





**Mémoire présenté le :  
pour l'obtention du diplôme  
de Statisticien Mention Actuariat  
et l'admission à l'Institut des Actuares**

Par : Monsieur Walid Rbaibi	
<b>Titre du mémoire :</b> Validation de la modélisation des sinistres graves de la LoB responsabilité civile automobile au sein d'un modèle interne partiel	
Confidentialité : <input checked="" type="checkbox"/> NON <input type="checkbox"/> OUI (Durée : <input type="checkbox"/> 1 an <input type="checkbox"/> 2 ans)	
Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus.	
<u>Membres présents du jury de la filière :</u>	Signature : <u>Entreprise : Groupama Assurances Mutuelles</u> Nom : Ali GOUMAR  Signature : 
<u>Membres présents du jury de l'Institut des Actuares :</u>	<u>Directeur de mémoire en entreprise</u> Nom : Ali GOUMAR Signature :   <u>Invité :</u> Nom : Signature :  <b>Autorisation de publication et de mise en ligne sur un site de diffusion de documents actuariels</b> (après expiration de l'éventuel délai de confidentialité)  <u>Signature du responsable entreprise :</u>   <u>Signature du candidat :</u> 



## Résumé

Afin de calculer le SCR de souscription non-vie de GMA, Groupama utilise un modèle interne partiel qui fait l'objet de revue annuelle. Lors d'une dernière revue, il a été recommandé de renforcer l'analyse et la modélisation des sinistres graves de la ligne d'activité responsabilité civile automobile pour le périmètre GMA.

L'objectif de ce mémoire est d'étudier et valider les méthodes actuelles de traitement des sinistres au sein du modèle interne, tout en proposant des méthodes alternatives permettant de challenger les méthodes actuelles.

Nous avons donc commencé par rappeler le contexte de l'étude, tout en décrivant la méthodologie actuelle de modélisation des sinistres graves de la LoB MTPL au sein du risque de souscription non-vie.

Dans un second temps, nous avons validé la méthode actuelle et proposé des méthodes alternatives comme le modèle individuel multi-états pour le calcul du risque de réserve et des méthodes issues de la théorie des valeurs extrêmes pour la séparation des sinistres graves et attritionnels au sein du risque de primes. Ces méthodes alternatives comme tous les modèles ont des limites que nous avons détaillées dans le mémoire.

Enfin, nous avons conclu par des perspectives pour les exercices futurs.

**Mots clés :** Modèle Interne Partiel, SCR, Risque de réserves, Risque de primes, Validation, Chain-Ladder, Merz-Wüthrich, Calcul miroir, Processus Markovien, Modèle multi-états, Seuil, Théorie des valeurs extrêmes, Modèle individuel.

## Abstract

In order to compute GMA's non-life underwriting risk, Groupama uses a partial internal model which is subject to annual review. During a last review, it was recommended to strengthen the analysis and the modeling of large claims from the motor third party vehicle liability business line on the GMA scope.

The purpose of this paper is to study and attest the current methods that handles the severe claims within the internal model, while proposing alternative methods to challenge the current ones.

We therefore started by recalling the background of the study, while describing the current approach for modeling large claims from the MTPL LoB within the non-life underwriting risk.

Afterwards, we validated the current method and proposed alternative approaches such as the individual multistate model for calculating reserve risk and methods from the extreme value theory for the separation of large and attritional claims within the premium risk. These alternative methods, like all models, have limitations that we have detailed in the brief.

Finally, we concluded with an outlook for the following years.

**Key words :** Partial Internal Model, SCR, Reserve Risk, Premiums Risk, Validation, Mirror Calculation, Chain-Ladder, Merz-Wüthrich, Markov Process, Multi-state Model, Threshold, Extreme Value Theory, Individual Model.

# Synthèse

## Contexte

Afin de mieux refléter sa structure et son profil de risque, Groupama a choisi d'utiliser un modèle interne partiel pour calculer son SCR<sup>1</sup> de souscription non-vie. Ce dernier comporte les risques de primes et de réserve.

L'utilisation d'un modèle interne partiel est réglementée et l'assureur doit être capable de justifier toutes les hypothèses sous-jacentes auprès du régulateur. Ainsi, une revue annuelle est réalisée pour chacun des deux risques, primes et réserves. Lors de l'examen précédent, il a été recommandé de renforcer l'analyse et la modélisation des sinistres graves automobile sur le périmètre GMA<sup>2</sup> au sein du modèle interne partiel.

Pour ce faire, nous allons valider la méthode actuelle de modélisation tout en proposant des méthodes alternatives au sein des risques primes et réserve. Notre démarche consistera donc à :

- Effectuer un calcul miroir
- Vérifier les hypothèses
- Effectuer des tests de sensibilité
- Proposer des méthodes alternatives

## Validation du calcul de la volatilité du risque de réserve

Le SCR du risque de réserve se calcule explicitement à l'aide de la formule suivante :

$$SCR_{res} = BE_0 \cdot \left( \frac{\exp\left(q \cdot \sqrt{\ln(1 + \phi^2)}\right)}{\sqrt{1 + \phi^2}} - 1 \right)$$

Avec :

- $BE_0$  : La provision pour sinistre Best Estimate en  $t = 0$
- $\phi$  : Le coefficient de variation relative au risque de réserve  $\phi = \frac{\sigma}{\sum_{t \geq 1} E(P_t)}$  où :
  - $\sigma$  : écart-type à 1 an
  - $\sum_{t \geq 1} E(P_t)$  : l'assiette de provisions Best Estimate hors escompte
- $q$  : Le quantile à 99.5% d'une loi normale centrée réduite

Au sein du modèle interne, le terme qui est modélisé et que nous allons valider est l'écart-type à 1 an. Ce dernier est calculé en appliquant sur le triangle de charges ou paiements les méthodes suivantes :

1. Chain Ladder afin de déterminer les facteurs de développement
2. Méthode de Mack pour calculer les sigmas de Mack
3. Méthode de Merz-Wüthrich par Bootstrap pour déterminer l'écart-type à 1 an

Dans le cas de la LoB<sup>3</sup> RC auto, la méthode est appliquée au triangle de charges.

---

<sup>1</sup> SCR : Solvency Capital Requirement

<sup>2</sup> GMA : Groupama Assurances Mutuelles

<sup>3</sup> LoB : Line of Business

les facteurs de développement sont complétés avec un facteur de queue modélisé à l'aide d'une loi de Weibull et les sigmas de Mack sont stabilisés à partir de l'année de développement 20 de la façon suivante :

$$\hat{\sigma}_{N-j} = \text{moyenne} (\hat{\sigma}_{N-5}, \hat{\sigma}_{N-6}, \hat{\sigma}_{N-7}) \text{ pour } j = 0, \dots, 4$$

Le calcul miroir donne les résultats suivants :

MTPL	BE Implicite	Ecart-type	CoV
Central	1 656.15	103.97	6.8%
Validation	1 656.15	101.55	6.6%
Ecart	0.0%	-2.3%	-2.3%

Figure 1: Résultats du calcul miroir

Nous constatons des écarts dus à la différence d'outils (R et ResQ) et aux caractères stochastiques de la modélisation.

Les hypothèses que nous allons vérifier sont les suivantes :

1. Les années de survenance sont indépendantes entre elles
2. Il existe un facteur de développement  $f_k$  tel que  $E[C_{i,k+1}|C_{i,1}, \dots, C_{i,k}] = f_k C_{i,k}$  avec  $0 \leq i \leq N$  et  $0 \leq k \leq N-1$
3.  $\forall i = 1, \dots, N, \forall j = 1, \dots, N-1$ , il existe un paramètre réel  $\sigma_j$  tel que :

$$\text{Var}(C_{i,j+1}|C_{i,j}) = \sigma_j^2 \cdot C_{i,j}$$

4. Les résidus du modèle sont i.i.d<sup>4</sup>

Nous avons vérifié les hypothèses à partir de tests graphiques et analytiques.

Ensuite, nous avons effectué des tests de sensibilité par rapport à :

1. La stabilisation des sigmas de Mack
2. Stabilité du modèle et convergence

MTPL	BE RA	Ecart-type	CoV	SCR (LoB)	SCR Total	Ecart SCR (LoB) N/N-1	Ecart SCR Total N/N-1
Central : Charges - TF sur 26 ans - Stab Moy àpd année de dev 20	1 535.48	103.97	6.77%	315.61	422.34	29.6%	14.3%
Charges		60.15	3.92%	205.68	320.76	-15.5%	-13.2%
Charges - Stab Max àpd année de dev 20		96.16	6.26%	295.35	403.29	21.3%	9.2%
Charges - Stab Min àpd année de dev 20		49.35	3.21%	180.27	298.12	-26.0%	-19.3%
Charges - Stab Moy àpd année de dev 20		67.02	4.37%	222.26	335.74	-8.7%	-9.1%
Charges - TF sur 26 ans		75.01	4.88%	241.86	353.64	-0.7%	-4.3%
Paiements		99.31	6.47%	303.49	410.93	24.7%	11.2%
Validation - Charges - TF sur 26 ans - Stabilisation 20 à 24 (1)		81.35	5.30%	257.69	368.22	5.8%	-0.3%
Validation - Charges - TF sur 26 ans - Stabilisation 21 à 25 (2)		137.58	8.96%	405.37	507.87	66.5%	37.5%

Figure 2: Comparaison des stabilisations

	Retraitements	Central - 100	Graine 1	Graine 2	Graine 3	Graine 4	Graine 5	Graine 6	Graine 7	Graine 8	Graine 9	Graine 10
04 - MTPL	TF sur 26 ans + Stabilisation moyenne àpd année de dev 20	6,8%	6,8%	6,7%	6,8%	6,8%	6,8%	6,8%	6,8%	6,9%	6,7%	6,8%

Figure 3: Test de stabilité

	Retraitements	Central - 20k	30k	50k	100k	200k
04 - MTPL	TF sur 26 ans + Stabilisation moyenne àpd année de dev 20	6,8%	6,8%	6,8%	6,8%	6,8%

Figure 4: Test de convergence

<sup>4</sup> i.i.d : Indépendants et identiquement distribués

L'étude de sensibilité ne remet pas en cause la méthode de stabilisation actuelle. Nous remarquons aussi que le modèle est stable et converge rapidement.

Enfin, nous avons proposé le modèle de projection individuelle par états afin de calculer les réserves et d'estimer le risque à 1 an. L'application de ce modèle sur les sinistres graves corporels des caisses régionales de Groupama qui dépassent le seuil des 500 k € donne le résultat suivant :

BE Chain Ladder (en M €)	BE modèle multi-états + binaire (en M €)	Ecart
1 789,9	986,1	803,8

Figure 5: Comparaison des réserves entre modèle multi états et Chain Ladder classique

Nous remarquons que les réserves du modèle que nous avons développé sont plus faibles que les résultats du Chain Ladder. Cela vient du fait d'un manque d'information précise au niveau du sous-dossier victime.

	Risque à 1 an
Bootstrap à 1 an avec un CoV de 5,8%	162,57
Modèles à états + binaire	80,7

Figure 6: Risque à 1 an en M €

Nous remarquons que le SCR risque de réserve calculé par la méthode alternative est inférieur à celui calculé par le Bootstrap à 1 an. Nous précisons que l'insuffisance des données constitue une limite à la pertinence du modèle de projection par états.

### Validation de la charge des graves au sein du risque de primes

Les travaux de validation du calibrage des graves sur la LoB MTPL consistent à :

1. Valider le seuil de séparation des sinistres attritionnels et graves.
2. Valider le calibrage des lois de fréquence et sévérité

Ces validations sont accompagnées de tests d'hypothèses et de sensibilités.

Deux méthodes à interprétation graphique ont été utilisées par l'équipe de modélisation pour obtenir le seuil de séparation à savoir les graphiques de la moyenne des excès et l'estimateur de Hill.

Nous avons testé des méthodes alternatives à savoir :

- Gerstengarbe and Werner plot
- Estimateur de Hill en minimisant l'erreur quadratique moyenne
- QQ estimateur

	Central et QQ estimator	DAMSE	GOMES	Gerstengarbe and Werner
Seuil reassurance (M Euro)	2,1	2,1	2,1	2,1
Seuil (M Euro)	1,38	2,00	1,44	3,04

Figure 7: Comparaison des méthodes

D'après le graphique du QQ estimateur, le seuil retenu est dans une région stable et il est donc pertinent.

Ensuite, Nous avons réalisé une sensibilité par rapport au seuil retenu en choisissant le seuil de 1.1 M €.

	Primes acquises brutes	Charge courante ultime brute	Solde réassurance (hors pool)	Solde pool rétrocession	Frais généraux techniques	Comm de réassurance (hors pool)	Participation aux bénéfices	Résultat tech courant hors escompte	Escompte Taux sans risque	Résultat technique éco courant
mean	0,0	-4,4	-3,3	0,0	0,0	-0,0	0,0	-7,7	-0,1	-7,8
standard_deviation	0,0	-3,9	-3,5	0,0	0,0	-0,0	0,0	-0,7		-0,7
return_period_1_10	0,0	1,7	8,7	0,0	0,0	-0,0	0,0	-6,9	0,0	-6,9
return_period_1_20	0,0	5,1	12,1	0,0	0,0	-0,0	0,0	-6,0	0,0	-6,0
return_period_1_50	0,0	5,7	15,5	0,0	0,0	-0,0	0,0	-6,1	-0,1	-6,0
return_period_1_100	0,0	8,1	15,8	0,0	0,0	-0,0	0,0	-5,8	0,3	-6,2
return_period_1_200	0,0	11,5	-17,8	0,0	0,0	-0,0	0,0	-7,2	-0,3	-7,5
Mean / Return period 99,5%	0,0%	1,1%	0,5%	0,0%	0	0	0,0%	-1,8%		-1,7%

Figure 8: Variation du résultat technique économique

Le résultat technique économique moyen baisse de 7,8 M €. La sensibilité a aussi un impact négatif sur le quantile à 200 ans du résultat technique avec une baisse de 7.5 M €. Le seuil actuel de séparation est donc prudent.

