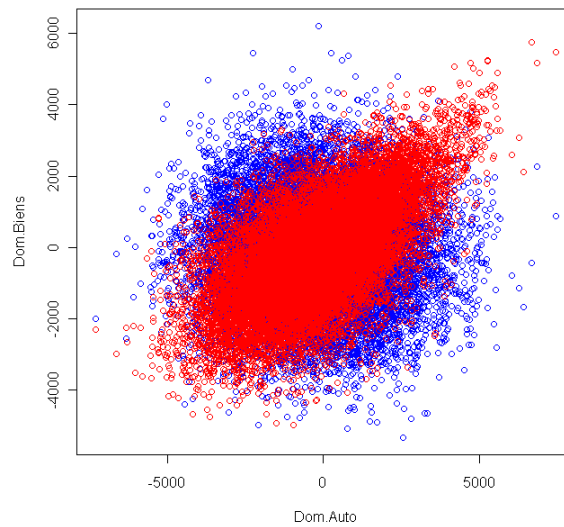
	Cas indépendant	Corrélation par la copule de Gumbel	Formule standard pour deux branches
SCR agrégé	5760	8011	9456

La prise en compte de la dépendance entre les deux branches augmente de manière significative le montant du SCR. Nous constatons une augmentation de 39% dans le cas corrélé par la copule. Le graphique suivant montre en effet une queue plus épaisse dans le cas dépendant.



Par ailleurs, la prise en compte de la dépendance augmente la variabilité du SCR agrégé. En effet, dans le cas indépendant, elle est de 6.30% alors que dans le cas corrélé, elle est de 8,02%. Nous constatons également que le nuage de points, qui était circulaire dans le cas indépendant, s'allonge en suivant une pente positive. La diffusion est accentuée dans le cadre supérieur droit, indiquant une dépendance importante entre les gros sinistres.



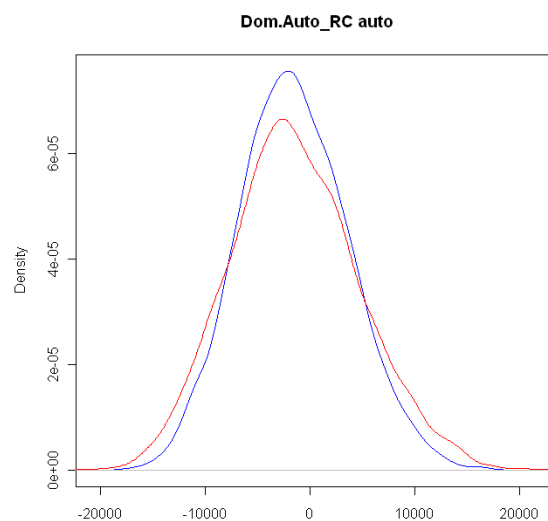


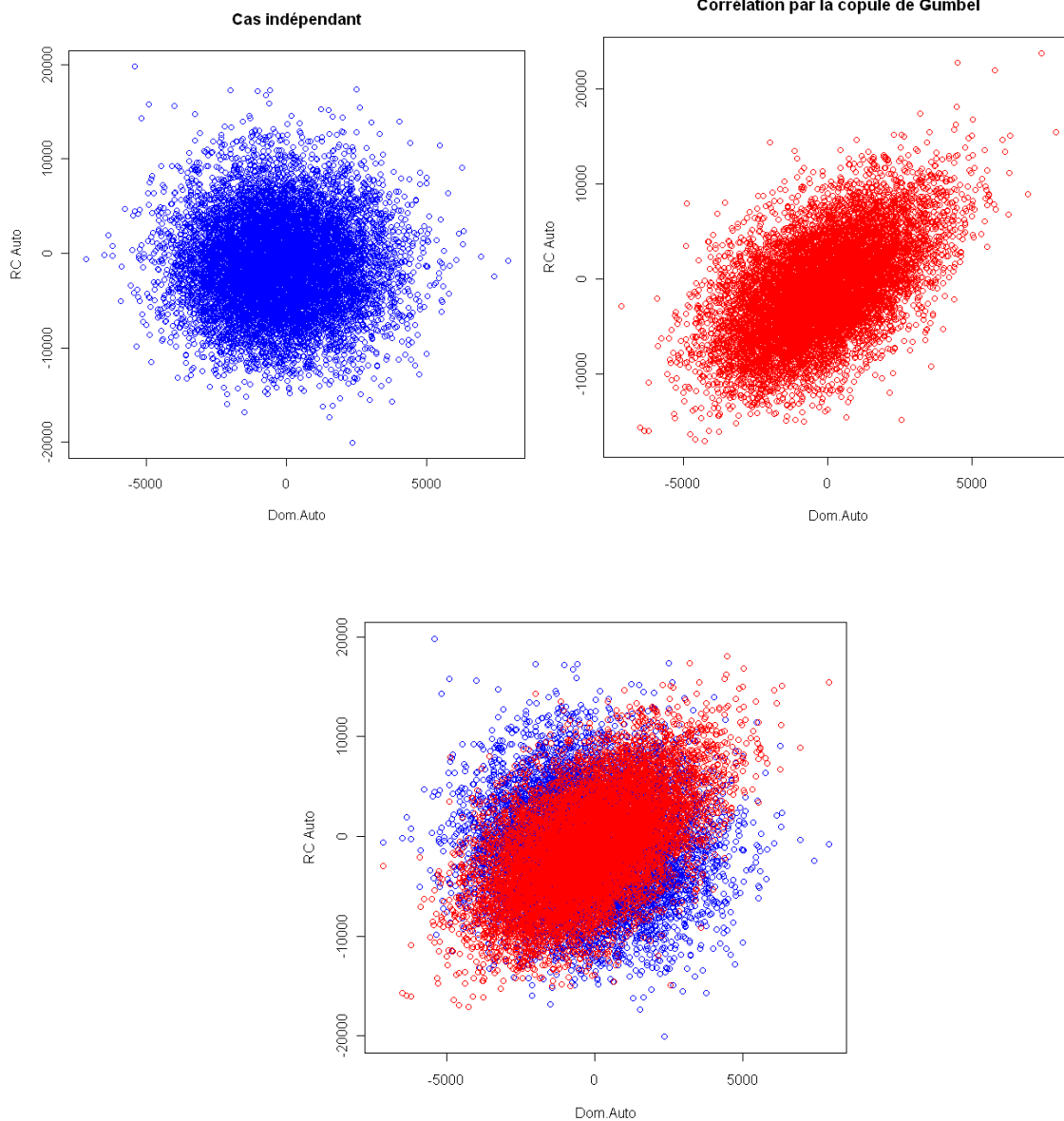
III.4. Application au calcul du SCR agrégé des branches Dom.Auto et RC Auto

(pour 15000 simulations)

Les matrices de corrélation de Kendall sur les triangles de paiements et les résidus de Poisson laissent tout de même apparaître un lien entre les Branches Dommages Auto et RC Auto. À titre d'exemple, nous avons décidé d'utiliser le coefficient de corrélation donné dans le QIS5 pour ces branches.

Si nous supposons un « tau » de Kendall de 0.5, le paramètre de la copule de Gumbel est égal à 2. Nous avons obtenus les résultats suivants :





Les résultats sont moins significatifs que pour les branches Dommages Auto et Dommages aux Biens. L'utilisation de la matrice de corrélation des coefficients de Kendall sur les résidus semble donc être un bon moyen de recherche de dépendance. Toutefois, son utilisation sous-estime les résultats par rapport à la formule standard.

	Cas indépendant	Corrélation par la copule de Gumbel	Formule standard pour deux branches
SCR agrégé	12648	15019	23407

CONCLUSION

Concernant les méthodes de provisionnement, nous ne pouvons pas dire que l'utilisation d'une méthode est préférable à toute autre dans tous les cas. Le choix d'une méthode dépend avant tout de son adéquation avec les données. Il ressort toutefois que la régression Log Normale n'est pas la méthode à privilégier. Elle tend à surestimer les provisions et semble très sensible aux faibles valeurs qui sont présentes dans les triangles à développement court. Elle peut néanmoins être utilisée à titre indicatif, principalement pour les branches à développement long. La méthode Chain-Ladder reste toujours la méthode la plus simple à mettre en œuvre et son utilisation dans la méthode bootstrap permet de diminuer le temps de calcul. Nous avons cependant préféré l'utilisation des modèles GLM, en particulier le modèle Log Poisson Surdispersé. Il semble s'adapter à tout type de branche et donne des erreurs de prédiction relativement faibles. Le modèle Log Gamma est quant à lui un bon outil de comparaison, car également applicable dans la méthode bootstrap. Par ailleurs, nous avons décidé de gérer la liquidation incomplète des triangles. Les résultats donnés par l'extrapolation du paramètre année de développement des modèles GLM nous ont semblé plus satisfaisants que ceux issus de l'extrapolation des coefficients de Chain-Ladder. Nous avons néanmoins fait des hypothèses fortes concernant l'arrêt de l'extrapolation. Il serait intéressant d'étudier l'impact que pourrait engendrer leur modification.

Pour l'évaluation du risque à un an, nous avons choisi d'adapter la méthode du bootstrap. Facile à mettre en œuvre et utilisant des simulations de Monte-Carlo, elle permet d'estimer la distribution des provisions, et par conséquent d'obtenir un montant du SCR. Pour les risques courts, elle donne des résultats plus faibles que la formule standard dans le cas Log Poisson et plus élevés dans le cas Log Gamma. En revanche, ils sont sensiblement équivalents pour les risques longs. Les recours n'ayant pas été pris en compte pour l'évaluation du risque à un an, il serait intéressant d'intégrer notre méthode de calcul de Best Estimate net de recours dans le calcul du SCR.

En ce qui concerne la dépendance entre branches, nous avons préféré l'utilisation des copules à d'autres modèles existants, comme le modèle à choc commun ou le bootstrap. La copule est en effet un outil flexible et qui représente la structure de dépendance des données indépendamment des lois marginales. Le faible nombre d'observations ne nous a permis de calibrer correctement une copule que dans un cas, la dépendance entre les branches Dommages Auto et Dommages aux Biens. À titre d'exemple, nous avons également étudié la dépendance entre les branches Dommages Auto et RC Auto en utilisant le coefficient de corrélation donné dans le QIS 5. Les résultats confirment l'importance de la prise en compte de la dépendance dans le calcul du SCR. Négliger la dépendance revient à sous-estimer le montant du SCR et prendre le risque de ne pas être couverts correctement en cas de situations défavorables. Enfin, nous nous sommes limités au cas bivarié. Une prochaine étude pourrait être menée pour traiter les cas multivariés.