Nous avons réalisé des tests d'indépendance concernant les deux hypothèses du modèles :

- Indépendance entre sévérité et fréquence des graves
- Indépendance entre sinistres attritionnels et sinistres graves

Les tests utilisés sont le test de Kruskal-Wallis et le test de Jonckheere-Terpstra.

hypothèse d'indépendance	Seuil(M Euro)	Jonckheere-Terpstra	Kruskal-Wallis	décision
Fréquence et Sévérité	1,45	11%	5%	Non rejetée
S/P attri et Sévérité	1,45	12%	24%	Non rejetée

Figure 9: Hypothèse d'indépendance

Nous avons effectué un calcul miroir pour le calibrage de la charge des graves.

Sinistralité Grave			CR	
Calibrage des sinistres Graves	Nombre	Distribution	Avec FGAO	Sans FGAO
			DAG	Binomiale négative
		Validation		Binomiale négative
		DAG	37,00	
		Esp	Validation	37,00
		écart		0%
	Var/Esp	DAG	1,83	
		Validation	1,83	
		écart		0%
		DAG	LogNormale	LogNormale
		Validation	LogNormale	LogNormale
	Sévérité	DAG	1448000	
		Min	Validation	1448000
		DAG	250000000	
		Max	Validation	250000000
		DAG	$\mu = 1,96 ; \sigma = 3,52$	$\mu = 10,45 ; \sigma = 2,091$
		Paramètres	$\mu = 1,39 ; \sigma = 3,59$	$\mu = 10,44 ; \sigma = 2,091$
	Charge espérée (M€)	DAG	221,5	186,1
		Validation	221,4	186,0
		écart	0,04%	0,05%
		DAG	557,8	438,1
	Quantile à 99,5% des charges (M€)	Validation	556,1	440,3
		écart	0,3%	-0,5%

Figure 10: Calibrage de la charge des graves

Nous observons des écarts non significatifs.

## Conclusion

Notre étude consiste en une analyse et une validation des méthodes actuelles de modélisation des sinistres graves de la LoB RC Auto. Nous avons vu que ces sinistres sont traités et identifiés de manière différente entre les deux risques, primes et réserve.

Nous avons étendu notre analyse à plusieurs méthodes alternatives en s'inspirant des méthodes nouvelles dans le monde d'actuariat tel que le modèle individuel par états et d'autres issues de la théorie des valeurs extrêmes.

Enfin, Les travaux de ce mémoire enrichissent le processus de validation du modèle interne au sein de Groupama.



# Executive summary

## Background

In order to better reflect its risk structure and profile, Groupama has chosen to use a partial internal model to calculate its non-life underwriting SCR. The latter includes premium and reserve risks.

The use of a partial internal model is regulated and the insurer must be able to justify all underlying assumptions to the supervisory authority. Thus, an annual review is performed for each of the two risks: premiums and reserves. During the previous review, it was recommended to strengthen the analysis and modeling of severe motor claims on the GMA perimeter within the partial internal model.

To do this, we will validate the current modeling method while proposing alternative methods within the premium and reserve risks. Our approach will therefore consist of:

- Perform a mirror calculation
- Verify assumptions
- Perform sensitivity tests
- Propose alternative methods

## Validation of the calculation of the reserve risk volatility

The SCR for reserve risk is explicitly calculated using the following formula:

$$SCR_{res} = BE_0 \cdot \left( \frac{\exp\left(q \cdot \sqrt{\ln(1 + \phi^2)}\right)}{\sqrt{1 + \phi^2}} - 1 \right)$$

With :

- $BE_0$  : Best Estimate claims reserve at  $t = 0$
- $\phi$  : one-year volatility for reserve risk  $\phi = \frac{\sigma}{\sum_{t \geq 1} E(P_t)}$  where :
  - $\sigma$  : one-year standard deviation
  - $\sum_{t \geq 1} E(P_t)$  : Best Estimate reserve base excluding discounting
- $q$  : The 99.5% quantile of a standard normal distribution

Within the internal model, the term that is modeled and that we will validate is the one-year standard deviation. The latter is calculated by applying the following methods to the incurred or paid triangles :

4. Chain Ladder method to determine development factors
5. Mack's method for calculating the Mack's sigmas
6. Merz-Wüthrich method by Bootstrap for determining the one-year standard deviation

In the case of the MTPL LoB, these methods are applied to the incurred triangle.

The development factors are supplemented with a Tail Factor modeled using a Weibull distribution and the Mack sigmas are stabilized from development year 20 as follows :

$$\hat{\sigma}_{N-j} = \text{mean}(\hat{\sigma}_{N-5}, \hat{\sigma}_{N-6}, \hat{\sigma}_{N-7}) \text{ for } j = 0, \dots, 4$$

The mirror calculation gives the following results:

MTPL	Implicit Best Estimate	standard deviation	CoV
Central	1656.15	103.97	6.8%
Validation	1656.15	103.97	6.6%
Difference	0.0%	-2.3%	-2.3%

Figure 1 : Results of the mirror calculation

We note discrepancies due to the difference in tools (R and ResQ) and the stochastic nature of the modeling.

The assumptions we will test are as follows:

1. The years of occurrence are independent of each other
2. There is a development factor  $f_k$  such as  $E[C_{i,k+1}|C_{i,1}, \dots, C_{i,k}] = f_k C_{i,k}$  with  $0 \leq i \leq N$  and  $0 \leq k \leq N - 1$
3.  $\forall i = 1, \dots, N, \forall j = 1, \dots, N - 1$ , there exists a real parameter  $\sigma_j$  such that :

$$Var(C_{i,j+1}|C_{i,j}) = \sigma_j^2 \cdot C_{i,j}$$

4. The residuals of the model are i.i.d <sup>5</sup>.

We tested the assumptions using graphical and analytical tests.

Next, we performed sensitivity tests against:

1. The stabilization of Mack's sigmas
2. Model stability and convergence

MTPL	BE actuarial report	Ecart-type	CoV	SCR (LoB)	SCR Total	Diff SCR (LoB) N/N-1	Diff SCR Total N/N-1
Central : Incurred - TF over 26 years - Mean Stab from year of dev 20	1 535,48	103,97	6,77%	315,61	422,34	29,6%	14,3%
Incurred		60,15	3,92%	205,68	320,76	-15,5%	-13,2%
Incurred - Max Stab from year of dev 20		96,16	6,26%	295,35	403,29	21,3%	9,2%
Incurred - Min Stab from year of dev 20		49,35	3,21%	180,27	298,12	-26,0%	-19,3%
Incurred - Mean Stab from year of dev 20		67,02	4,37%	222,26	335,74	-8,7%	-9,1%
Incurred - TF over 26 years		75,01	4,88%	241,86	353,64	-0,7%	-4,3%
Paid		99,31	6,47%	303,49	410,93	24,7%	11,2%
Validation - Incurred - TF over 26 years - Stab from dev 20 to 24 (1)		81,35	5,30%	257,69	368,22	5,8%	-0,3%
Validation - Incurred - TF over 26 years - Stab from dev 21 to 25 (2)		137,58	8,96%	405,37	507,87	66,5%	37,5%

Figure 2 : Comparison of the different stabilizations

	Adjustments	Central - 100	Seed 1	Seed 2	Seed 3	Seed 4	Seed 5	Seed 6	Seed 7	Seed 8	Seed 9	Seed 10
04 - MTPL	TF over 26 years + Average stabilization from year of dev 20	6,8%	6,8%	6,7%	6,8%	6,8%	6,8%	6,8%	6,8%	6,9%	6,7%	6,8%

Figure 3 : Stability test

	Adjustments	Central - 20k	30k	50k	100k	200k
04 - MTPL	TF over 26 years + Average stabilization from year of dev 20	6,8%	6,8%	6,8%	6,8%	6,8%

Figure 4 : Convergence test

The sensitivity study does not challenge the current stabilization method. We also note that the model is stable and converges quickly.

<sup>5</sup> i.i.d : independents and identically distributed

Finally, we proposed the individual multi-state model to calculate reserves and estimate the one-year reserve risk.

The application of this model on the serious bodily injury claims of Groupama's regional mutuals that exceed the 500k Euros threshold gives the following result:

Chain Ladder BE (in M Euros)	Multi-state + binary model (in M Euros)	Difference
1 789,9	986,1	803,8

Figure 5 : Comparison of reserves between multi-state model and classical Chain Ladder\*

We notice that the reserves of the model we developed are lower than the results of the Chain Ladder. This is due to a lack of precise information at the victim sub-file level.

	one-year reserve risk SCR
one-year Bootstrap with a 5,8% volatility	162,57
Multi-state + binary model	80,7

Figure 6 : one-year reserve risk in M Euros

We note that the reserve risk SCR calculated by the alternative method is lower than that calculated by the one-year Bootstrap. We point out that insufficient data is a limitation to the relevance of the multi-state projection model.

### Validation of serious claims loss within the premium risk

The validation work on the MTPL LoB consists of :

1. Validate the threshold for separating attritional and severe claims.
2. Validate the calibration of the frequency and severity laws

These validations are accompanied by hypothesis and sensitivity tests.

Two methods with graphical interpretation were used by the modeling team to obtain the separation threshold namely the graphs of the mean of the excesses and the Hill estimator.

We have tested alternative methods such as :

- Gerstengarbe and Werner plot
- Hill estimator minimizing the mean square error
- QQ estimator

	Central and QQ estimator	DAMSE	GOMES	Gerstengarbe and Werner
Reinsurance threshold (M Euros)	2,1	2,1	2,1	2,1
Threshold (M Euro)	1,38	2,00	1,44	3,04

Figure 7 : Comparison of methods

From the graph of the QQ estimator, the selected threshold is in a stable region and is therefore relevant.

Then, we performed a sensitivity analysis with respect to the selected threshold by choosing the threshold of 1.1 M €.

	Gross earned premiums	Gross ultimate current loss	Reinsurance balance (excluding pool)	Retrocession pool balance	Technical overheads	Reinsurance commission (excluding pool)	Profit sharing	Current technical result excluding discounting	Discount Risk-free rate	Current economic technical result
mean	0,0	-4,4	-3,3	0,0	0,0	0,0	0,0	-7,7	-0,1	-7,8
standard deviation	0,0	-3,9	-3,5	0,0	0,0	0,0	0,0	-0,7		-0,7
return_period_1_10	0,0	1,7	-8,7	0,0	0,0	0,0	0,0	-6,9	0,0	-6,9
return_period_1_20	0,0	5,1	-12,1	0,0	0,0	0,0	0,0	-6,0	0,0	-6,0
return_period_1_50	0,0	5,7	-15,5	0,0	0,0	0,0	0,0	-6,1	0,1	-6,0
return_period_1_100	0,0	8,1	-15,8	0,0	0,0	0,0	0,0	-5,8	-0,3	-6,2
return_period_1_200	0,0	11,5	-17,8	0,0	0,0	0,0	0,0	-7,2	-0,3	-7,5
Mean / Return period	0,0%	1,1%	0,5%	0,0%	0	0	0,0%	-1,8%		-1,7%

Figure 8 : Sensitivity of the technical-economic result

The average economic technical result decreased by 7.8 million Euros. The sensitivity also has a negative impact on the 200-year quantile of the technical result with a decrease of 7.5 million Euros. The current separation threshold is therefore conservative.

We performed tests of independence for the two hypotheses of the model:

- Independence between severity and frequency of severe claims
- Independence between attritional claims and severe claims

The tests used are the Kruskal-Wallis test and the Jonckheere-Terpstra test.

Assumption	Threshold(M Euro)	Jonckheere-Terpstra	Kruskal-Wallis	decision
Frequency and Severity	1,45	11%	5%	Not rejected
Attritional S/P and Severity	1,45	12%	24%	Not rejected

Figure 9 : independence hypothesis

We performed a mirror calculation for the calibration of the large loss.

Serious loss			CR	
			with FGAO	without FGAO
Calibration of the serious loss	Numbre	Distribution	DAG	Negative binomial
			Validation	Negative binomial
		mean	DAG	37,00
			Validation	37,00
		Var/mean	Difference	0%
			DAG	1,83
	Severity	Distribution	Validation	1,83
			Difference	0%
		Min	DAG	LogNormal
			Validation	LogNormal
		Max	DAG	1448000
			Validation	1448000
		Parameters	DAG	250000000
			Validation	250000000
		Parameters	DAG	$\mu = 1,96 ; \sigma = 3.52$
			Validation	$\mu = 10.45 ; \sigma = 2.09$
		Expected loss (in M€)	DAG	$\mu = 1.39 ; \sigma = 3.59$
			Validation	$\mu = 10.44 ; \sigma = 2.09$
			Difference	
		Quantile to 99.5% of losses (in M€)	DAG	221,5
			Validation	186,1
			Difference	0,04%
			DAG	0,05%
			Validation	557,8
Difference	438,1			
	Validation	556,1		
	DAG	440,3		
	Validation	0,3%		
	DAG	-0,5%		
	Validation			

Figure 10 : Calibration of the serious loss

We observe non-significant differences.

## Conclusion

Our study consists of an analysis and a validation of the current methods for modeling severe claims in the LoB auto liability. We have seen that these claims are treated and identified differently between the two risks, premiums and reserves.

We have extended our analysis to several alternative methods, drawing on methods that are new to the actuarial world, such as the individual state model and others from extreme value theory.

Finally, the work in this thesis enriches the internal model validation process within Groupama.

## Remerciements

Je tiens d'abord à remercier Ali GOUMAR de m'avoir fait confiance en m'offrant cette opportunité chez Groupama Assurances Mutuelles, pour ses qualités de manager, son suivi et ses conseils tout au long de mon alternance.

De plus, je remercie Denis ROUSSET, directeur de la Direction Risques Contrôle et Conformité Groupe ainsi que Renaud BRUNEL, directeur de la Direction des Risques Opérationnels et Contrôle Permanent pour leur confiance ainsi que leurs disponibilités lors des réunions mensuelles de la direction.

Je remercie également Youva MANSOUR, Valentine NAPOLEON et Zakaria MOULIM qui m'ont aidé dans les différentes parties de l'étude.

Au même titre, je remercie toute l'équipe de la Direction Risques Contrôle et Conformité Groupe ainsi que les équipes de la Direction Actuariat Groupe pour leurs aides.

Je remercie également l'ensemble de l'équipe pédagogique de l'ISUP pour leurs enseignements.

# Table des matières

Résumé .....	3
Synthèse .....	5
Remerciements .....	13
Introduction.....	17
I. Contexte du mémoire .....	18
1 Modèle Interne Partiel .....	19
1.1 Cadre Solvabilité 2 .....	19
1.2 Modèle interne partiel .....	19
1.3 Définition et calcul du SCR du risque de souscription non-vie.....	20
1.3.1 Principes de calcul .....	20
1.3.2 Décomposition du SCR non-vie .....	21
1.3.3 Distribution du résultat technique à 1 an .....	23
2 Méthode actuelle de calcul de la volatilité du risque de réserves.....	25
2.1 Principes du calcul du SCR : .....	25
2.2 Estimation des paramètres : .....	27
2.2.1 Chain Ladder .....	27
2.2.2 Méthode de Mack .....	28
2.2.3 Méthode de Merz-Wüthrich par Bootstrap .....	29
2.2.4 Traitements spécifiques .....	31
3 Modélisation de la charge des sinistres graves pour le risque de primes.....	33
3.1 Choix du seuil de séparation des sinistres.....	33
3.1.1 Rappel de la théorie des valeurs extrêmes .....	33
3.1.2 Mean Excess plot.....	35
3.1.3 Estimateur de Hill .....	36
3.2 Modélisation de la charge des graves .....	36
3.2.1 Mise en as if des coûts de sinistres graves.....	37
3.2.2 Mise en as if du nombre de sinistres graves .....	38
3.2.3 Calibrage et choix de lois.....	38
II. Validation de la modélisation des graves.....	39
4 Politique de la validation .....	40
4.1 Objectifs de la validation .....	40
4.2 Processus de la politique de validation .....	40
4.2.1 Champ d'application de la validation .....	40
4.2.2 Plan de validation .....	41

4.3	Outils de la validation .....	41
5	Validation calcul de la volatilité du risque de réserve.....	43
5.1	Spécificité de la modélisation.....	43
5.2	Rejeu et calcul miroir de la volatilité à 1 an .....	44
5.3	Tests de sensibilité mis en place .....	45
5.3.1	Sensibilité par rapport à la stabilisation .....	45
5.3.2	Tests de stabilité et convergence .....	46
5.3.3	Tests de lois .....	47
5.4	Tests d'hypothèses .....	48
5.4.1	Non-corrélation entre les coefficients de développement .....	48
5.4.2	Analyse d'effets d'année calendaire .....	52
5.4.3	Test de Kolmogorov-Smirnov à deux échantillons .....	53
5.4.4	Application.....	55
5.5	Comparaison Bootstrap / Merz-Wüthrich.....	59
5.6	Analyse de variation N/N-1 .....	59
5.7	Analyse des boni/mali et Backtesting .....	60
6	Modèle de projection individuelle par états .....	62
6.1	Modèles de provisionnement individuel.....	62
6.1.1	Approche multi-états en temps discret.....	63
6.2	Modélisation.....	64
6.2.1	Probabilités de transition .....	64
6.2.2	Survenance des sinistres et délai de déclaration .....	65
6.2.3	Modélisation des paiements .....	65
6.2.4	Calcul de la charge ultime et des réserves .....	66
6.2.5	Modélisation du risque à un an.....	67
6.3	Application.....	67
6.3.1	Analyse descriptive des données utilisées dans le modèle.....	68
6.3.2	Modélisation des probabilités de transitions.....	70
6.3.3	Modélisation des paiements .....	73
6.3.4	Estimation du nombre de sinistres déclarés chaque année.....	76
6.3.5	Charge ultime et réserve .....	77
6.3.6	Calcul du risque à 1 an.....	80
7	Validation du calibrage de la charge des graves .....	82
7.1	Choix du seuil de séparation des sinistres.....	82
7.1.1	Méthodologie de l'équipe de modélisation .....	82
7.1.2	Méthodes alternatives pour la détermination du seuil .....	84

7.1.3	Sensibilité du seuil .....	88
7.2	Analyse de l'indépendance.....	89
7.2.1	Indépendance Fréquence / Sévérité .....	89
7.2.2	Indépendance sinistres attritionnels / graves .....	90
7.3	Calibrage de la charge des graves .....	90
7.3.1	Données.....	91
7.3.2	Calibrage de la loi de fréquence .....	91
7.3.3	Calibrage de la loi du coût des graves .....	92
7.3.4	Comparaison de la calibration N/N-1.....	95
	Conclusion .....	97
	Lexique des abréviations.....	98
	Annexes .....	99
	Table des figures.....	107
	Bibliographie.....	109



# Introduction

Dans le cadre prudentiel de Solvabilité 2, l'évaluation des flux de trésorerie futurs et des provisions techniques est de plus en plus importante pour évaluer la solidité financière des portefeuilles d'assurance ou de réassurance. Cette évaluation couvre un vaste champ de scénarios futurs et notamment les plus extrêmes. L'objectif est de bien mesurer les exigences en capital nécessaires pour faire face aux engagements avec un niveau de confiance et un horizon donnés.

Le SCR est le montant de capital dont doit disposer l'assureur pour absorber les pertes potentielles à horizon 1 an avec une probabilité d'accepter une ruine une fois tous les 200 ans. Il peut être déterminé à l'aide de trois méthodes différentes : Formule standard, USP et modèle interne (partiel ou complet).

Afin de mieux refléter sa structure et son profil de risque, Groupama a choisi d'utiliser un modèle interne partiel (MIP) pour calculer son SCR de souscription non-vie. Son utilisation est réglementée et l'assureur doit être capable de justifier toutes les hypothèses sous-jacentes auprès du régulateur. C'est pourquoi Groupama a mis en place une gouvernance afin de garantir l'efficacité des méthodes et des calculs, et de challenger les résultats notamment par des méthodes alternatives.

Lors de la dernière revue du modèle interne partiel, il a été recommandé de renforcer l'analyse et la modélisation des sinistres graves automobile sur le périmètre de la caisse centrale de réassurance « Groupama Assurances Mutuelles » (GMA) au sein du modèle interne partiel.

Pour ce faire, nous allons étudier et valider les méthodes actuelles de modélisation des sinistres graves de la ligne d'activité responsabilité civile automobile. Ensuite nous allons proposer des méthodes alternatives afin de challenger l'approche actuelle.

Dans une première partie, nous allons présenter le contexte de l'étude en nous attachant sur les méthodes actuelles de modélisation des sinistres afin de calculer le SCR de souscription non-vie. En deuxième partie, nous analyserons les méthodes actuelles de traitement des graves tout en proposant différentes méthodes alternatives, à savoir le modèle de projection des sinistres individuels par états pour calculer le risque de réserve à 1 an et les méthodes issues de la théorie des valeurs extrêmes permettant d'identifier le seuil de séparation entre les sinistres attritionnels et les sinistres graves des modèles risque de primes.

## Première Partie

### Contexte du mémoire

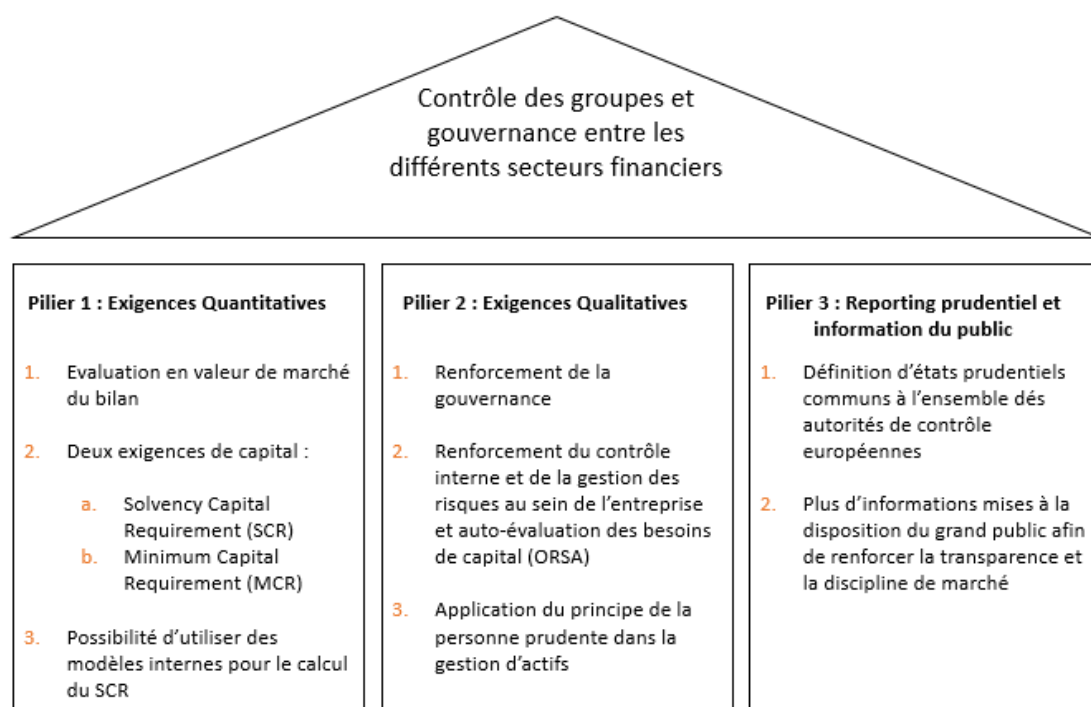
## Chapitre 1

### 1 Modèle Interne Partiel

#### 1.1 Cadre Solvabilité 2

Solvabilité 2 est une réforme réglementaire européenne entrée en application le 1er janvier 2016. Son but est de fixer et mieux adapter le régime de solvabilité des entreprises d'assurance et de réassurance dans l'Union européenne.

Les règles prudentielles de Solvabilité 2 sont réparties en 3 Piliers décrits ci-dessous :



Extrait du cours « Réglementation des Assurances » – Eva Faisse

Figure 1-1: Piliers de Solvabilité 2

Dans ce mémoire, nous allons nous intéresser au calcul du capital de solvabilité requis (SCR) qui correspond au capital exigible pour faire face à une situation de ruine à horizon 1 an dans 99,5% des cas.

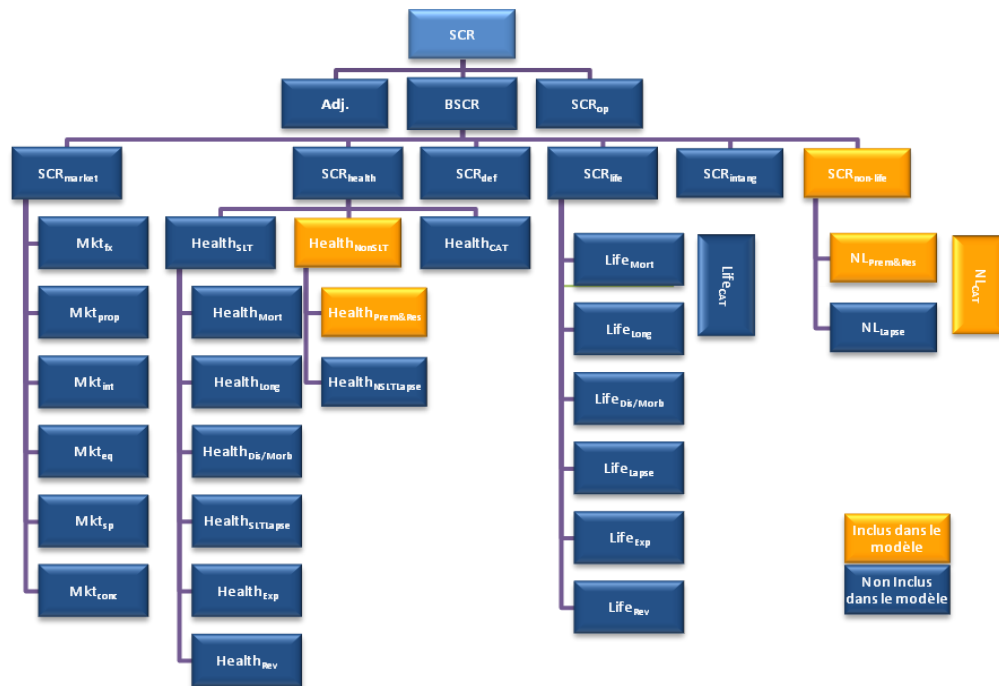
Les assureurs peuvent avoir recours à un modèle interne complet ou partiel pour calculer leurs exigences : Groupama a fait le choix d'évaluer avec un modèle interne partiel.

#### 1.2 Modèle interne partiel

Le modèle interne partiel de Groupama couvre principalement les 2 risques suivants :

- Risque de souscription non-vie
- Risque de souscription de santé non similaire à de la vie

Le schéma ci-dessous distingue les modules de risques inclus dans les modèles internes partiels :



Extrait du corps méthodologique - Groupama

Figure 1-2: Modules de risques du modèle interne partiel

La mesure de risque utilisée pour calculer le SCR non-vie est la *Value at Risk* (*VaR*) avec un niveau de confiance de 99,5%. Cette mesure est définie de la manière suivante :

$$VaR_{\alpha}(X) = \inf\{x | P(X \leq x) \geq \alpha\}$$

### 1.3 Définition et calcul du SCR du risque de souscription non-vie

Le SCR non-vie est défini comme l'opposé de la Value-at-Risk à 0,5% de la variation de situation nette (notée *SN*) à 1 an, qui est la différence entre les actifs et les passifs en valeur économique, avec une nouvelle année de souscription.