ANNEXES

L'année 0 correspond à l'année 1997

A. Branche Dommages Auto :

Triangle de règlements décumulé :

	dvp 0	dvp 1	dvp 2	dvp 3	dvp 4	dvp 5	dvp 6	dvp 7	dvp 8	dvp 9
surv 0	89037	14262	830	458	311	321	47	24	10	2
surv 1	87077	15814	895	522	349	390	100	29	12	
surv 2	83689	23776	890	202	14	255	58	30		
surv 3	82186	19173	964	362	365	5	12			
surv 4	85840	22316	753	385	427	146				
surv 5	91965	17885	1146	533	248					
surv 6	84282	15630	1220	636						
surv 7	78270	15775	1538							
surv 8	71708	16077								
surv 9	76604									

I. Chain-ladder :

Coefficients de développement :

j	0	1	2	3	4	5	6	7	8
\hat{f}_j	1,2131	1,01	1,0042	1,0027	1,0021	1,0005	1,0003	1,0001	1

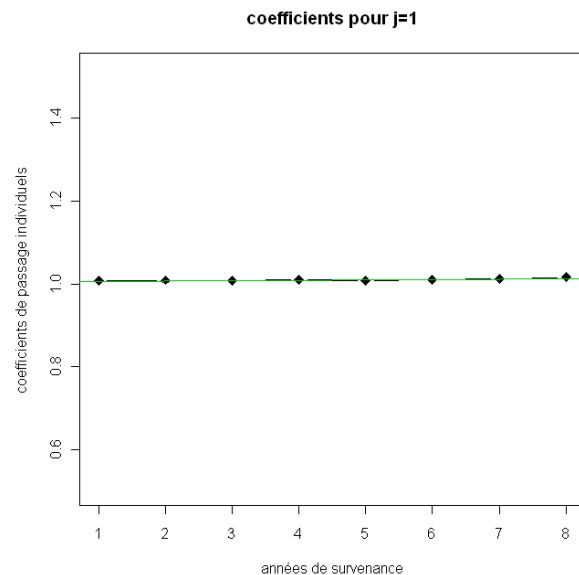
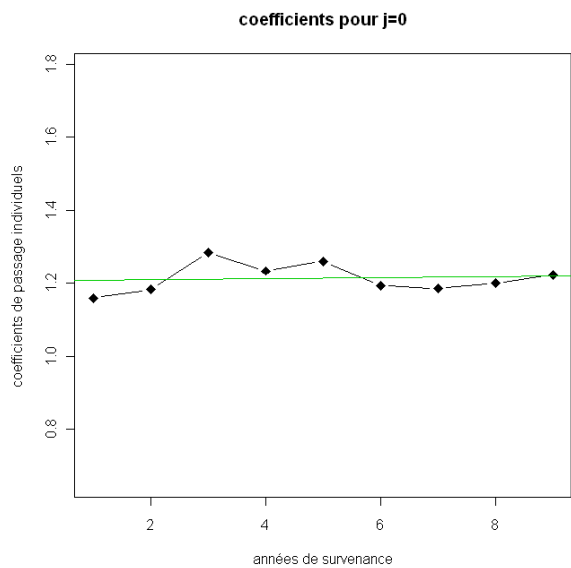
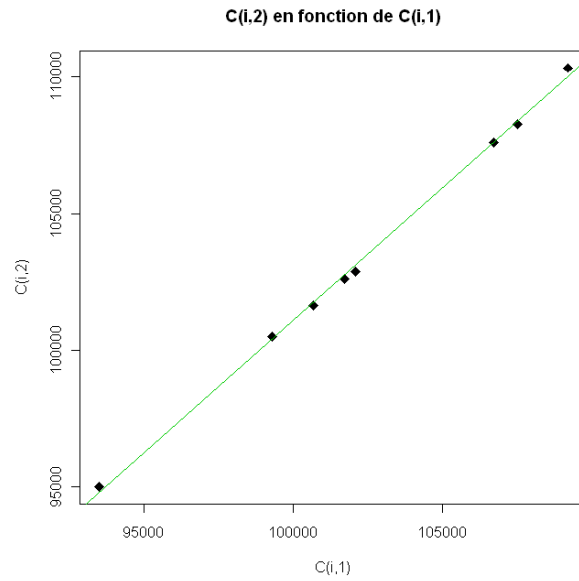
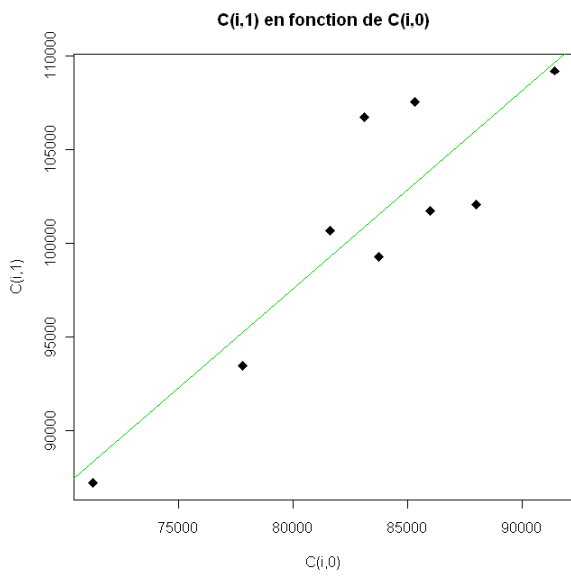
cadences	80,819%	98,043%	99,020%	99,435%	99,701%	99,910%	99,962%	99,988%	99,998%
-----------------	---------	---------	---------	---------	---------	---------	---------	---------	---------

Provisions par année de survenance :

i	\hat{R}_i
0	0
1	2
2	13
3	40
4	99
5	336
6	579
7	946
8	1752
9	18181

Provisions totales: 21 948

Validation graphique des hypothèses :



Les conditions ne sont pas vraiment vérifiées pour l'année de développement 0, toutefois elles le sont à partir de l'année de développement 1.

II. Mack

Estimation des $\hat{\sigma}_j^2$:

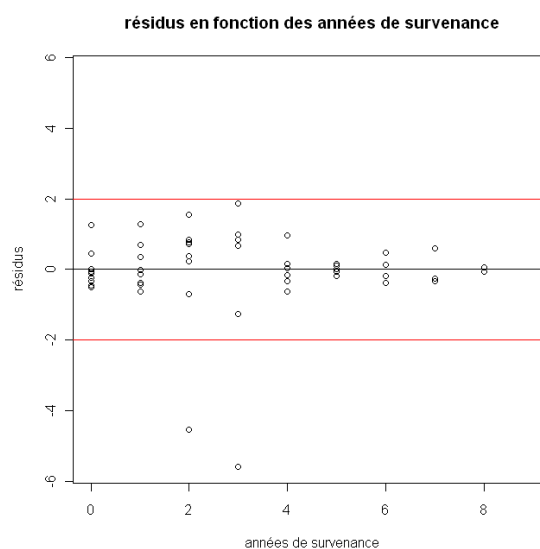
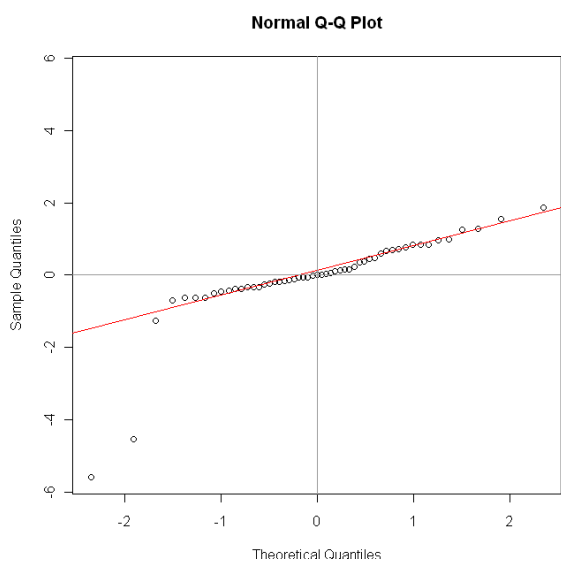
j	0	1	2	3	4	5	6	7	8
$\hat{\sigma}_j^2$	136.91	0,888	0,207	0,206	0,220	0,012	8.02^{-05}	$1,92e^{-05}$	$4.61e^{-06}$

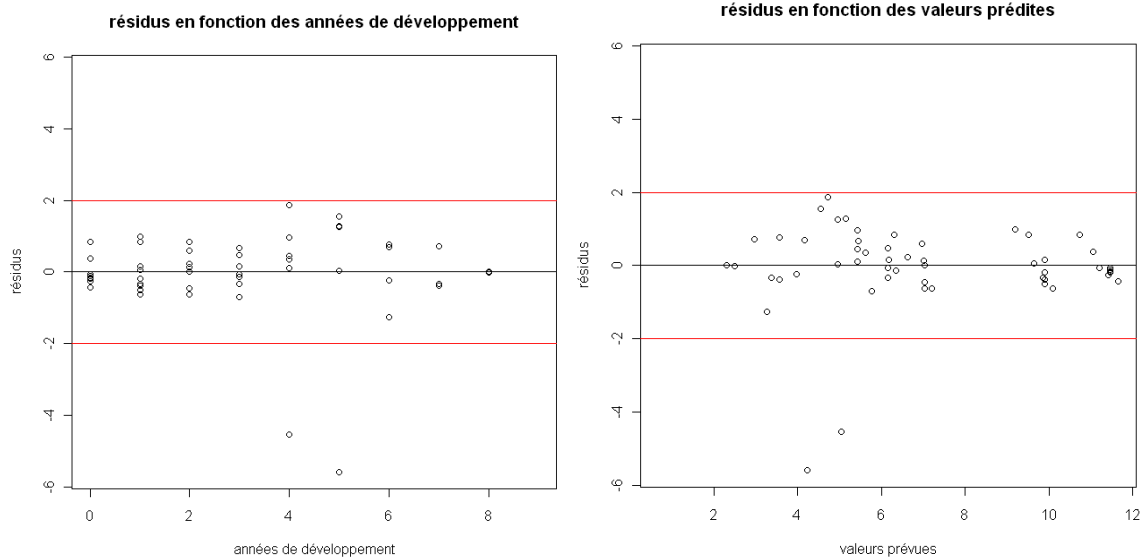
i	\hat{R}_i	$\widehat{MSEP}(\hat{R}_i)$	$\sqrt{\widehat{MSEP}(\hat{R}_i)/\hat{R}_i}$
1	2	0,97018	49,25%
2	13	4,2013	15,77%
3	40	14,828	9,63%
4	99	1722,9	41,93%
5	336	31548	52,86%
6	579	52750	39,67%
7	946	71905	28,35%
8	1752	154000	22,40%
9	18181	12183000	19,20%
total	21948	12652352	16,21%

III. Modèles stochastiques

1. Régression log normale

test	p-value	acceptation/rejet (seuil à 5%)
indépendance	0,637	on accepte
normalité	0,0415	on rejette





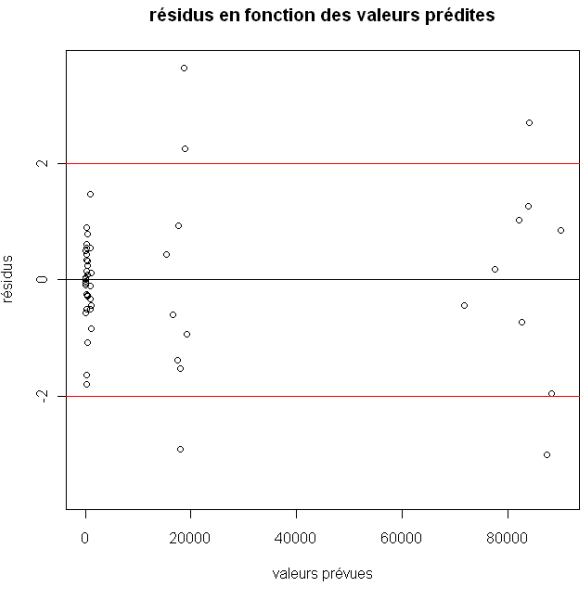
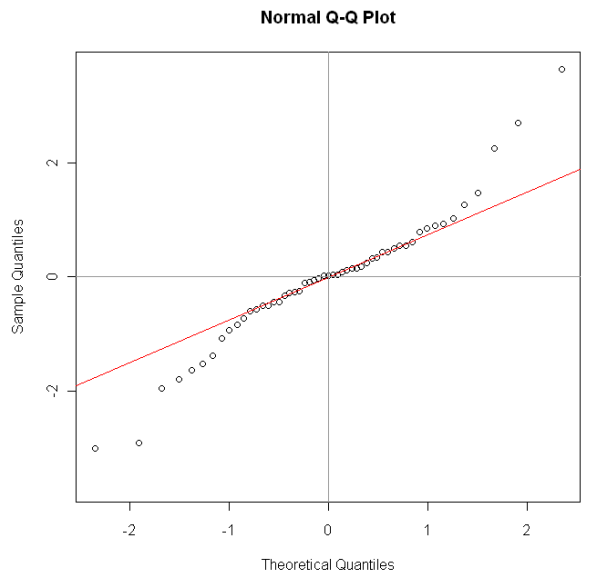
On rejette l'hypothèse de normalité. La présence de valeurs atypiques entraîne un mauvais ajustement du modèle.