La mesure de risque choisie en MIP<sup>6</sup> à savoir la Value-at-Risk est la même que celle de la Formule Standard, avec un niveau de confiance de 99,5% et sur le même horizon d'un an.

#### 1.3.1 Principes de calcul

Le SCR de souscription non-vie s'écrit :  $SCR = -VaR_{0.5\%}(SN_1 - SN_0)$

En notant :

- *A* : les actifs en valeur de marché
- *BE* : la somme escomptée des cash-flows futurs espérés sur la base de la courbe des taux sans risque.
- *RM* : la marge pour risque qui est la part des provisions techniques notées PT permettant de s'assurer que le montant des PT est équivalent au montant que l'assureur doit disposer pour

<sup>6</sup> MIP : Modèle Interne Partiel

honorer ses engagements.

La situation nette s'écrit :  $SN = A - BE - RM$

En supposant que la compagnie n'est soumise qu'au risque de souscription non-vie donc une variation d'actif qui est nulle, la variation de situation nette s'écrit :

$$SN_1 - SN_0 = -R - S - F + PE + [BE_0^{Pr} - BE_1^{Pr}] + RM_0 - RM_1$$

Avec :

- R : les boni/mali sur le stock de provisions pour sinistres Best Estimate représentant le risque de provisionnement ou risque de réserve
- S : la charge sinistre relative aux nouveaux sinistres qui donne lieu à un risque de prime
- F : l'ensemble des frais (administration, acquisition et gestion)
- PE : les primes émises pour la nouvelle année de souscription
- $BE_i^{Pr}$  : Les provisions pour primes Best Estimate à la date  $t = i$

En prenant pour notation :

$$BE_i^{Pr} = PNA_i \times \tau_i^{i+1} + PVFP_i \times (\tau_i^{i+1} - 1)$$

$$PE = PA + PNA_1 - PNA_0$$

$$TR = -R - S - F + PA$$

Où :

- PA : Primes acquises de l'année 1
- $\tau_i^{i+1}$  : Meilleure estimation, vue en fin d'année  $i$ , du ratio combiné net de réassurance de l'année  $i + 1$  ;
- $PNA_i$  : Provisions pour primes non acquises comptabilisées en fin d'année  $i$  ;
- $PVFP_i$  : Assiette de primes futures (*present value of future premiums*) à comptabiliser en fin d'année  $i$ .
- TR : Résultat technique en vision économique vu en fin d'année 1

Nous avons donc :

$$SN_1 - SN_0 = TR + (1 - \tau_1^2) \times (PNA_1 + PVFP_1) - (1 - \tau_0^1) \times (PNA_0 + PVFP_0) - \Delta RM$$

Le SCR de souscription non-vie est simplifiée avec les hypothèses suivantes :

- 1) La variation de marge de risque est négligeable
- 2) Les volumes sous risque de provision pour primes nettes de réassurance en fin d'année 1 ( $PNA_1$  et  $PVFP_1$ ) sont déterministes

La formule du SCR devient donc :

$$SCR = -VaR_{0.5\%} \left( TR + (\tau_0^1 - \tau_1^2) \times (PNA_1 + PVFP_1) \right) + (1 - \tau_0^1) \times (PNA_0 + PVFP_0 - PNA_1 - PVFP_1)$$

### 1.3.2 Décomposition du SCR non-vie

Le résultat technique peut être vu par ligne d'activité Solvabilité 2 ou bien au global toutes LoBs confondues.

Le résultat par LoB s'écrit :  $TR_i = Pr_i - S_i - R_i - F_i$

Où :

- $Pr_i$  : Volume de primes acquises sous risque de la LoB  $i$  (déterministe) nettes de réassurance
- $S_i$  : Charge sinistre vue en fin de 1ère année relative aux nouveaux sinistres (risque de prime)
- $R_i$  : Boni/Mali sur antérieurs (risque de réserve) avec  $E(R_i) = 0$
- $F_i$  : Frais d'administration, acquisition et gestion (déterministes)

Le SCR total s'exprime donc par la formule :

$$SCR = -VaR_{0.5\%} \left( \sum_{i \in LoB} [TR_i + (\tau_{0,i}^1 - \tau_{1,i}^2) \times (PNA_{1,i} + PVFP_{1,i})] \right) + \sum_{i \in Lob} [(1 - \tau_{0,i}^1) \times (PNA_{0,i} + PVFP_{0,i} - PNA_{1,i} - PVFP_{1,i})]$$

Dans le modèle interne, il n'est pas possible d'obtenir par tirages de Monte Carlo la distribution complète de la variable  $\sum_{i \in Lob} TR_i + (\tau_{0,i}^1 - \tau_{1,i}^2) \times (PNA_{1,i} + PVFP_{1,i})$ . C'est pourquoi une approche modulaire est adoptée en calculant :

- Un SCR risque de réserves relatif à l'incertitude sur les montants et dates des paiements des sinistres qui sont déjà survenus ;
- Un SCR risque de primes relatif aux sinistres non encore survenus liés à la souscription pour l'année à venir.

Dans le cas d'une seule LoB S2<sup>7</sup>, le SCR s'exprime donc par :

$$SCR_i = VaR_{99.5\%} (R_i + S_i - Pr_i + F_i - (\tau_{0,i}^1 - \tau_{1,i}^2) \times (PNA_{1,i} + PVFP_{1,i})) + K$$

Avec  $K$  la partie déterministe de la formule.

Vu que les primes, les frais et  $\tau_{0,i}^1$  sont déterministes, ce SCR devient :

$$SCR_i = VaR_{99.5\%} (R_i + S_i + \tau_{1,i}^2 \times (PNA_{1,i} + PVFP_{1,i})) + K$$

Notons :  $A = R_i + S_i + \tau_{1,i}^2 \times (PNA_{1,i} + PVFP_{1,i})$

En centrant les variables  $S_i$  et  $\tau_{1,i}^2$  et en associant un écart-type aux variables  $R_i$  et  $[S_i + \tau_{1,i}^2 \times (PNA_{1,i} + PVFP_{1,i})]$  (noté respectivement  $\sigma_i^R$  et  $\sigma_i^P$ ), l'écart-type de  $A$  est :

$$\sqrt{(\sigma_i^R)^2 + (\sigma_i^P)^2 + 2 \cdot \rho_{R,P} \cdot \sigma_i^R \cdot \sigma_i^P}$$

où  $\rho_{R,P}$  est le coefficient de corrélation entre les variables  $R_i$  et  $[S_i + \tau_{1,i}^2 \times (PNA_{1,i} + PVFP_{1,i})]$

Dans un cadre elliptique, nous avons :

$$VaR_{99.5\%}(A - E(A)) = q \cdot \sqrt{(\sigma_i^R)^2 + (\sigma_i^P)^2 + 2 \cdot \rho_{R,P} \cdot \sigma_i^R \cdot \sigma_i^P}$$

Où  $q$  correspond au quantile à 99,5% d'une loi normale centrée réduite.

---

<sup>7</sup> S2 : Solvabilité 2

Le SCR vaut alors :

$$SCR_i = q \cdot \sqrt{(\sigma_i^R)^2 + (\sigma_i^P)^2 + 2 \cdot \rho_{R,P} \cdot \sigma_i^R \cdot \sigma_i^P} + K$$

Ainsi, dans le cas d'une LoB soumise au risque de prime et de réserve, nous avons :

$$SCR_i = \sqrt{SCR_{res}^i{}^2 + \widetilde{SCR}_{pr}^i{}^2 + 2 \cdot \rho_{R,P} \cdot SCR_{res}^i \cdot \widetilde{SCR}_{pr}^i} - K$$

Le SCR a donc trois composantes :

- Un SCR risque de réserve correspondant au quantile à 99.5% de la quantité  $R_i$  :  $SCR_{res}^i = VaR_{99,5\%}(R_i)$
- La variation de la sinistralité à 99,5% au-delà de la moyenne :  $\widetilde{SCR}_{pr}^i = VaR_{99,5\%}([S_i + \tau_{1,i}^2 \times (PNA_{1,i} + PVFP_{1,i})] - [E(S_i) + E(\tau_1^2) \times (PNA_{1,i} + PVFP_{1,i})])$
- La partie déterministe :  $-E(TR_i) + (1 - \tau_{0,i}^1) \times (PNA_{0,i} + PVFP_{0,i} - PNA_{1,i} - PVFP_{1,i})$

### 1.3.3 Distribution du résultat technique à 1 an

En vision marginale, le SCR relatif au risque de prime repose sur la distribution du résultat technique net de réassurance évalué par LoB S2.

Le SCR relatif au risque de primes ne peut être directement obtenu par le quantile d'une distribution continue. En effet, l'application des traités de réassurance induit des effets non continus que nous ne pouvons pas modéliser à partir des distributions continues.

Ainsi, la distribution du résultat technique en fin de 1<sup>ère</sup> année est obtenue par tirages de Monte Carlo à partir des éléments suivants :

- La sinistralité future qui se compose de :
  - La charge sinistre en brut de réassurance selon trois catégories correspondant à trois types de modélisation différents :
    - Les sinistres attritionnels correspondant à la sinistralité récurrente pour des montants de sinistre inférieurs à un certain seuil ;
    - Les sinistres graves dépassant un certain seuil et correspondant à une sinistralité de fréquence faible et de sévérité importante ;
    - Les sinistres catastrophes correspondant à la sinistralité de pointe, de fréquence extrêmement faible, mais de sévérité extrêmement importante. Cette sinistralité nécessite des approches de modélisation spécifiques car elle n'est pas observée dans les historiques sinistres.
  - L'inflation future ;
  - L'application des traités de réassurance ;
- Les primes reçues et les primes cédées en réassurance (déterministes) – les éventuelles primes de reconstitution sont vues comme une charge au même titre que les sinistres ;
- Les frais d'acquisition, d'administration et de gestion (déterministes) ;
- Les taux d'escompte ;
- La structure de dépendance entre les risques.

Le processus de modélisation du risque de prime est décrit dans le schéma ci-dessous :

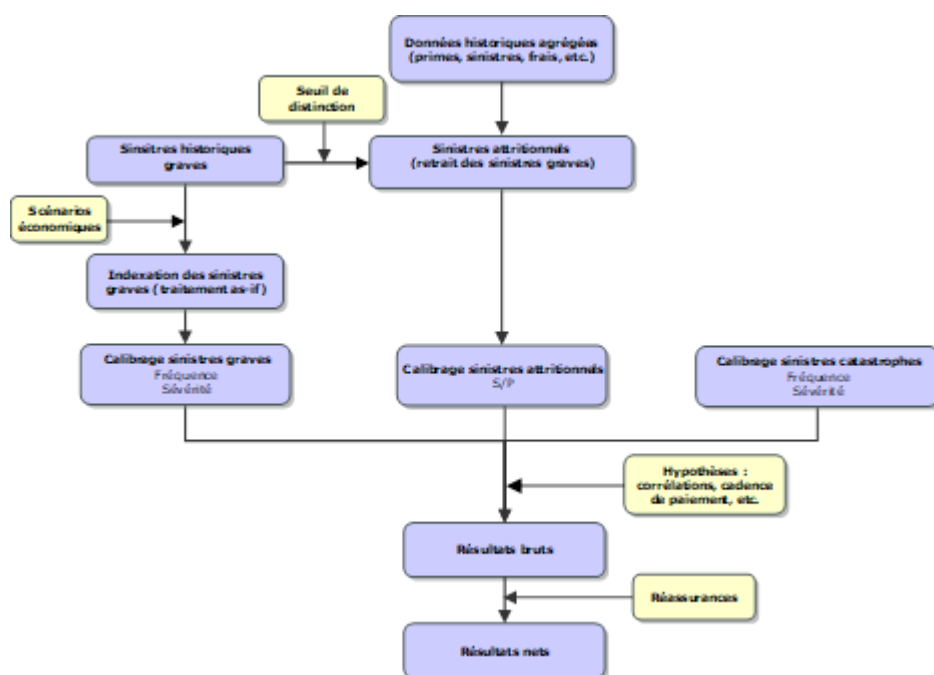


Figure 1-3: Processus de modélisation du risque de primes



## Chapitre 2

### 2 Méthode actuelle de calcul de la volatilité du risque de réserves

Le risque de réserve résulte de fluctuations affectant la date et le montant des paiements de sinistres déjà survenus. À horizon 1 an, la compagnie doit être en mesure de faire face à une déviation défavorable de la sinistralité par rapport à l'anticipation faite dans les provisions Best Estimate des sinistres à payer en  $t=0$ . Il s'agit de capter la distribution des malis de liquidation à un an.

#### 2.1 Principes du calcul du SCR :

L'approche repose sur la variation de la charge ultime Best Estimate entre deux estimations successives en  $t=0$  et en  $t=1$ .

Cette variation correspond à la différence entre la provision pour sinistres Best Estimate  $BE_0$  évaluée en  $t=0$  et la quantité  $P_1 + BE_1$  correspondant aux paiements réalisés en 1<sup>ère</sup> année et la réévaluation des provisions Best Estimate en fin de 1<sup>ère</sup> année.

Nous précisons que le risque de taux est hors périmètre du modèle interne. Les facteurs d'escompte  $D$  sont donc déterministes.

Le SCR de réserves s'écrit :  $SCR_{res} = VaR_{99.5\%}(BE_1 + P_1) \cdot D_{1/0} - BE_0$

Notons :

- $BE_i$  : Somme escomptée en  $t = i$  des cash-flows futurs espérés sur la base de la courbe des taux sans risque.
- $P_i$  : Paiements réalisés en  $i^{ème}$  année
- $D_{i/0}$  : Déflateur en  $t = i$  vu en  $t = 0$  du taux sans risque.

Dans le cadre où un calcul de provisions Best Estimate a été réalisé en  $t=0$ . L'espérance des cash-flows futurs est notée  $E(P_i)$ . La provision Best Estimate est donnée par :

$$BE_0 = \sum_{i \geq 1} E(P_i) \cdot D_{i|0}$$

En fin d'année 1, un 1er paiement a lieu (noté  $P_1$ ) et les cash-flows sont réévalués en fonction de la nouvelle information connue en fin d'année 1. Ces cash-flows conditionnels sont notés  $E(P_i|P_1)$  et nous avons :

$$P_1 + BE_1 = P_1 + \sum_{i \geq 2} E(P_i|P_1) \cdot D_{i|1}$$

En supposant une cadence déterministe, le cash-flow de l'année  $i$  rapporté à la charge ultime (somme de l'intégralité des cash flows) peut s'exprimer de deux manières selon la date d'évaluation ( $t=0$  et  $t=1$ ) :

	t = 0	t = 1
Pourcentage de paiements de 1ère année	$\%\_Paid_1 = \frac{E(P_1)}{\sum_{t \geq 1} E(P_t)}$	$\%\_Paid_1 = \frac{P_1}{P_1 + \sum_{t \geq 2} E(P_t P_1)}$
Pourcentage de cash-flow payé l'année i	$\%\_Paid_i = \frac{E(P_i)}{\sum_{t \geq 1} E(P_t)}$	$\%\_Paid_i = \frac{E(P_i P_1)}{P_1 + \sum_{t \geq 2} E(P_t P_1)}$

Figure 2-1: Extrait du corpus méthodologique - Groupama

Ainsi :

- $BE_0 = \sum_{i \geq 1} E(P_i) \cdot D_{i|0} = [\sum_{t \geq 1} E(P_t)] \cdot \sum_{i \geq 1} \%\_Paid_i \cdot D_{i|0}$
- $P_1 + BE_1 = P_1 + \sum_{i \geq 2} E(P_i|P_1) \cdot D_{i|1} = [P_1 + \sum_{t \geq 2} E(P_t|P_1)] \cdot \sum_{i \geq 1} \%\_Paid_i \cdot D_{i|1}$

En remplaçant ces deux termes dans l'équation du SCR, nous obtenons:

$$SCR_{res} = \left( \sum_{i \geq 1} \%\_Paid_i \cdot D_{i|0} \right) \cdot \left[ Var_{99.5\%} \left( P_1 + \sum_{t \geq 2} E(P_t|P_1) \right) - \sum_{t \geq 1} E(P_t) \right]$$

La seule variable aléatoire dans l'expression du SCR est  $P_1 + \sum_{t \geq 2} E(P_t|P_1)$  que nous notons X.

L'espérance de X est :

$$E \left[ P_1 + \sum_{t \geq 2} E(P_t|P_1) \right] = \sum_{t \geq 1} E(P_t)$$

L'écart type de X est noté  $\sigma$ .

En faisant l'hypothèse d'une loi log normale pour la variable X, les paramètres de cette loi sont :

$$\begin{cases} a = \ln \left( \sum_{t \geq 1} E(P_t) \right) - \ln(\sqrt{1 + \phi^2}) \\ b = \sqrt{\ln(1 + \phi^2)} \end{cases}$$

Où  $\phi$  est le coefficient de variation noté CoV avec  $\phi = \frac{\sigma}{\sum_{t \geq 1} E(P_t)}$

Le facteur  $\phi$  correspond à l'écart type à 1 an ramené à l'assiette de provisions Best Estimate hors escompte

Le quantile Q à 99.5% de X est tel que :  $99.5\% = P(X < Q) = P(\ln(X) < \ln(Q)) = \Phi \left( \frac{\ln(Q) - a}{b} \right)$  où  $\Phi$  est la fonction de distribution d'une loi normale centrée réduite.

Ainsi :  $Q = \exp[a + q \cdot b]$  où q désigne le quantile à 99.5% d'une loi normale centrée et réduite. En remplaçant a et b par leurs valeurs, nous obtenons :

$$Q = \left[ \sum_{t \geq 1} E(P_t) \right] \cdot \frac{1}{\sqrt{1 + \phi^2}} \cdot \exp \left( q \cdot \sqrt{\ln(1 + \phi^2)} \right)$$

Donc :

$$SCR_{res} = BE_0 \cdot \left( \frac{\exp\left\langle q \cdot \sqrt{\ln(1 + \phi^2)} \right\rangle}{\sqrt{1 + \phi^2}} - 1 \right)$$

**Remarque :**

Le SCR de réserve est calculé avec la formule ci-dessus que ce soit en brut ou net de réassurance avec respectivement une provision Best Estimate brute ou nette. Concernant le coefficient de volatilité  $\phi$ , du fait de l'indisponibilité de triangles nets sur le périmètre GMA, le facteur en brut est également utilisé en net de réassurance.

Le SCR net est donc :

$$SCR_{res}^{NET} = BE_0^{NET} \cdot \left( \frac{\exp\left\langle q \cdot \sqrt{\ln(1 + \phi^2)} \right\rangle}{\sqrt{1 + \phi^2}} - 1 \right)$$

Nous notons que le  $BE_0^{NET}$  en modèle interne est appelé BE rapport actuariel et noté BE RA. Ce dernier est issu directement des équipes provisionnement. Le calcul de ce dernier est effectué à une maille plus fine que la LoB Solvabilité 2 (segment de revue).

## 2.2 Estimation des paramètres :

La volatilité à 1 an du risque de réserve est composée de la provision Best Estimate et de l'écart-type à 1 an des charges ultimes.

L'estimation du coefficient de variation s'appuie généralement sur la méthode de Merz-Wüthrich décrite dans les articles :

- England, PD and Verrall, RJ (2002), *Stochastic Claims Reserving in General Insurance*, British Actuarial Journal 8, III, pp443-544
- England, PD and Verrall, RJ (2006), *Predictive distributions of outstanding liabilities in general insurance*, Annals of Actuarial Science 1, II, pp221-270

Au sein du modèle interne, l'approche de Merz-Wüthrich est adaptée par une procédure Bootstrap décrite dans l'article : One-year reserve risk including a tail factor : closed formula and bootstrap approaches

Dans cette partie, nous allons expliciter les principales méthodes sur lesquelles se base la méthode de Merz-Wüthrich par Bootstrap.

### 2.2.1 Chain Ladder

La méthode Chain-Ladder est une méthode déterministe permettant de calculer le montant des réserves. Cette méthode, fréquemment utilisée, est facile à mettre en œuvre. Elle s'applique à des triangles de paiements ou de charges cumulés.

Soit un triangle de paiements ou de charges cumulés avec  $N$  périodes de survénances.

Notons :

- $i$  : année de survénance,  $\forall i \in \mathbb{N}$  tel que  $0 \leq i \leq N$
- $j$  : année de développement,  $\forall j \in \mathbb{N}$  tel que  $0 \leq j \leq N$
- $C_{i,j}$  : montant des paiements ou des charges cumulés entre l'année de survénance  $i$  et l'année de développement  $j$

La méthode de Chain Ladder repose sur les hypothèses suivantes :

- Les années de survénance sont indépendantes entre elles
- Il existe un facteur de développement  $f_k$  tel que  $E[C_{i,k+1}|C_{i,1}, \dots, C_{i,k}] = f_k C_{i,k}$  avec  $0 \leq i \leq N$  et  $0 \leq k \leq N-1$

Nous allons alors considérer des coefficients de développement, commun pour les années de survénance, et dont l'estimation est donnée par :

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

Grâce à ces facteurs nous pouvons estimer :

- Les charges futures :

$$\hat{C}_{i,j} = C_{i,j} \times \hat{f}_{j-1}, \forall i + j \geq N$$

- La charge ultime :

$$\hat{C} = C_{0,N} + \sum_{i=1}^N \hat{C}_{i,N}$$

- Les provisions totales :

$$\hat{R} = \sum_{i=0}^N \hat{C}_{i,N} - \hat{C}_{i,N-i}$$

## 2.2.2 Méthode de Mack

La méthode de Mack est la première méthode faisant intervenir la notion d'incertitude dans la méthode déterministe Chain-Ladder. Introduite en 1993 par T.Mack dans l'article « *Distribution-free calculation of the standard error of the Chain-Ladder reserve estimates*, *Astin Bulletin*, Vol. 23, n°2, 1993 », elle permet de mesurer l'incertitude associée à la prédiction du montant des provisions.

Cette méthode repose sur les mêmes hypothèses que celles de la méthode Chain Ladder avec une supplémentaire :

$\forall i = 1, \dots, N, \forall j = 1, \dots, N-1$ , il existe un paramètre réel  $\sigma_j$  tel que :

$$Var(C_{i,j+1}|C_{i,j}) = \sigma_j^2 \cdot C_{i,j}$$

Les sigmas de Mack mesurent l'incertitude autour des facteurs de développement. Mack propose un estimateur sans biais de  $\sigma_j$  noté  $\hat{\sigma}_j$  :

$$\begin{cases} \hat{\sigma}_j^2 = \frac{1}{N-j-1} \sum_{i=0}^{N-j} C_{i,j} \left( \frac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j \right)^2, & 0 \leq j \leq N-2 \\ \hat{\sigma}_{N-1}^2 = \min \left( \frac{\hat{\sigma}_{N-2}^4}{\hat{\sigma}_{N-3}^2}; \min(\hat{\sigma}_{N-3}^2; \hat{\sigma}_{N-2}^2) \right) \end{cases}$$

À partir de ces sigmas, la méthode de Mack permet de calculer l'écart type des provisions à l'ultime, également appelé MSEP ( *Mean Square Error Prediction* ) .

$$MSEP(R_i) = E \left[ (\hat{R}_i - R_i)^2 | D \right] = E \left[ (\hat{C}_{i,N} - C_{i,N})^2 | D \right] = MSEP(C_{i,N})$$

Avec :

- $\hat{R}_i$  : Le montant de réserve de l'année de survenance  $i$
- $\hat{C}_{i,N}$  : L'estimation des charges futures à l'ultime
- $D$  : L'ensemble d'informations connues au moment du calcul

Après développement, nous avons :

$$MSEP(R_i) = \hat{C}_{i,N}^2 \sum_{k=N-i+1}^{N-1} \frac{\hat{\sigma}_k^2}{\hat{f}_k^2} \left[ \frac{1}{\hat{C}_{i,k}} + \frac{1}{\sum_{j=0}^{N-k} C_{j,k}} \right]$$

Enfin, la MSEP totale est :

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

### 2.2.3 Méthode de Merz-Wüthrich par Bootstrap

Cette méthode réplique par une procédure Bootstrap l'approche de Merz-Wüthrich décrite dans l'article « Modelling the claims development result for solvency purposes » (2008).

La méthode développée par Merz et Wüthrich, appelée Claims Development Result (CDR), est une approche similaire au modèle de Mack, adaptée au calcul du risque sur un horizon d'un an. Les CDR correspondent à la différence entre les réserves au temps  $t=0$  et au temps  $t=1$ , ajustée des paiements effectués durant l'année calendaire 1.

La procédure Bootstrap présente l'avantage de pouvoir obtenir la distribution empirique des provisions Best Estimate à 1 an.

À partir des méthodes Chain Ladder et Mack décrites précédemment, nous pouvons calculer les résidus du modèle comme suit :

$$\frac{\sqrt{C_{i,j}} \cdot (f_{i,j} - \hat{f}_j)}{\sigma_j}$$

Afin de rendre les estimateurs non biaisés, les résidus sont calculés de la façon suivante :

$$r_{i,j} = \sqrt{\frac{N-j}{N-j-1}} \cdot \frac{\sqrt{C_{i,j-1}} \cdot (f_{i,j} - \hat{f}_j)}{\hat{\sigma}_j}$$

Ces résidus sont centrés par construction colonne à colonne, mais pas au global sur le triangle supérieur. Par défaut, ces résidus sont recentrés au global.

Le Bootstrap suppose que les résidus sont i.i.d<sup>8</sup>

Ensuite, la méthode consiste à répéter X fois les étapes suivantes :

1. Rééchantillonnage des résidus  $r_{ij}^*$  afin de créer les triangles de sinistres fictifs à partir du jeu de pseudo-facteurs de développement obtenu à partir du triangle initial et des résidus tirés :

$$f_{i,j}^* = r_{ij}^* \cdot \frac{\hat{\sigma}_j}{\sqrt{C_{i,j-1}}} + \hat{f}_j$$

2. Application d'un Chain-Ladder classique pour obtenir une nouvelle cadence de développement  $\hat{f}_j^*$
3. Simulation de la diagonale de l'année suivante  $N + 1$  à l'aide d'une loi Log Normale dont les paramètres sont déterminés à partir de diagonale  $N$  et les facteurs de développement calculés dans l'étape 2 :

$$C_{ij}^{N+1} \rightarrow \text{LogNormale}(\text{Mean} = C_{i,j-1} \cdot \hat{f}_j^* | \text{Variance} = C_{i,j-1} \cdot \hat{\sigma}_j^2)$$

4. Recalcul des facteurs de développement sachant les informations contenues dans la nouvelle diagonale simulée en  $N+1$  : ces facteurs sont notés  $\hat{f}_j^{*N+1}$
5. Détermination des nouvelles charges ultimes avec la méthode de Chain-Ladder suivant la dernière diagonale  $N+1$  simulée et les coefficients de passage recalculés.

Enfin, notre MSEP est la variance des X charges ultimes calculées.