Valeurs atypiques :
(2,4) et (3,5)

2. Modèle Log Poisson Surdispersé

test	p-value	acceptation/rejet (seuil à 5%)
indépendance	0,157	on accepte
normalité	0,807	on accepte

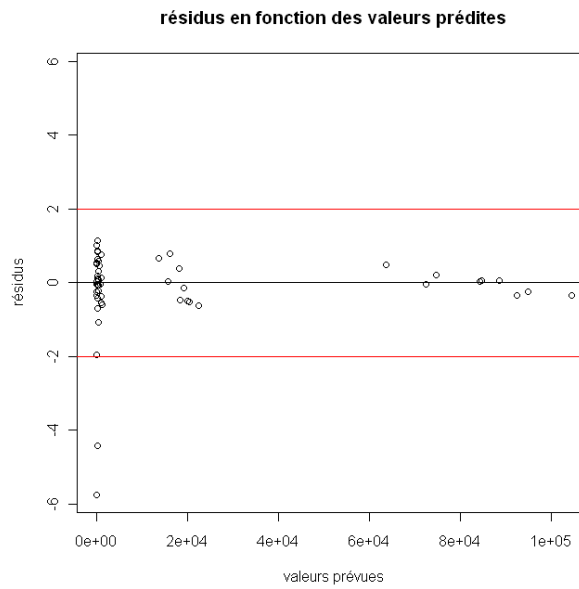
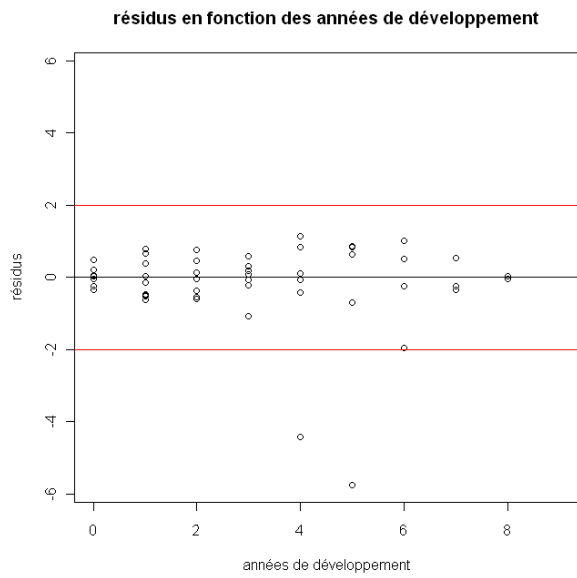
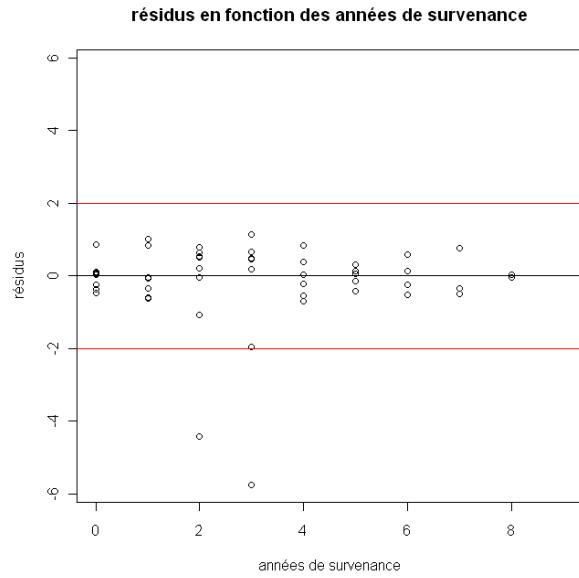
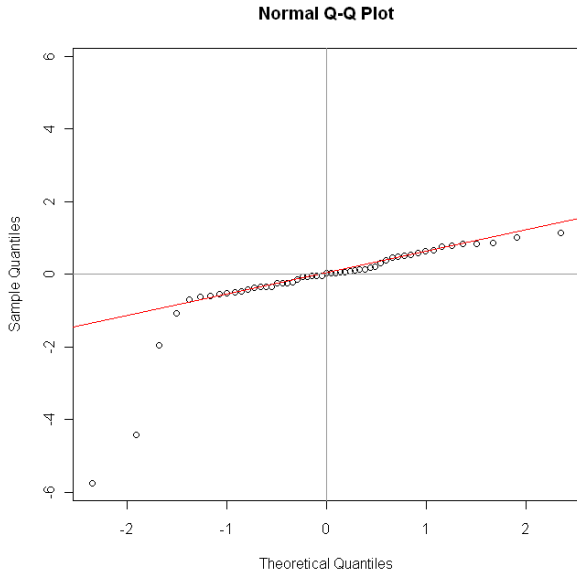
Le test accepte la normalité des résidus. Néanmoins, le graphique QQ plot pourrait la remettre en doute.



Valeurs atypiques :
 (0,0), (0,1), (2,0), (2,1), (4,1)

3. Modèle Log Gamma

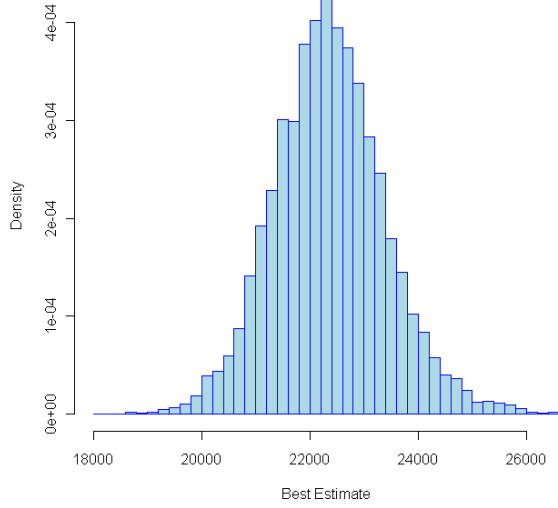
test	p-value	acceptation/rejet (seuil à 5%)
indépendance	0,157	on accepte
normalité	0,0898	on accepte



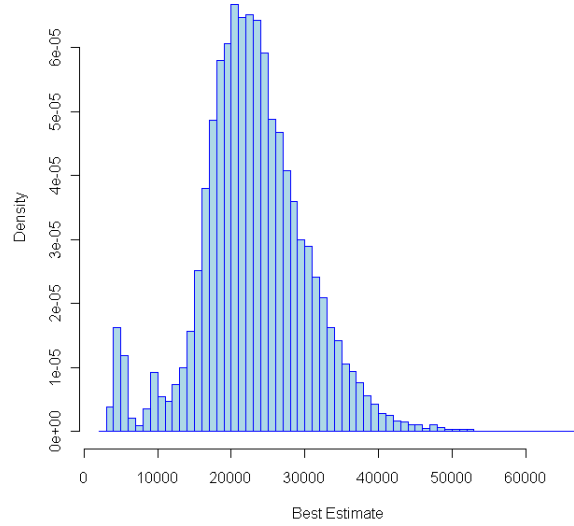
Valeurs atypiques :
(2,4) et (3,5)

IV. Bootstrap

**densité des provisions
modèle log poisson surdis. par bootstrap**

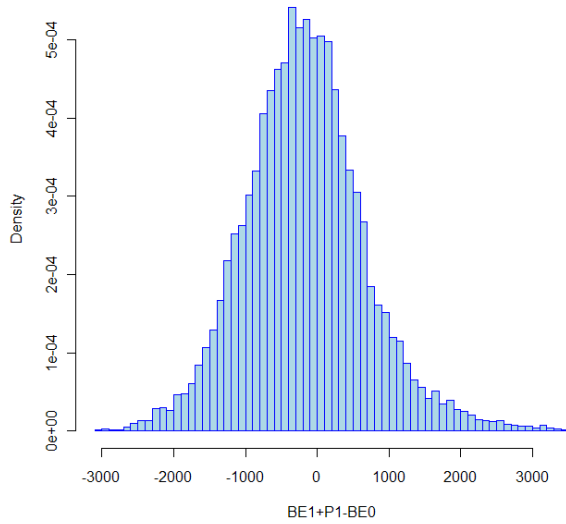


**densité des provisions
modèle log gamma par bootstrap**

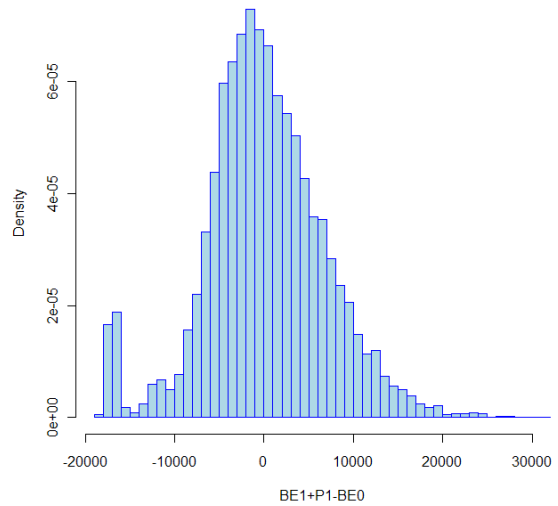


V. Bootstrap à un an

**densité de BE1+P1-BE0
modèle log poisson surdis. par bootstrap**



**densité de BE1+P1-BE0
modèle log gamma par bootstrap**



B. Branche Dommages aux Biens :

Triangle de règlements décumulé :

	dvp 0	dvp 1	dvp 2	dvp 3	dvp 4	dvp 5	dvp 6	dvp 7	dvp 8	dvp 9
surv 0	27757	10757	721	424	302	108	22	151	17	6
surv 1	23061	8666	435	367	220	116	114	4	3	
surv 2	21184	10586	934	403	329	189	120	30		
surv 3	18169	11470	677	390	255	72	97			
surv 4	22071	14993	981	270	92	168				
surv 5	25343	10576	844	577	122					
surv 6	28854	11612	1110	429						
surv 7	22675	12528	1324							
surv 8	21079	11791								
surv 9	25653									

I. Chain-ladder :

Coefficients de développement :

j	0	1	2	3	4	5	6	7	8
\hat{f}_j	1,4899	1,0251	1,0114	1,0062	1,0037	1,0026	1,0017	1,0003	1,0002