Les étapes de la procédure Bootstrap sont décrites dans le schéma ci-dessous :

---

<sup>8</sup> Indépendants et identiquement distribués

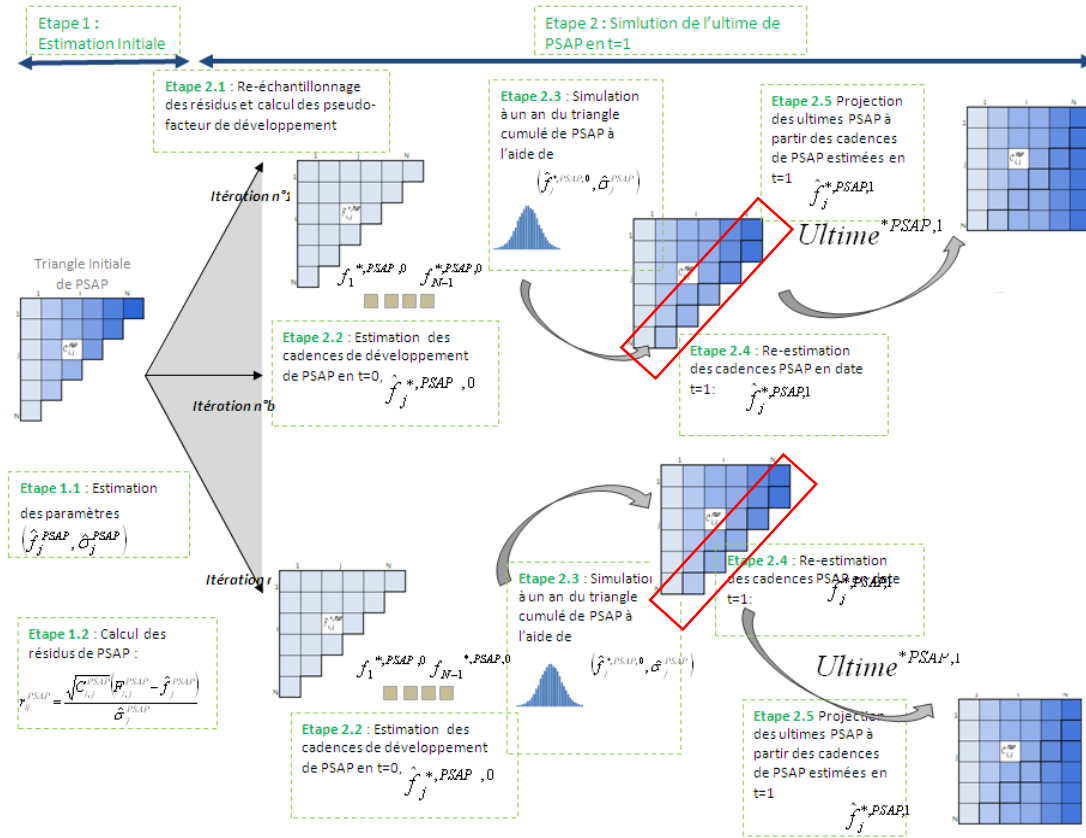


Figure 2-2: Extrait du corpus méthodologique - Groupama

## 2.2.4 Traitements spécifiques

Sur les branches à développement très long comme notre cas d'étude à savoir la LoB MTPL, plusieurs traitements sont effectués sur les facteurs de développements et les sigmas de Mack afin de tenir compte des spécificités de ces branches. Ces traitements concernent l'ajout d'un « tail factor » et la stabilisation des sigmas de Mack.

### 2.2.4.1 Application d'un facteur de queue

Sur une branche longue, il est parfois utile d'utiliser un « Tail Factor » pour estimer la charge ultime afin de ne pas la sous-estimer. Pour cela il faut estimer de nouveaux facteurs appelés les facteurs de la queue de développement.

La méthode utilisée en modèle interne est une extrapolation par un modèle paramétrique. Les facteurs sont déterminés par l'utilisation de lois à deux paramètres.

Les paramètres de la loi sont obtenus à partir d'une régression linéaire.

Par exemple le modèle Weibull, défini par la fonction :

$$f_t = \frac{1}{1 - e^{-at^b}}$$

Peut aussi s'écrire :

$$\ln\left(\ln\left(\frac{f_t}{f_t - 1}\right)\right) = \ln(a) + b * \ln(t)$$

Les paramètres  $a$  et  $b$  sont respectivement l'exponentiel de la constante la pente de la régression linéaire de  $\ln\left(\ln\left(\frac{f_t}{f_{t-1}}\right)\right)$  sur  $\ln(t)$ .

D'autres lois sont utilisées et sont présentées dans le tableau suivant :

	$f_j$
Puissance	$f_t = a^{b^t}$
Exponentielle	$f_t = 1 + a * e^{b^t}$
Puissance inverse	$f_t = 1 + a(c + t)^b$

Figure 2-3: Estimation du facteur de queue

#### 2.2.4.2 Stabilisation des sigmas de Mack

Le calcul des sigmas de Mack donne des estimations atypiques sur les années de développement les plus grandes. Il est donc retenu de mettre en œuvre des méthodes de stabilisation du sigma de Mack sur ces années et de choisir la méthode qui permet d'observer une évolution plus cohérente de la volatilité à 1 an entre deux années de calibrage du SCR de réserve.

Les sigmas de Mack sont stabilisés en prenant soit le maximum, soit le minimum ou la moyenne des 3 sigmas de Mack qui précèdent l'année de développement à partir de laquelle les sigmas sont stabilisés.



## Chapitre 3

### 3 Modélisation de la charge des sinistres graves pour le risque de primes

La modélisation des risques graves se base sur un modèle fréquence/sévérité impliquant donc la calibration de deux lois. Celles-ci sont calibrées à partir des sinistres graves historiques observés au-delà du seuil des graves et permettent de mieux prendre en compte la spécificité de leur aléa d'occurrence (faible fréquence et forte sévérité).

#### 3.1 Choix du seuil de séparation des sinistres

Le seuil de séparation des sinistres est la clé pour la modélisation du risque de primes car il détermine à la fois les données individuelles en input du module sinistres graves et les chroniques de S/P en input du module sinistres attritionnels.

Le seuil optimal est déterminé selon une approche statistique, en application de la théorie des valeurs extrêmes que nous allons parcourir dans cette partie.

##### 3.1.1 Rappel de la théorie des valeurs extrêmes

La théorie des valeurs extrêmes s'intéresse à la loi limite du maximum d'un échantillon avec le Théorème de Fisher-Tippett. Puis, pour ne pas réduire l'information de la queue de distribution à un seul point (le maximum), elle a considéré la loi limite des excédents au-delà d'un certain seuil, avec le Théorème de Pickands.

Avant d'introduire les deux théorèmes, nous précisons que :

- Les distributions  $F$  et  $G$  sont dites de même type s'il existe des constantes  $a > 0$  et  $b \in \mathbb{R}$  telles que  $F(ax + b) = G(x)$
- Soient  $X_i$   $n$  v.a.r.<sup>9</sup> i.i.d. de variable parente  $X$  de fonction de répartition  $F$ ,  $G$  une fonction de répartition.

Le domaine d'attraction maximum de  $G$  noté  $MDA(G)$  est défini comme l'ensemble de lois de fonction de répartition  $F$  telles qu'il existe deux suites réelles  $a_n \in \mathbb{R}$  et  $b_n > 0$  telles que  $\frac{M_n - a_n}{b_n}$  converge en loi vers une v.a.r. de fonction de répartition  $G$ .

Avec :  $M_n = \max(X_1, \dots, X_n)$

Alors  $F$  appartient au max-domaine d'attraction de  $G$ , et nous notons  $F \in MDA(G)$

- La fonction de répartition des excès au-delà du seuil  $u$  est définie par :

$$F_u(x) = Pr(X - u \leq x | X > u) = \frac{F(u + x) - F(u)}{1 - F(u)}$$

---

<sup>9</sup> Variables aléatoires réelles

si  $x \geq 0, u < x_F$ , où  $x_F$  est le point terminal droit de la distribution de  $X$ ,

$$x_F = \sup \{x \in \mathbb{R} : F(x) < 1\}.$$

### Théorème de Fischer-Tippett

Supposons  $n$  variables aléatoires  $X_1, \dots, X_n$  indépendantes et de même loi de distribution  $F$ . S'il existe des constantes  $a_n \in \mathbb{R}$  et  $b_n > 0$ , et une loi non dégénérée de fonction de répartition  $H$  telles que :

$$\frac{M_n - a_n}{b_n} \xrightarrow{\text{loi}} H \quad \text{avec } M_n = \max(X_1, \dots, X_n)$$

Alors  $H$  est du même type que l'une des trois lois suivantes :

1. Loi de Fréchet,  $\Phi_\alpha(x) = \exp(-x^{-\alpha})1_{(x>0)}, \alpha > 0$  ;
2. Loi de Weibull,  $\Psi_\alpha(x) = \exp(-x^{-\alpha})$  si  $x \leq 0$ , et 1 sinon,  $\alpha > 0$  ;
3. Loi de Gumbel,  $\Lambda(x) = \exp(-\exp(-x)), x \in \mathbb{R}$

Ce théorème fournit, une alternative au théorème central limite (TCL) dans le cas d'événements extrêmes. Mais, contrairement au TCL, où la loi asymptotique est la loi normale, avec ce théorème, trois types de lois asymptotiques sont possibles.

Ces trois lois sont en fait, les trois cas particuliers de la distribution GEV (Generalized Extreme Value Distribution) qui est définie par :

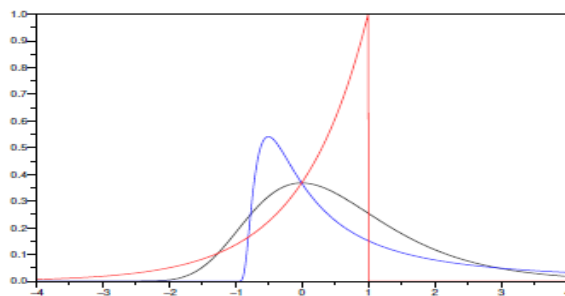
$$H_{\xi, \mu, \sigma}(x) = \exp \left\{ - \left( 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right)^{\frac{-1}{\xi}} \right\}$$

si  $\mu + \xi \frac{x}{\sigma} > 0$ , avec  $\mu \in \mathbb{R}$  et  $\sigma > 0$  respectivement des paramètres de localisation et de dispersion.

$\xi$  est un paramètre de forme appelé indice de queue qui contrôle l'épaisseur de la queue de loi. Plus cet indice est élevé, plus le poids des valeurs extrêmes est important dans notre échantillon et la distribution est dite à queue épaisse.

En effet, nous avons trois cas de loi :

1. Loi de Fréchet,  $\xi = \alpha^{-1} > 0$
2. Loi de Gumbel,  $\xi = 0$
3. Loi de Weibull,  $\xi = \alpha^{-1} < 0$



$\xi = -1$  (Weibull)

$\xi = 0$  (Gumbel)

$\xi = 1$  (Fréchet)

Figure 3-1: les densités des 3 lois de valeurs extrêmes

### Théorème de Pickands

$F \in \text{MDA}(H_\xi)$  où  $H_\xi$  est la loi  $\text{GEV}(\xi)$  si et seulement si : pour  $u$  assez grand :

$$\lim_{u \rightarrow x_f^+} \sup_{y \in [0; x_f^+ - u)} |F_u(y) - G_{\xi, \sigma(u)}(y)| = 0$$

Où  $G_{\xi, \sigma(u)}(y)$  est la fonction de répartition de la loi de Pareto Généralisée (Generalized Pareto Distribution) définie par :

$$G_{\xi, \sigma(u)}(y) = \begin{cases} 1 - \left(1 + \xi \frac{y}{\sigma(u)}\right)^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - \exp\left(-\frac{y}{\sigma(u)}\right), & \xi = 0 \end{cases}$$

Pour  $y \geq 0$  si  $\xi \geq 0$  et  $0 \leq y \leq \frac{-\sigma(u)}{\xi}$  si  $\xi < 0$

Avec  $\sigma(\cdot)$  une fonction positive,  $F_u(x) = \Pr(X - u \leq x | X > u)$

Le théorème de Pickands montre que pour un seuil  $u$  suffisamment élevé, la fonction de répartition de la loi des excès peut être approchée par  $G_{\xi, \sigma(u)}(y)$ ,  $\xi$  indiquant le paramètre de queue et  $\sigma$  celui de l'échelle. Donc pour un seuil assez grand, nous nous attendons à ce que les données au-dessus du seuil suivent une loi GPD<sup>10</sup>.

Ces deux théorèmes seront la base pour les différentes méthodes permettant de déterminer le seuil des graves.

#### 3.1.2 Mean Excess plot

Un premier outil graphique permettant de déterminer le seuil de séparation des graves est le graphe de la moyenne des excès.

Nous définissons d'abord la fonction moyenne des excès (Mean-Excess Function) :

$$e(u) = E(X - u | X > u)$$

Nous pouvons démontrer rapidement que si  $X \sim \text{GPD}(\xi, \sigma)$  avec  $\sigma > 0$  et  $\xi < 1$  alors :

$$e(u) = \frac{\sigma + u\xi}{1 - \xi} 1_{\{\sigma + u\xi > 0\}}$$

Cette fonction est donc affine.

L'estimateur empirique de la fonction d'excès moyen est donné par :

$$\hat{e}(u) = \frac{\sum_{i=1}^n \max(0, X_i - u)}{\sum_{i=1}^n 1_{\{X_i > u\}}}$$

<sup>10</sup> Loi de Pareto généralisée (Generalized Pareto distribution)

Le seuil optimal des graves est déterminé comme celui à partir duquel la courbe  $\hat{e}(u)$  a une forme linéaire.

### 3.1.3 Estimateur de Hill

L'estimateur de Hill est l'estimateur le plus utilisé dans l'estimation du paramètre de queue  $\xi$ . Soient  $n$  variables aléatoires i.i.d.  $X_1, \dots, X_n$ . Rangeons ces variables aléatoires par ordres croissants, et notons :

$$X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n-1:n} \leq X_{n:n}$$

$X_{i:n}$  est donc la  $i^{\text{ème}}$  statistique d'ordre.

L'estimateur de Hill repose sur l'idée que les points  $(\ln(X_{n-k+1:n}), \ln(\frac{k}{n+1}))$  pour  $k = 1, \dots, j$  doivent être approximativement alignés sur une droite de pente  $\xi$  dont l'estimateur est :

$$\hat{\xi}_{j,n}^{Hill} = \frac{\frac{1}{j} \sum_{i=1}^j \ln(X_{n-i+1:n}) - \ln(X_{n-j:n})}{\frac{1}{j} \sum_{i=1}^j \ln\left(\frac{i}{n+1}\right) - \ln\left(\frac{j}{n+1}\right)}$$

Pour  $j$  suffisamment grand, le dénominateur tend vers 1 et nous avons la formule connue de cet estimateur :

$$\hat{\xi}_{j,n}^{Hill} = \frac{1}{j} \sum_{i=1}^j \ln(X_{n-i+1:n}) - \ln(X_{n-j:n})$$

L'estimateur de Hill est volatile lorsque le seuil est faible, puis se stabilise : le seuil optimal des graves est déterminé par  $\hat{k} = X_{n-j:n}$  comme celui à partir duquel l'estimateur  $\hat{\xi}_{j,n}^{Hill}$  se stabilise.

## 3.2 Modélisation de la charge des graves

Dans la modélisation du risque de primes, l'objectif est de quantifier le risque que les montants des sinistres futurs dépassent les primes acquises. Il faut donc étudier le comportement de la survenance des sinistres et établir un modèle statistique permettant de prédire le mieux possible la sinistralité.

Les sinistres graves sont modélisés à l'aide d'un modèle fréquence/coût. Ce modèle va permettre de calculer le quantile de niveau 99,5% des pertes annuelles dues aux sinistres graves, en faisant des tirages de Monte-Carlo d'un nombre  $N$  d'années de sinistralité.

Pour une année de sinistralité, la charge des sinistres est estimée individuellement à l'aide de la loi de distribution de leur nombre annuel et la loi du coût individuel.

La charge ultime des sinistres graves survenus pendant la période est modélisée à partir d'un modèle collectif de « fréquence / sévérité » de la forme  $\sum_{i=1}^N X_i$  où :

- $X$  et  $N$  doivent être indépendantes
- $N$  est la variable aléatoire du nombre de sinistres graves qui surviendront pendant l'année à venir, dépassant le seuil, en cohérence avec le nombre de contrats si disponible, le volume de

primes anticipé dans le plan d'affaires éventuellement retraité des variations tarifaires historiques ;

- $X$  est la variable aléatoire du coût du sinistre  $i$ , brut de réassurance et dépassant le seuil à l'ultime, vu à l'ultime, y compris inflation future moyenne.

Les flux de paiements futurs sont obtenus en appliquant à chaque sinistre individuel une cadence de paiements déterministe avec la méthode Chain-Ladder. Cette cadence de paiements est spécifique à la sinistralité grave de la branche concernée.

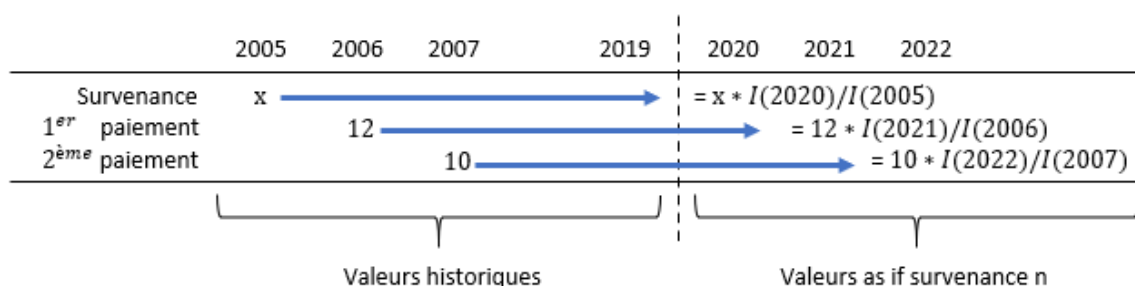
Le calibrage s'appuie donc sur le triangle de paiements des sinistres graves rapportés à leur montant ultime.

### 3.2.1 Mise en as if des coûts de sinistres graves

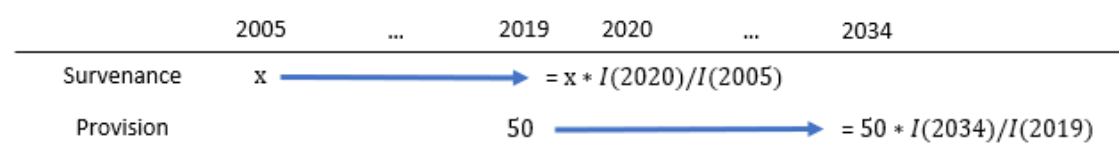
L'indexation des sinistres historiques permet de créer un échantillon de sinistres de survénance as if  $N+1$  contenant implicitement l'inflation future. Les sinistres sont supposés survenir en  $N+1$ , et leurs paiements à venir sont inflatés en conséquence.

En s'appuyant sur le fait que la provision dossier / dossier contient implicitement l'inflation future. Les montants de sinistres historiques sont corrigés de l'inflation de la façon suivante (Exemple de mise en as if en 2020) :

- Les paiements passés sont vieillis en fonction de l'année de survénance du sinistre et de l'année de développement. Par exemple, pour un sinistre survenu en 2005, un paiement effectué en 2006 (année de développement  $t+1$ ), est mis en valeur 2020+1, i.e. 2021 :



- Le dernier montant de la provision évaluée au 31/12/2019 pour le même sinistre survenu en 2005 (année de développement  $t+14$ ) est vieilli à la date 2020+14, i.e. 2034 :



Ainsi, par cette méthode de mise en as if, nous disposons d'un ensemble de sinistres de survénance équivalente  $N+1$ , et dont les montants (de paiements, de provisions et de coûts) contiennent implicitement une inflation future en cohérence avec l'inflation historique constatée sur les portefeuilles.

Groupama s'appuie sur une méthode dite rétrospective et consiste à :

- Mettre en as if uniquement la dernière charge connue au 31/12/ $N$  : paiements individuels + provisions

- Répliquer les boni/mali constatés historiquement en partant du dernier montant mis en as if 31/12/N et en remontant au 1er montant : par exemple, si le sinistre historique est multiplié par 1,2 entre le 31/12/N-1 et le 31/12/N, le montant as-if 31/12/N-1 est égal au montant as if 31/12/N divisé par 1,2. Ainsi, de proche en proche, nous obtenons l'historique complet de chaque sinistre as if.

### 3.2.2 Mise en as if du nombre de sinistres graves

Maintenant que nous avons le déroulé as-if des sinistres graves, les triangles de nombres de sinistres graves sont construits à partir des coûts de sinistres ultimes qui dépassent le seuil de modélisation.

Les nombres de sinistres sont traités de la façon suivante :

- À partir des nombres de sinistres graves historiques, nous appliquons la méthode Chain-Ladder, afin d'en déduire les nombres ultimes de sinistres graves ;
- L'ultime est retraité de variations d'exposition du portefeuille (via un indice de variation tarifaire, le nombre de contrats, à défaut le montant revalorisé des primes).

### 3.2.3 Calibrage et choix de lois

Le choix des lois et l'estimation des paramètres reposent sur des méthodes statistiques usuelles en assurance non-vie : tests d'adéquation ou mesures de statistiques d'adéquation, maximum de vraisemblance, méthode des moments, tests de sensibilité des paramètres, avec éventuellement jugements d'experts.

Les statistiques d'ajustement suivantes sont utilisées au cours du calibrage :

- La statistique d'Akaike ;
- La statistique des moindres carrés ;
- La statistique de Kolmogorov-Smirnov ;
- La statistique de Kuiper ;
- La statistique d'Anderson-Darling.

Les expressions des statistiques utilisées sont détaillées dans l'annexe 1.

## Deuxième Partie

### Validation de la modélisation des graves

## Chapitre 4

### 4 Politique de la validation

L'article 115 de la Directive 2009/138/CE suggère de mettre en place une gouvernance autour du modèle interne partiel. Cette gouvernance comprend la validation du modèle interne. En effet, les modèles de calcul du capital économique sont construits dans un objectif précis et la validation permet de s'assurer qu'ils le remplissent effectivement. Cette validation consiste à faire une revue indépendante et critique des travaux effectués par la DAG<sup>11</sup>.

#### 4.1 Objectifs de la validation

Les objectifs de la politique de validation sont définis dans les articles de la section 7 du Règlement délégué (UE) 2015/35 de la Commission du 10 octobre 2014. D'après ces objectifs, il faut :

- Spécifier les processus, règles et méthodes utilisées pour valider le modèle interne
- S'assurer que le modèle satisfait les exigences déterminées dans les articles 101, 112, 113, 120 à 123, 125 et 126 de la *Directive 2009/138/CE*
- S'assurer de disposer d'un modèle interne à jour et qui reflète au mieux le profil de risque de Groupama. Cela nécessite d'être informé des évolutions pertinentes des sciences actuarielles et des pratiques de marché en ce qui concerne la modélisation des risques
- Garantir la compréhension et renforcer la confiance dans les résultats du modèle interne par le conseil d'administration et les principaux responsables opérationnels
- Identifier les points forts et faibles du modèle afin d'améliorer le modèle auprès des responsables
- Distinguer et résoudre les points de non-conformité réglementaire

#### 4.2 Processus de la politique de validation

##### 4.2.1 Champ d'application de la validation

Le processus de validation du modèle s'applique à toutes les parties du modèle interne et couvre les exigences énoncées dans la *Directive 2009/138/CE* :

- Article 101 et 122 : Calcul du SCR, analyse de ses composantes et de la variation N/N-1
- Article 112 et 113 : Le capital de solvabilité requis qui en résulte du modèle interne reflète mieux le profil de risque de l'entreprise concernée
- Article 123 : Les méthodes utilisées pour l'attribution des profits et pertes et la manière de classification des risques utilisés

---

<sup>11</sup> Direction Actuariat Groupe



- Article 120 : Test relatif à l'utilisation du modèle qui consiste à s'assurer que les entreprises d'assurance ou de réassurance utilisent largement leur modèle interne et que celui-ci joue un rôle important dans leur système de gouvernance
- Article 123 : S'assurer de la présence d'une documentation décrivant la conception et le fonctionnement du modèle interne
- Contrôle opérationnel des résultats fournis par l'équipe modèle interne
- Contrôle de qualité des données

#### 4.2.2 Plan de validation

Les missions de validation suivent un plan défini par des valideurs indépendant (internes ou externes). Ce plan comprend :

- Les travaux à réaliser et les personnes mobilisées
- Documents et fichiers devant être fournis par l'équipe de modélisation
- Les outils de validation utilisés (voir partie 4.3) et une liste des tests envisagés
- Définition de l'effort de validation apporté à chaque élément étudié en fonction de l'importance relative de l'élément
- Le délai estimé pour chaque tâche de validation
- Le sommaire et conclusion prévu dans le rapport rédigés par l'équipe validation à la fin de sa revue
- Restitutions intermédiaires

À la fin de ces travaux, un rapport de validation est rédigé. Ce dernier comprend les éléments suivants :

- Risque revu (primes ou réserves)
- Composant du modèle interne revu (modélisation, documentation, ...)
- Description des tests, leurs finalités et interprétations
- Description des problèmes identifiés en fonction de leur matérialité et proposition d'actions à suivre pour les corriger
- Personnes en charge de la revue

Le rapport doit contenir en plus les conclusions de la validation par rapport aux problèmes majeurs et les recommandations associées. À l'issue des travaux, les résultats sont remontés au CGMI<sup>12</sup> qui doit analyser les recommandations proposées et mesures correctrices envisagées. De plus, il doit s'approprier les remarques matérielles et les points de non-conformité constatés et éventuellement proposer un plan d'action.

### 4.3 Outils de la validation

Les différents outils et tests utilisés dans le cadre de la validation sont les suivants :

- Calculs miroirs : Cela consiste à refaire les modélisations et calculs de manière indépendante
- Tests de modèles alternatifs : Il s'agit de s'assurer que des méthodes alternatives ont été testées et que le modèle retenu est pertinent

---

<sup>12</sup> CGMI : Comité Groupe du Modèle Interne

- Backtesting : Il s'agit de comparer les prédictions effectuées le 31/12/N-1 au titre de l'année N avec les résultats observés réellement au cours de l'année N.
- Sensibilités aux hypothèses du modèle : Ces tests permettent de mesurer l'impact sur les résultats du modèle d'une variation des hypothèses retenues
- Tests de stabilité du modèle : Il s'agit de vérifier que le modèle produit les mêmes résultats sachant les mêmes données d'entrée et le même jeu d'hypothèses. En particulier, vérifier que le SCR estimé ne varie pas significativement lorsque le modèle est lancé à plusieurs reprises en utilisant des jeux de nombres aléatoires différents
- Stress Tests et analyses de scénarios : Ces tests mesurent l'impact d'un événement isolé extrême mais possible (stress test) ou d'un ensemble d'événements (analyse de scénarios). L'approche retenue pour choisir les scénarios et stress tests est documentée dans le rapport de validation.
- Reverse stress test : Ce test sert à identifier les scénarios les plus probables qui pourraient mettre en péril la solvabilité de l'entreprise.
- Benchmarking : Ce test vise à comparer les méthodes et hypothèses utilisée dans le modèle interne avec celles du marché
- Attribution des profits et des pertes : Ce test consiste à revoir les causes ainsi que les sources des profits et des pertes pour chacune des principales unités opérationnelles et à les comparer à celles attendues à la fois sur le périmètre du Modèle Interne et le périmètre de la Formule Standard
- Tests de cohérence : Il s'agit de comparer les données et hypothèses utilisées dans le modèle interne à celles utilisées pour autres fins au sein de Groupama. Cela consiste aussi à analyser les écarts entre les résultats issus du modèle interne et ceux issus de la formule standard.
- Analyse des mouvements : Ce test vise à analyser et comprendre la manière avec laquelle les résultats du modèle évoluent d'une année sur l'autre.
- Revues qualitatives : Ce test permet de valider des éléments tels que la documentation, la gouvernance ou encore la satisfaction du test d'utilisation réglementaire (Use-test)

## Chapitre 5

### 5 Validation calcul de la volatilité du risque de réserve

Nous avons décrit dans la première partie la méthodologie de calcul du SCR risque de réserve. Le calcul se base sur un coefficient de variation  $\phi$  qui est estimé selon le processus suivant :

1. Calcul des facteurs de développement avec la méthode Chain-Ladder
2. Estimation des sigmas de Mack par la méthode de Mack
3. Simulation de la charge ultime avec la méthode de Merz-Wüthrich par Bootstrap
4. Calcul de l'écart-type à 1 an sur la distribution simulée

Dans ce chapitre, nous allons présenter les différents travaux de validation du calcul du SCR de réserve sur la LoB MTPL et sur le périmètre de GMA.