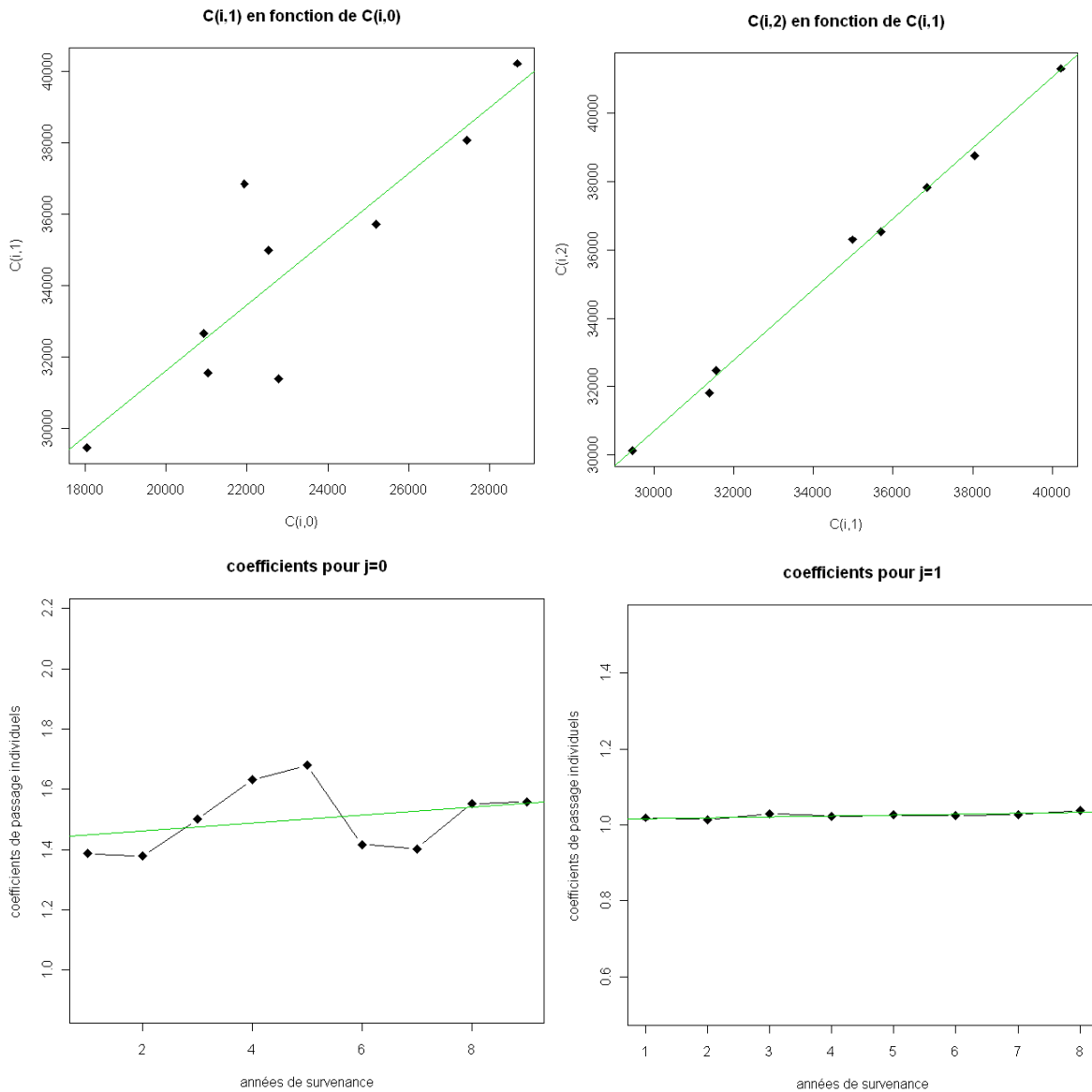
cadences	63,797%	95,052%	97,435%	98,546%	99,161%	99,530%	99,785%	99,958%	99,985%
-----------------	---------	---------	---------	---------	---------	---------	---------	---------	---------

Provisions par année de survenance :

i	\hat{R}_i
0	0
1	5
2	14
3	67
4	182
5	317
6	620
7	962
8	1711
9	14558

Provisions totales: 18436

Validation graphique des hypothèses :



Les conditions ne sont pas vraiment vérifiées pour l'année de développement 0, toutefois elles le sont à partir de l'année de développement 1.

II. Mack

Estimation des $\hat{\sigma}_j^2$:

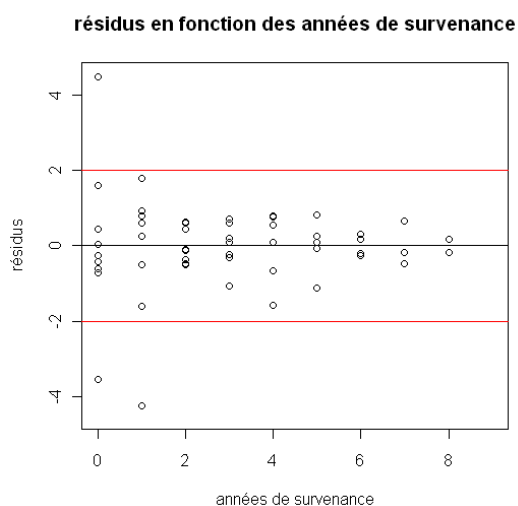
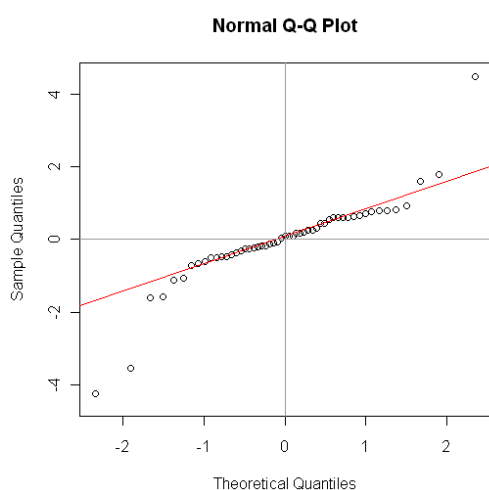
j	0	1	2	3	4	5	6	7	8
$\hat{\sigma}_j^2$	286.63	1.759	0,256	0,312	0,061	0,078	0.138	0.002	$2.88 e^{-05}$

i	\hat{R}_i	$\widehat{MSEP}(\hat{R}_i)$	$\sqrt{\widehat{MSEP}(\hat{R}_i)/\hat{R}_i}$
1	5	1,7286	26,30%
2	14	100,12	71,47%
3	67	5632,2	112,01%
4	182	11240	58,25%
5	317	13685	36,90%
6	620	31893	28,80%
7	962	38387	20,37%
8	1711	102980	18,76%
9	14558	9253700	20,90%
total	18436	9622128	16,83%

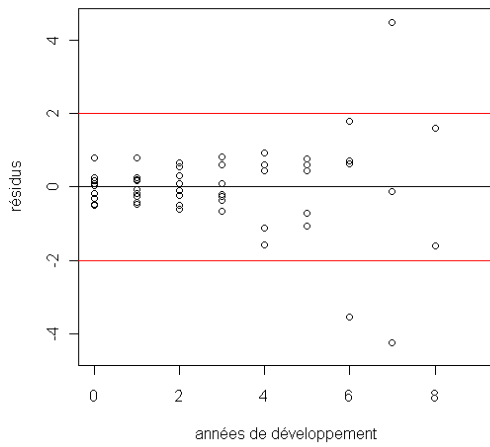
III. Modèles stochastiques

1. Régression log normale

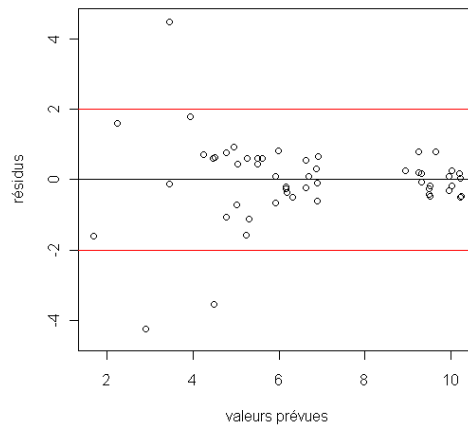
test	p-value	acceptation/rejet (seuil à 5%)
indépendance	0,637	on accepte
normalité	0,269	on accepte



résidus en fonction des années de développement



résidus en fonction des valeurs prédites

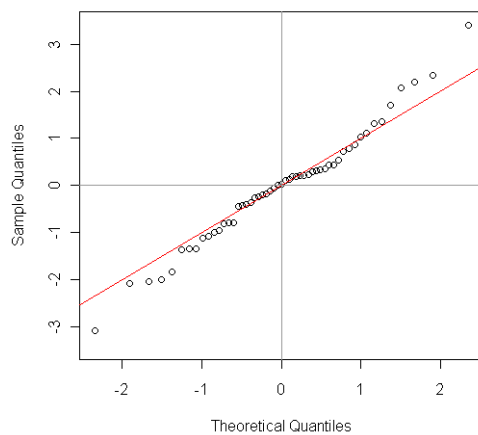


Valeurs atypiques :
(0,6), (0,7), (1,7)

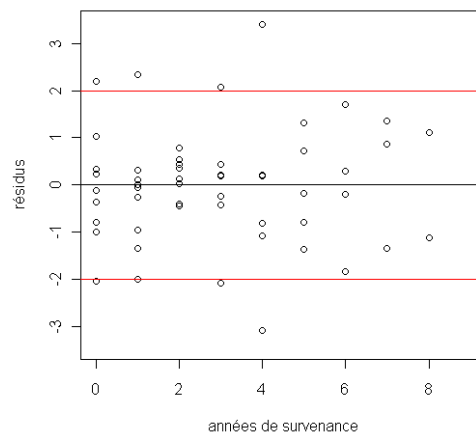
2. Modèle Log Poisson Surdispersé

test	p-value	acceptation/rejet (seuil à 5%)
indépendance	0,637	on accepte
normalité	0,779	on accepte

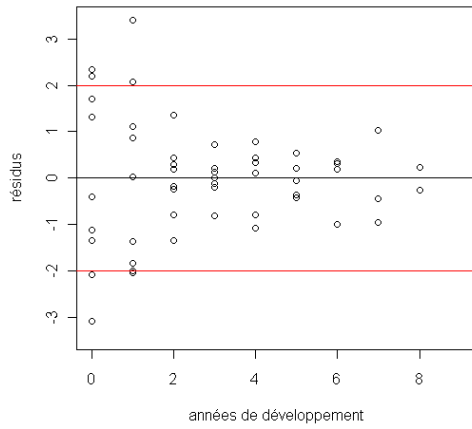
Normal Q-Q Plot



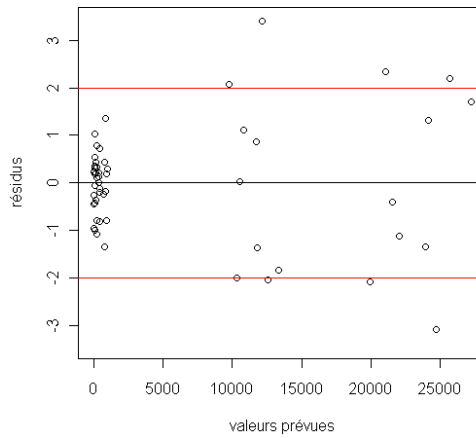
résidus en fonction des années de survéance



résidus en fonction des années de développement



résidus en fonction des valeurs prédites

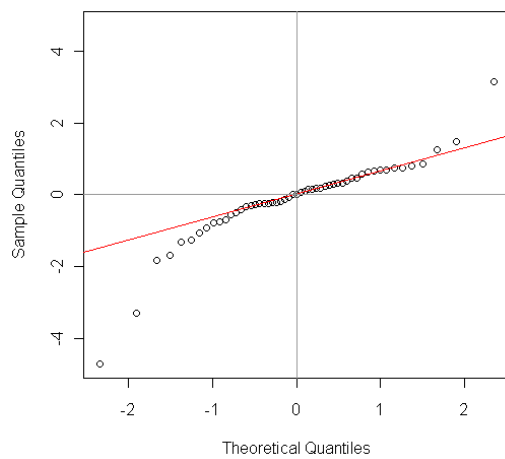


Valeurs atypiques :
 année de survenance : 0, 1, 3, 4
 année de développement : 0, 1

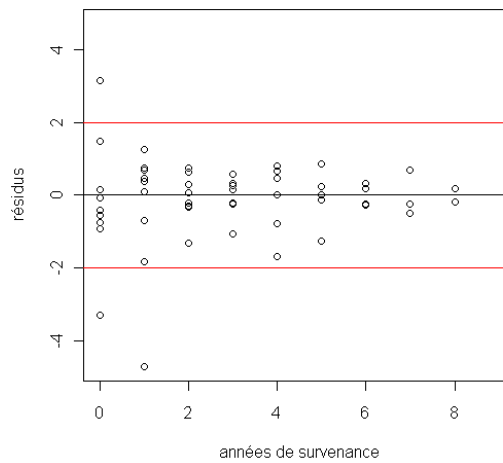
3. Modèle Log Gamma

test	p-value	acceptation/rejet (seuil à 5%)
indépendance	0,637	on accepte
normalité	0,241	on accepte

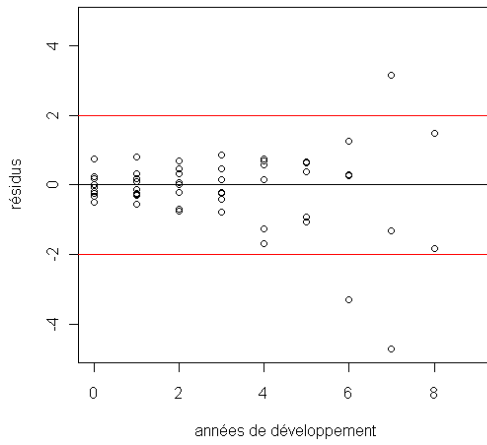
Normal Q-Q Plot



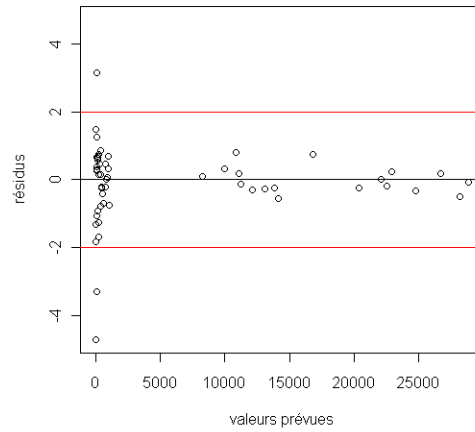
résidus en fonction des années de survenance



résidus en fonction des années de développement



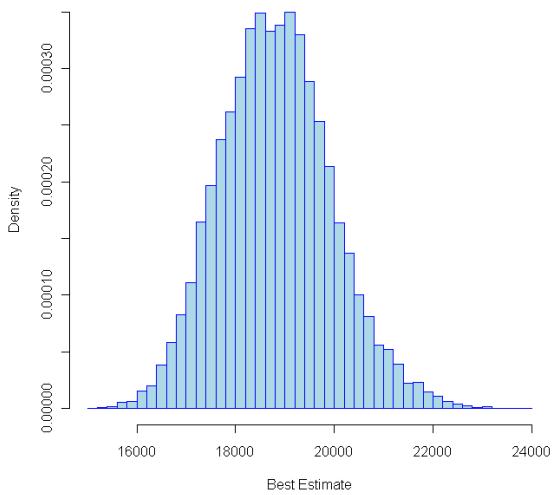
résidus en fonction des valeurs prédites



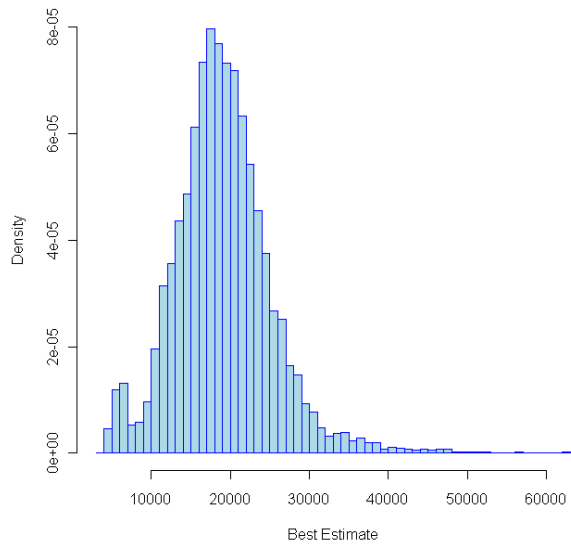
Valeurs atypiques :
(0,6), (0,7), (1,7)

IV. Bootstrap

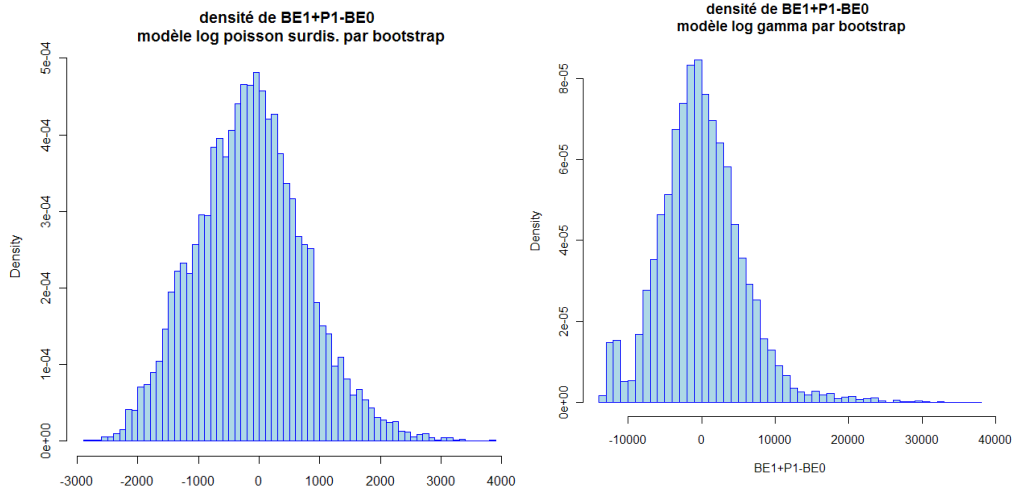
**densité des provisions
modèle log poisson surdis. par bootstrap**



**densité des provisions
modèle log gamma par bootstrap**



V. Bootstrap à un an



C. Branche RC Auto :

Triangle de règlements décumulé :

	dvp 0	dvp 1	dvp 2	dvp 3	dvp 4	dvp 5	dvp 6	dvp 7	dvp 8	dvp 9
surv 0	22456	22351	6474	4391	1857	3056	249	89	423	804
surv 1	22954	23416	5798	3094	1844	891	588	931	336	
surv 2	21312	21039	7058	2915	2105	2790	903	357		
surv 3	20626	24529	6878	3388	1960	564	1986			
surv 4	21037	23173	6665	4027	3579	1160				
surv 5	22708	20480	7375	3549	3606					
surv 6	18416	17896	5732	2228						
surv 7	19479	17416	5804							
surv 8	18006	17049								
surv 9	17559									

I. Chain-ladder :

Coefficients de développement :

j	0	1	2	3	4	5	6	7	8
\hat{f}_j	2,0019	1,1526	1,0677	1,0456	1,0297	1,0159	1,0078	1,0063	1,0131

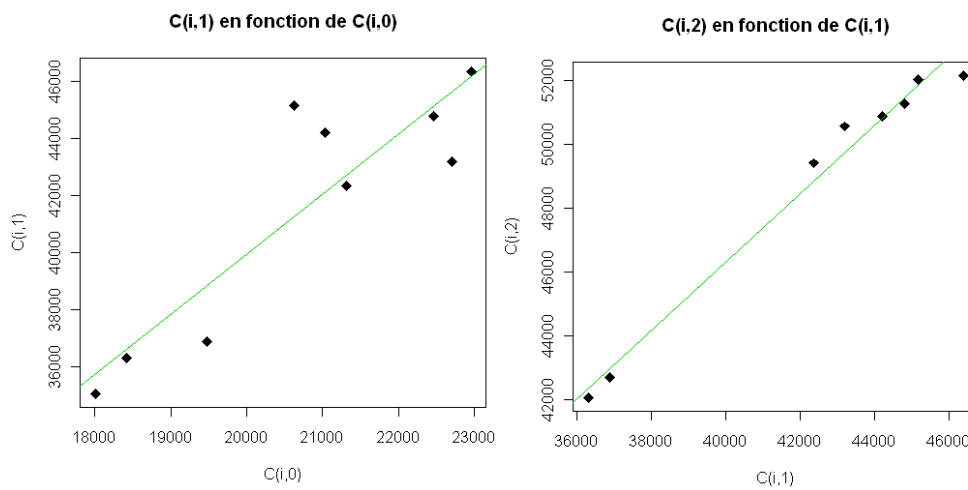
cadences	36,118%	72,304%	83,339%	88,983%	93,043%	95,806%	97,333%	98,088%	98,706%
-----------------	---------	---------	---------	---------	---------	---------	---------	---------	---------

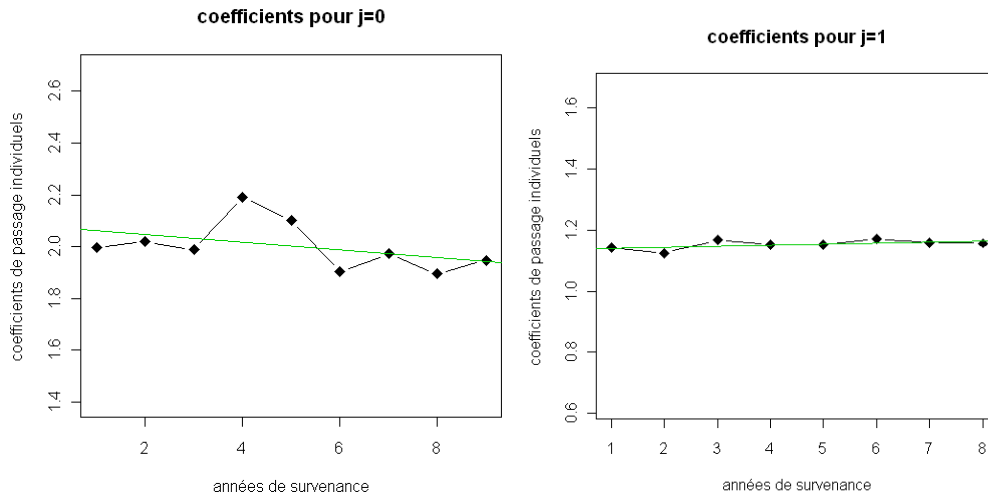
Provisions par année de survenance :

i	\hat{R}_i
0	0
1	784
2	1140
3	1642
4	2611
5	4316
6	5481
7	8536
8	13428
9	31057

Provisions totales: 68995

Validation graphique des hypothèses :





Les hypothèses semblent être vérifiées, mais de manière très approximative

II. Mack

Estimation des $\hat{\sigma}_j^2$:

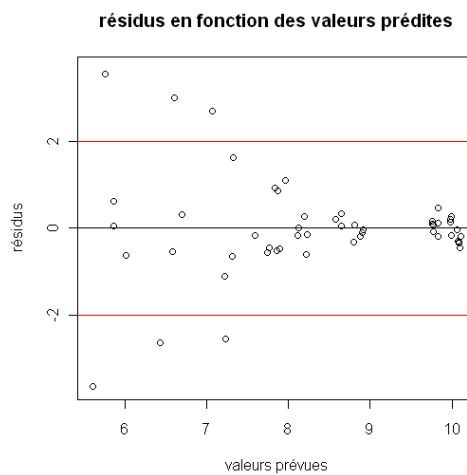
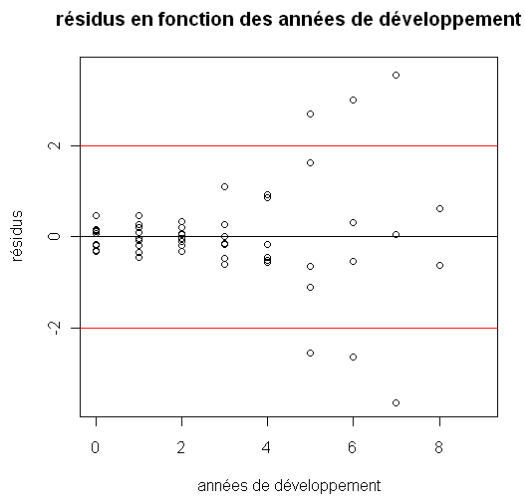
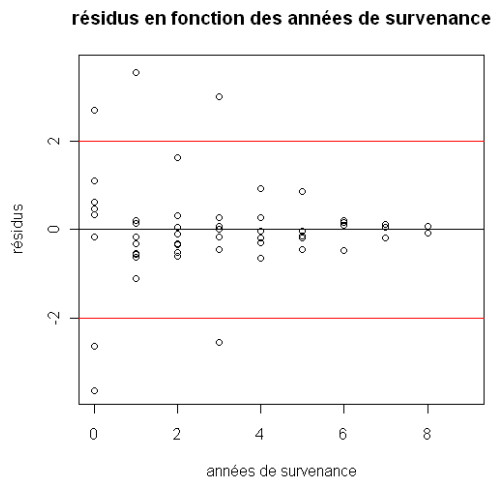
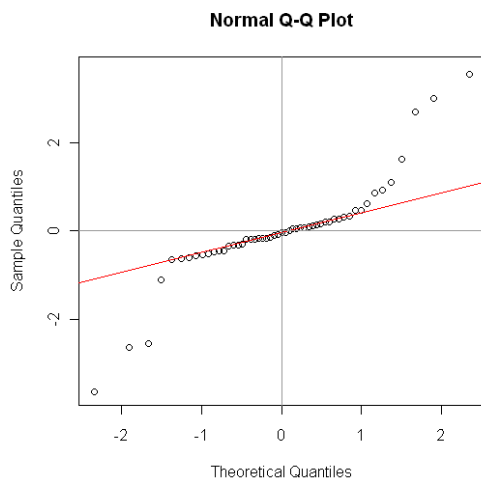
j	0	1	2	3	4	5	6	7	8
$\hat{\sigma}_j^2$	183,84	8,97	6,72	13,79	24,14	9,97	3,22	0,05	$7,98e^{-04}$

i	\hat{R}_i	$\widehat{MSEP}(\hat{R}_i)$	$\sqrt{\widehat{MSEP}(\hat{R}_i)/\hat{R}_i}$
1	784	94	1,24%
2	1140	4612	5,96%
3	1642	273010	31,82%
4	2611	1064600	39,52%
5	4316	2885400	39,36%
6	5481	3029400	31,76%
7	8536	3540000	22,04%
8	13428	3823200	14,56%
9	31057	10589000	10,48%
total	68995	36893587	8,80%

III. Modèles stochastiques

1. Régression log normale

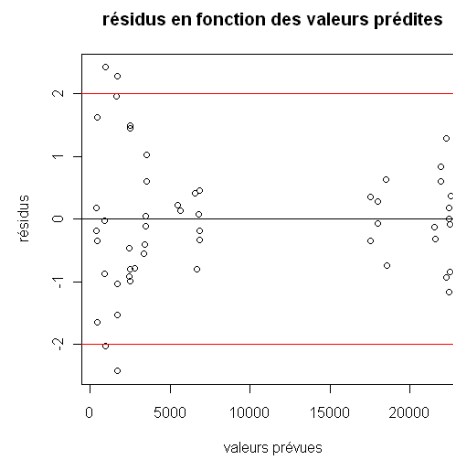
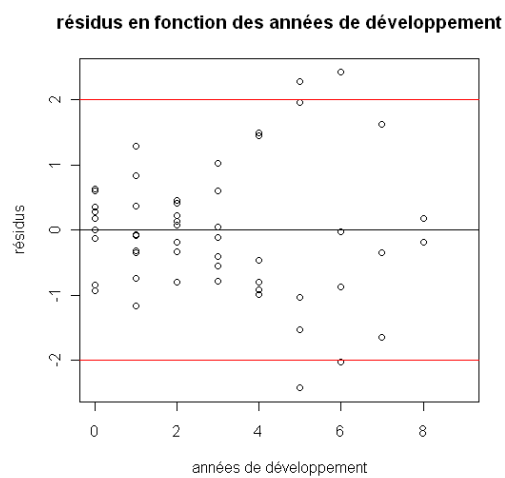
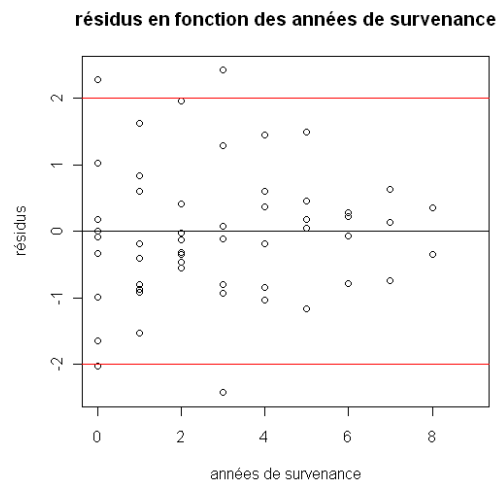
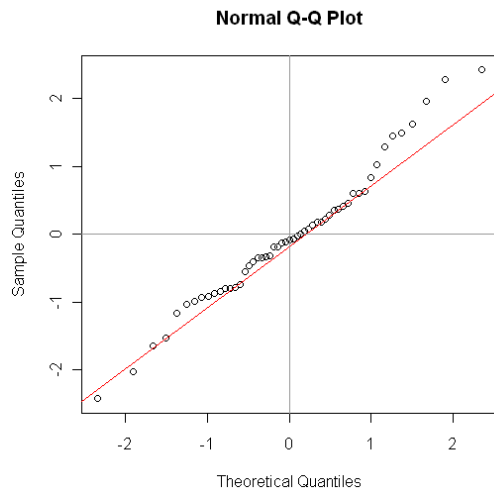
test	p-value	acceptation/rejet (seuil à 5%)
indépendance	0,157	on accepte
normalité	0,0491	on rejette



Valeurs atypiques : (0,5), (0,6), (0,7),
(1,7) , (3,5), (3,6)

2. Modèle Log Poisson Surdispersé

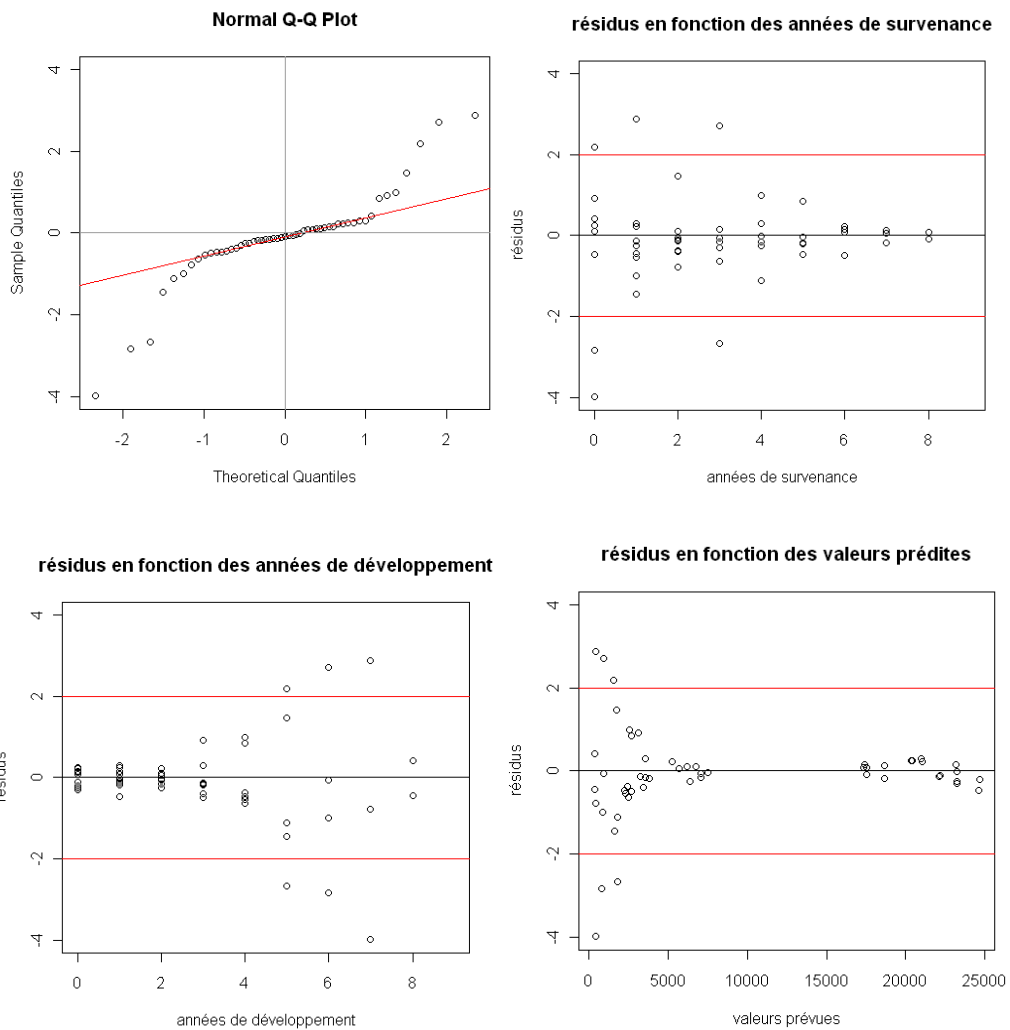
test	p-value	acceptation/rejet (seuil à 5%)
indépendance	0,637	on accepte
normalité	0,661	on accepte



Valeurs atypiques : (0,5), (0,6), (3,5),
(3,6)

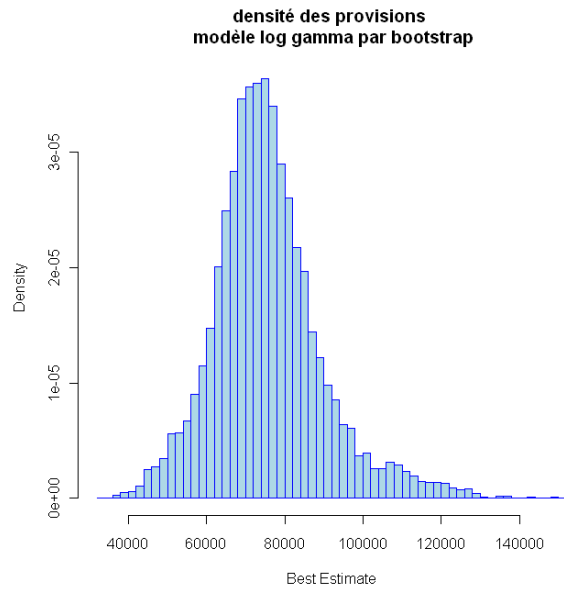
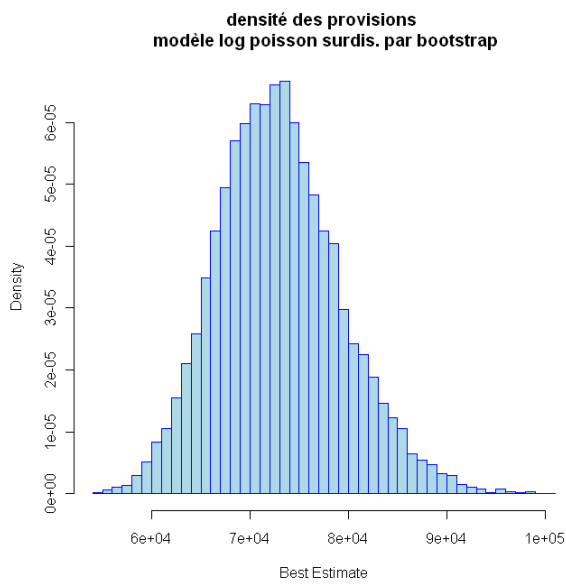
3. Modèle Log Gamma

test	p-value	acceptation/rejet (seuil à 5%)
indépendance	0,346	on accepte
normalité	0,0054	on rejette

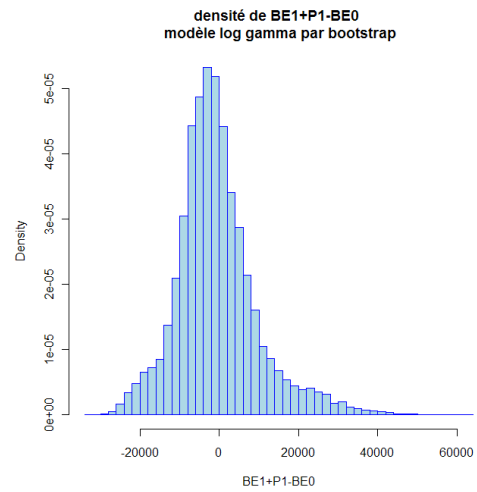
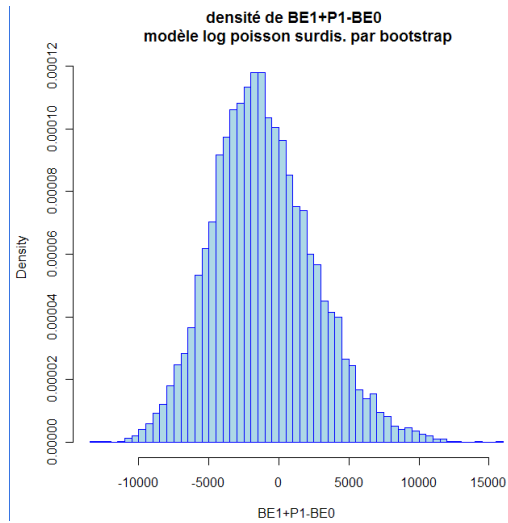


Valeurs atypiques : (0,5),
(0,6), (0,7), (1,7), (3,5),(3,6)

IV. Bootstrap



V. Bootstrap à un an



D. Branche RC Divers :

Triangle de règlements décumulé :

	dvp 0	dvp 1	dvp 2	dvp 3	dvp 4	dvp 5	dvp 6	dvp 7	dvp 8	dvp 9
surv 0	18013	14248	3616	1811	4312	1945	179	921	419	1979
surv 1	17335	16368	4079	1898	749	784	712	2212	1410	
surv 2	17358	12403	3607	2843	2563	1001	3055	1062		
surv 3	14663	16392	6306	5590	1812	14527	3872			
surv 4	15130	16742	5884	2852	3034	4125				
surv 5	14545	13362	4578	3050	894					
surv 6	14520	13335	5190	2681						
surv 7	14019	15558	3851							
surv 8	13885	15135								
surv 9	11085									

I. Chain-ladder :

Coefficients de développement :

j	0	1	2	3	4	5	6	7	8
\hat{f}_j	1,9575	1,1521	1,0837	1,0574	1,1068	1,0424	1,0325	1,0205	1,0435

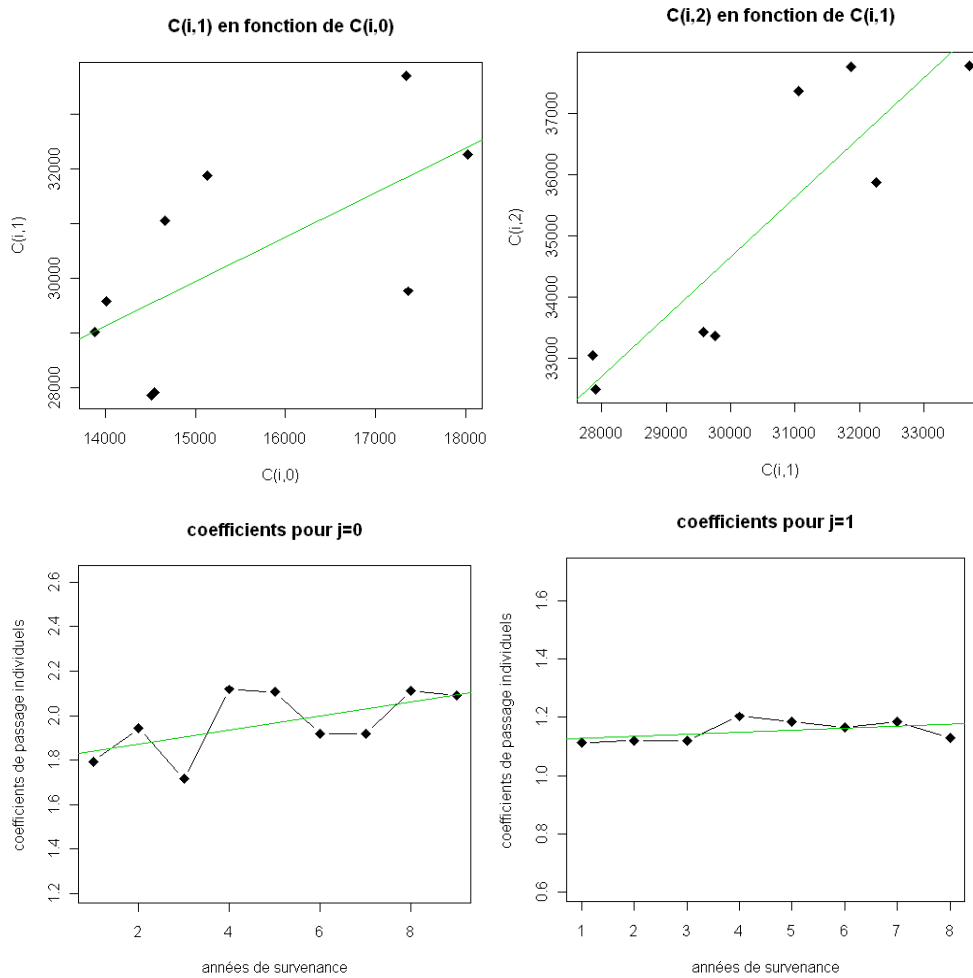
cadences	30,500%	59,705%	68,786%	74,542%	78,824%	87,240%	90,943%	93,903%	95,829%
-----------------	---------	---------	---------	---------	---------	---------	---------	---------	---------

Provisions par année de survenance :

i	\hat{R}_i
0	0
1	1983
2	2850
3	6291
4	6986
5	9787
6	12201
7	15169
8	19586
9	25259

Provisions totales: 100112

Validation graphique des hypothèses :



Les hypothèses ne sont pas vérifiées sur ce triangle, même après plusieurs années de développement.

II. Mack

Estimation des $\hat{\sigma}_j^2$:

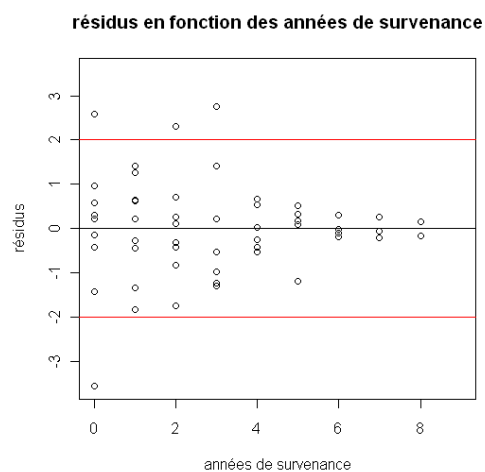
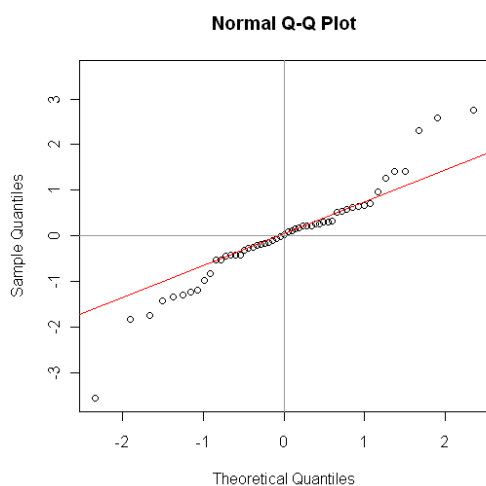
j	0	1	2	3	4	5	6	7	8
$\hat{\sigma}_j^2$	356,71	39,68	41,74	49,39	711,38	56,25	12,86	11,43	10,16

i	\hat{R}_i	$\widehat{MSEP}(\hat{R}_i)$	$\sqrt{\widehat{MSEP}(\hat{R}_i)/\hat{R}_i}$
1	1983	926540	48,54%
2	2850	1718900	46,00%
3	6291	4444400	33,51%
4	6986	7255600	38,56%
5	9787	45814000	69,16%
6	12201	51072000	58,57%
7	15169	54754000	48,78%
8	19586	57489000	38,71%
9	25259	53209000	28,88%
total	100112	479813911	21.88%

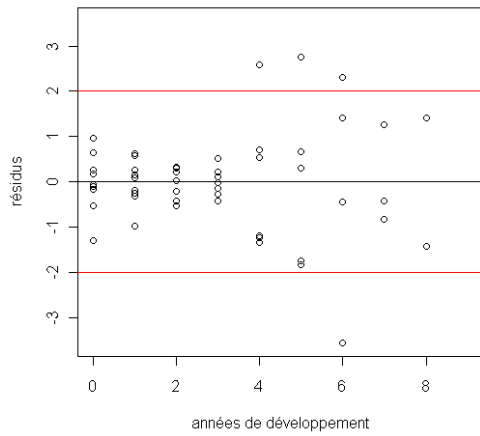
III. Modèles stochastiques

1. Régression log normale

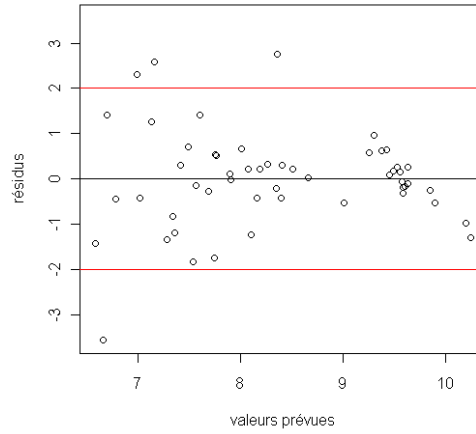
test	p-value	acceptation/rejet (seuil à 5%)
indépendance	0,346	on accepte
normalité	0,508	on accepte



résidus en fonction des années de développement



résidus en fonction des valeurs prédites

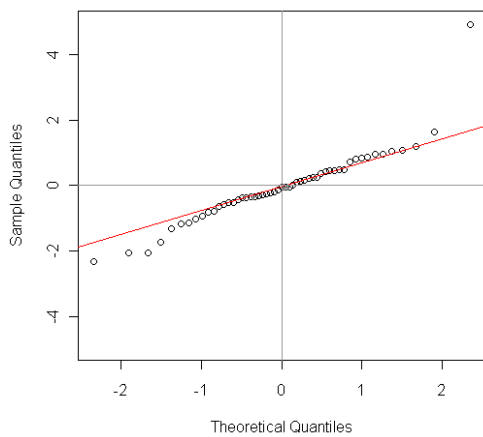


Valeurs atypiques : (0,4), (0,6), (2,6),
(3,5)

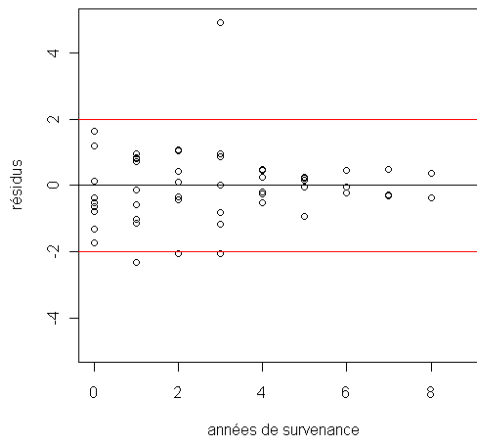
2. Modèle Log Poisson Surdispersé

test	p-value	acceptation/rejet (seuil à 5%)
indépendance	0,346	on accepte
normalité	0,605	on accepte

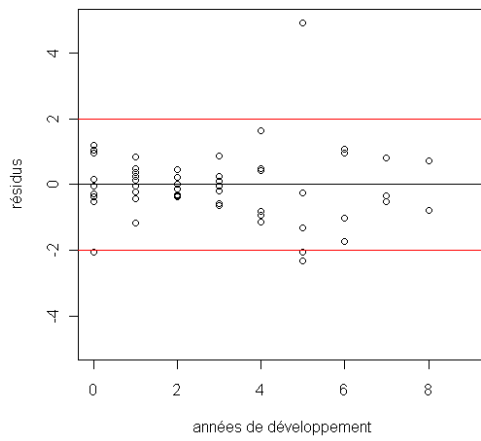
Normal Q-Q Plot



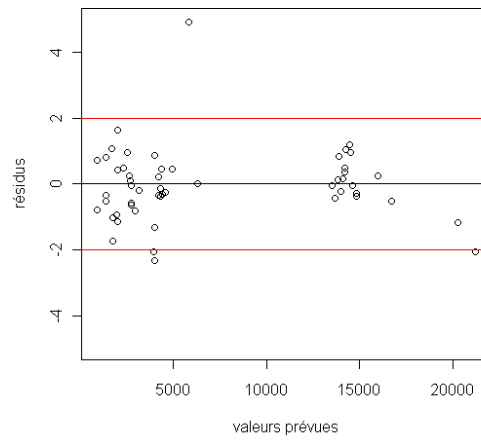
résidus en fonction des années de survénance



résidus en fonction des années de développement



résidus en fonction des valeurs prédites

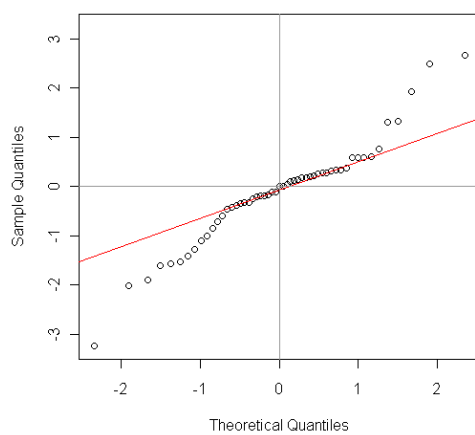


Valeurs atypiques : (3,5)

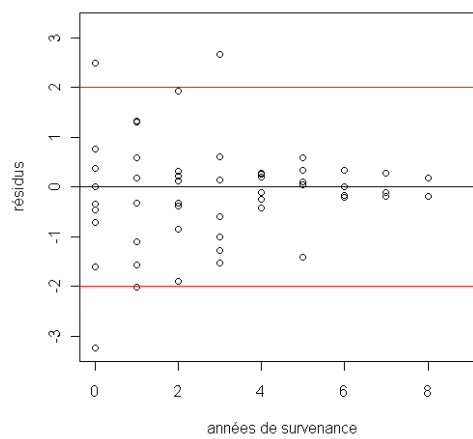
3. Modèle Log Gamma

test	p-value	acceptation/rejet (seuil à 5%)
indépendance	0,059	on accepte
normalité	0,095	on accepte

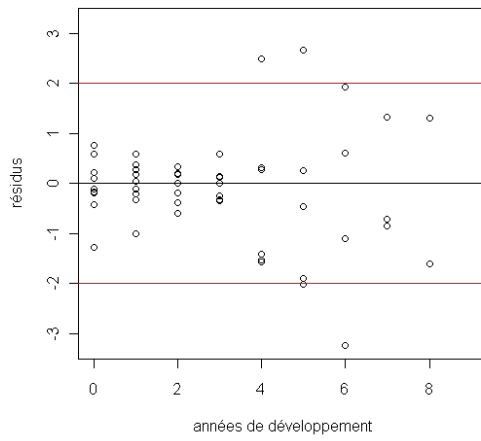
Normal Q-Q Plot



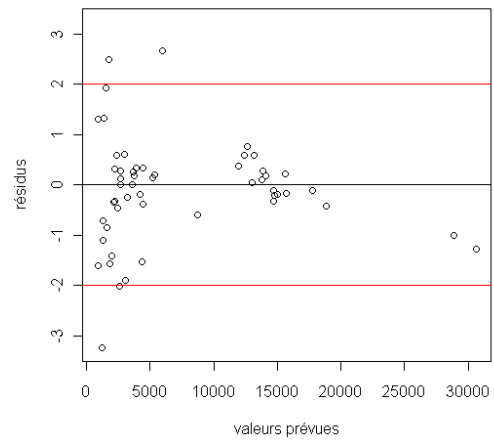
résidus en fonction des années de survivance



résidus en fonction des années de développement



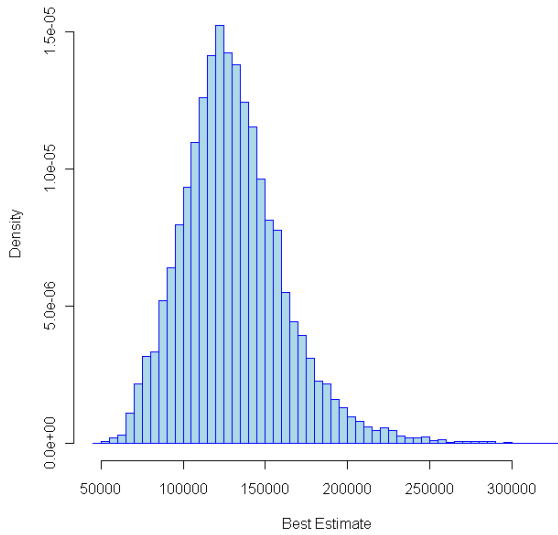
résidus en fonction des valeurs prédites



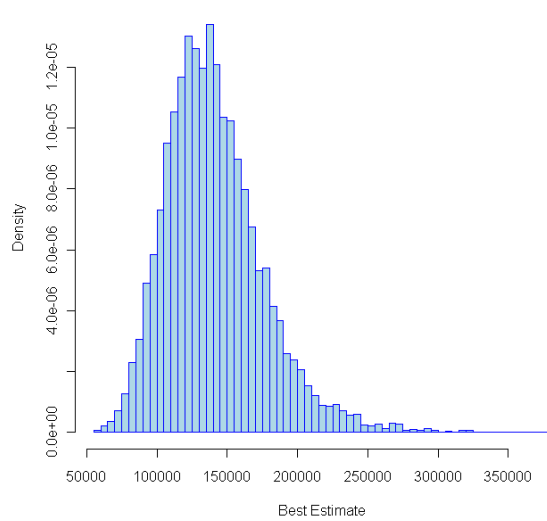
Valeurs atypiques : (0,4),
(0,6), (3,5)

IV. Bootstrap

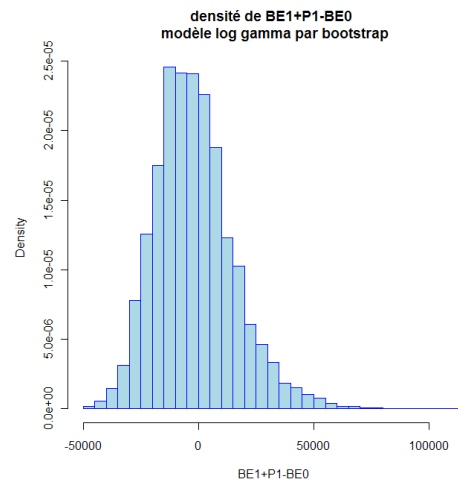
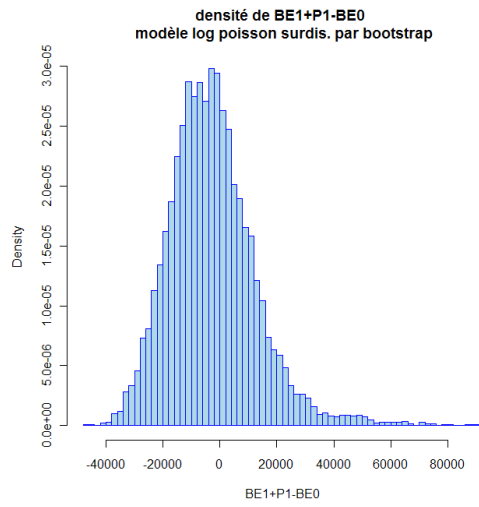
**densité des provisions
modèle log poisson surdis. par bootstrap**



**densité des provisions
modèle log gamma par bootstrap**



V. Bootstrap à un an



BIBLIOGRAPHIE

PARTRAT C. et alii, *Provisionnement technique en assurance non-vie*, Collection Assurance Audit Actuariat, Economica, 2007

DENUIT M., CHARPENTIER A., *Mathématiques de l'assurance non-vie*, Collection Economie et Statistiques Avancées, Economica, 2005

MACK T., *Distribution-free calculation of the standard error of Chain Ladder reserve estimates*, Astin bulletin, Vol 23, 1993

MACK T., *The standard error of Chain Ladder reserve estimates: recursive calculation and inclusion of a tail factor*, Astin bulletin, Vol. 29, 1999

WÜTHRICH M., MERZ M., LYSENKO N., *Uncertainty of the claims development result in the Chain Ladder method*, Scandinavian Actuarial Journal, 2008

GILLET A., SERRA B., *Effets de la dépendance entre différentes branches sur le calcul des provisions*, Mémoire d'actuariat ENSAE, 2002

CADOUX D., LOIZEAU J-M., *Copules et dépendances : application pratique à la détermination du besoin en fonds propres d'un assureur non-vie*, Mémoire d'actuariat CEA, 2004

BELGUISE O., LEVI C., *Tempêtes : étude des dépendances entre les branches Auto et Incendie avec la théorie des Copulas*, Mémoire ULP, 2001

KRAUTH G., *Provisionnement et corrélation entre branches*, Mémoire d'actuariat CNAM, 2007

LAGENEBRE V., *Prise en compte des dépendance entre risques par la théorie des copules : impact sur le besoin en fonds propres de l'assureur*, Mémoire EURIA, 2009

LACOUME A., *Mesure du risque de réserve sur un horizon un an*, Mémoire ISFA 2008

CAMBARROT T., *Provisionnement non-vie : le modèle de Poisson Surdispersé*, Mémoire ISFA, 2009

HESS C., *Modélisation de la dépendance entre risques*, Cours de Théorie du risque est réassurance, Dauphine, 2009

CEIOPS, *QIS 5 Technical Specifications*, 2010