#### 5.1 Spécificité de la modélisation

Les données correspondent aux acceptations de réassurance des 9 caisses régionales métropolitaines. Le calcul de la volatilité de GMA pour la LoB MTPL se fait sur les triangles de charges nets de recours.

L'équipe de modélisation complète les facteurs de développement avec un facteur de queue modélisé à l'aide d'une loi de Weibull et retraite les sigmas de Mack à partir de l'année de développement 20 (voir tableau ci-dessous). Ces retraitements par rapport aux sigmas de Mack sont liés à des coefficients considérés comme atypiques survenus sur les années avec un fort développement. Compte tenu de l'historique limité sur ces années, un élément atypique a un impact très important sur le calcul de l'écart-type.

Périmètre	Méthode 2019	Méthode 2020
GMA	Stabilisation Moyenne à partir de l'année de développement 20	Tail Factor sur 26 ans + Stabilisation Moyenne à partir de l'année de développement 20

Figure 5-1: Retraitements entre 2019 et 2020

L'équipe de modélisation a choisi d'appliquer une stabilisation moyenne sur les derniers sigma de Mack à partir de l'année de développement 20 qui sont retraités de la façon suivante :

$$\hat{\sigma}_{N-j} = \text{moyenne} (\hat{\sigma}_{N-5}, \hat{\sigma}_{N-6}, \hat{\sigma}_{N-7}) \text{ pour } j = 0, \dots, 4$$

Les sigmas de Mack associés aux développements les plus lointains sont présentés dans le tableau suivant pour la clôture 2020 et les clôtures précédentes :

Closing	Méthode	Années de développement								
		17	18	19	20	21	22	23	24	25
2018	Stab Moy N-3	85.28	91.13	84.39	86.93	86.93	86.93			
2019	Stab Moy à partir de l'année de développement 20	77.55	160.28	337.54	191.79	191.79	191.79	191.79		
2020	TF sur 26 ans + Stab Moy à partir de l'année de développement 20	72.23	163.48	314.94	183.55	183.55	183.55	183.55	183.55	183.55

Figure 5-2: Sigma de Mack les plus récentes des 3 dernières clôtures

Ci-dessous sont représentés les sigmas de Mack avant et après retraitement :

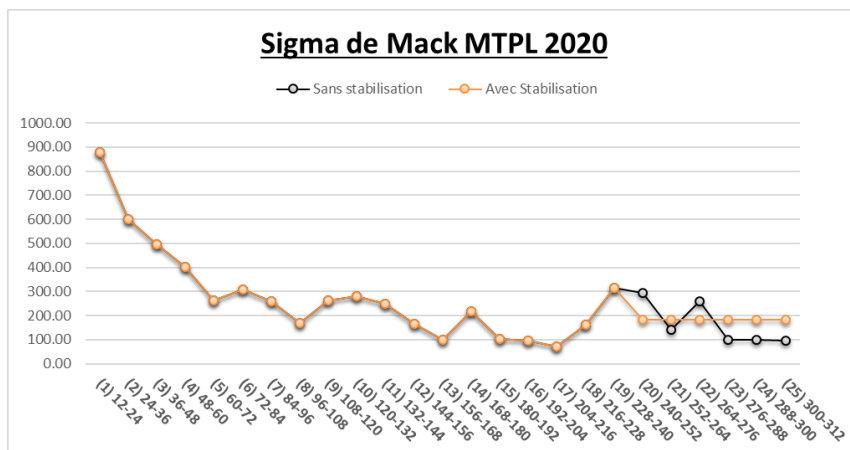


Figure 5-3: Effet de stabilisation sur les sigma de Mack

## 5.2 Rejeu et calcul miroir de la volatilité à 1 an

Les résidus du modèle qui vont servir comme input de la méthode de Merz-Wüthrich par Bootstrap sont calculés à partir des facteurs de développements et les sigma de Mack de la façon suivante :

$$r_{i,j} = \sqrt{\frac{N-j}{N-j-1}} \cdot \frac{\sqrt{C_{i,j-1}} \cdot (f_{i,j} - \hat{f}_j)}{\hat{\sigma}_j}$$

Les résidus calculés sont présentés dans le tableau suivant :

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	-0,479	-0,125	-0,982	-0,517	-0,277	-0,013	-0,440	0,213	0,095	-0,074	-0,194	-0,405	0,939	0,305	-0,865	-1,146	1,339	0,152	-0,010	2,039	1,692	-1,4267103	-1,062
2	-0,271	-1,220	-1,020	0,342	0,949	0,632	1,159	1,219	-0,277	0,553	2,200	-1,637	2,302	1,401	0,915	-0,560	-0,809	1,589	-0,746	-0,437	-0,933	0,8939501	0,933
3	-0,167	-0,475	-1,198	-0,252	-0,230	0,408	0,869	-0,799	0,293	-0,399	0,118	1,599	-0,547	-0,270	-0,539	-0,134	1,500	0,195	-0,405	-0,200	-0,025	0,404	
4	0,010	-0,372	-0,491	0,011	0,467	0,040	0,534	0,097	-0,824	-0,019	0,060	0,288	-0,979	1,575	0,412	0,368	-0,079	0,852	-0,514	-0,445	-0,520		
5	0,784	-0,184	0,530	0,551	0,862	-2,395	-0,446	-0,079	0,578	0,293	0,715	-1,198	-0,566	-0,194	0,874	-0,520	-0,146	0,028	2,185	-0,647			
6	0,516	0,776	-0,494	1,715	-1,197	-0,793	0,067	0,748	0,795	-0,249	1,034	1,711	-0,339	0,768	-1,443	1,780	-1,787	-1,509	-0,490				
7	-0,329	0,198	1,445	-0,679	0,802	-0,824	0,150	-0,272	-0,383	2,741	-1,244	0,094	0,069	-2,172	1,701	1,519	0,291	-1,188					
8	0,136	0,217	-0,915	-0,936	-0,446	0,013	-0,608	0,233	0,220	-0,064	-0,045	-0,320	-0,766	0,258	0,527	-0,466	0,046						
9	-0,235	0,040	-0,577	0,739	-0,121	-1,136	1,333	-0,247	0,279	-0,153	-0,783	-0,292	0,460	-0,433	-1,326	-1,118							
10	0,117	-0,436	-1,452	-2,169	-0,277	-0,739	-0,012	-0,672	0,461	-0,998	-0,450	1,064	-0,832	-0,665	-0,419								
11	-0,018	-0,835	-0,622	0,469	0,749	0,107	-1,462	-1,632	0,401	0,448	-0,373	0,724	1,277	-0,587									
12	1,113	-0,201	-0,280	1,588	2,054	1,077	-1,860	-0,633	-0,878	-1,283	-0,128	-1,043	-1,000										
13	0,036	-0,294	0,072	0,176	1,588	0,215	-1,590	-1,445	-2,404	-1,149	-0,135	-0,462											
14	0,113	0,907	0,464	0,201	-1,051	0,454	-0,198	2,715	2,482	-1,181	0,894												
15	1,219	1,106	-0,659	0,134	-1,372	0,448	1,681	-0,271	-0,032	1,129													
16	0,885	-0,648	-0,100	0,273	-1,086	-0,933	-0,669	-0,393	-0,907														
17	-1,206	0,424	-0,489	-0,368	-0,498	-0,344	0,030	0,923															
18	-0,727	-1,219	1,282	0,119	0,180	0,686	1,017																
19	1,261	1,312	2,823	-2,370	-1,398	2,542																	
20	0,793	-1,812	0,155	1,112	1,397																		
21	-1,630	-1,384	0,164	0,911																			
22	-0,971	1,143	1,174																				
23	-1,206	2,651																					
24	2,392																						

Figure 5-4: Résidus du Bootstrap

Afin de nous assurer de la bonne implémentation du modèle par l'équipe de modélisation, nous avons reproduit le calcul de la volatilité à 1 an avec le même outil de modélisation que l'équipe de

modélisation (ResQ) et ensuite effectué un calcul miroir sous R. Les résultats détaillés de ces deux calculs sont présentés ci-dessous :

MTPL	BE RA	BE Implicite (M€)	Ecart-type (M€)	CoV Retenu
Central	1 535,48	1 656,15	103,97	6,8%
Validation		1 656,15	103,97	6,8%
Ecart		0,00	0,00	0,0%

Figure 5-5: Rejeu du calcul de volatilité à 1 an

Nous n’observons aucun écart en utilisant l’outil de l’équipe de modélisation.

MTPL	BE Implicite	Ecart-type	CoV
Central	1 656.15	103.97	6.8%
Validation	1 656.15	101.55	6.6%
Ecart	0.0%	-2.3%	-2.3%

Figure 5-6: Calcul miroir de volatilité à 1 an

Nous constatons des écarts dus à la différence d’outils (R et ResQ) et aux caractères stochastiques de la modélisation.

Le SCR du risque de réserve au niveau de cette LoB est enfin calculé selon la formule explicitée dans la première partie :

$$SCR_{res}^{NET} = BE_0^{NET} \cdot \left( \frac{\exp\left(q \cdot \sqrt{\ln(1 + \phi^2)}\right)}{\sqrt{1 + \phi^2}} - 1 \right)$$

### 5.3 Tests de sensibilité mis en place

Afin de mesurer l’impact des choix de calibrage effectués par l’équipe modélisation, différentes études de sensibilités ont été réalisées sur les hypothèses du modèle du risque de réserve de la LoB MTPL.

Cette partie comprend à la fois les sensibilités réalisées par l’équipe de modélisation et revue par l’équipe validation et d’autres proposés par la validation.

#### 5.3.1 Sensibilité par rapport à la stabilisation

Nous avons étudié ici l’impact de la stabilisation des sigmas de Mack selon les années de développement :

- Stabilisation des 5 derniers sigmas de Mack hors le sigma de Mack issu du Tail Factor en prenant la moyenne des 3 précédents (17, 18, 19)
- Stabilisation des 5 derniers sigmas de Mack y compris le sigma de Mack du Tail Factor en prenant la moyenne des 3 précédents (18, 19, 20)

Les derniers sigmas de Mack pour ces deux tests de stabilisation sont les suivants :

	17	18	19	20	21	22	23	24	25
Sans retraitement	72.23	163.48	314.94	292.89	141.01	258.05	100.32	100.32	95.65
Central	72.23	163.48	314.94	183.55	183.55	183.55	183.55	183.55	183.55
Validation - Charges - TF sur 26 ans - Stabilisation 20 à 24 (1)	72.23	163.48	314.94	183.55	183.55	183.55	183.55	183.55	95.65
Validation - Charges - TF sur 26 ans - Stabilisation 21 à 25 (2)	72.23	163.48	314.94	292.89	257.10	257.10	257.10	257.10	257.10

Figure 5-7: Sigma de Mack stabilisés

Le choix de ces deux nouveaux tests vient du fait que la stabilisation pour cette année est effectuée sur un sigma de Mack de plus que ce qui devrait être effectué : la méthode de stabilisation doit s'appliquer aux 5 dernières années de développement au maximum

Ensuite, nous avons revu les tests de divers retraitements effectués par l'équipe de modélisation. Les résultats de ces tests ainsi que les 2 précédents sont résumés dans le tableau ci-dessous avec le calcul du SCR au niveau de la LoB et du SCR total et une comparaison avec les résultats N-1.

MTPL		BE RA	CoV	SCR N-1 (LoB)	SCR Total N-1
Central : Charges - Stab Moy àpd année de dev 20		1 342.44	5.7%	243.46	369.38

MTPL	BE RA	Ecart-type	CoV	SCR (LoB)	SCR Total	Ecart SCR (LoB) N/N-1	Ecart SCR Total N/N-1
Central : Charges - TF sur 26 ans - Stab Moy àpd année de dev 20	1 535.48	103.97	6.77%	315.61	422.34	29.6%	14.3%
Charges		60.15	3.92%	205.68	320.76	-15.5%	-13.2%
Charges - Stab Max àpd année de dev 20		96.16	6.26%	295.35	403.29	21.3%	9.2%
Charges - Stab Min àpd année de dev 20		49.35	3.21%	180.27	298.12	-26.0%	-19.3%
Charges - Stab Moy àpd année de dev 20		67.02	4.37%	222.26	335.74	-8.7%	-9.1%
Charges - TF sur 26 ans		75.01	4.88%	241.86	353.64	-0.7%	-4.3%
Paielements		99.31	6.47%	303.49	410.93	24.7%	11.2%
Validation - Charges - TF sur 26 ans - Stabilisation 20 à 24 (1)		81.35	5.30%	257.69	368.22	5.8%	-0.3%
Validation - Charges - TF sur 26 ans - Stabilisation 21 à 25 (2)		137.58	8.96%	405.37	507.87	66.5%	37.5%

Figure 5-8: Sensibilité sur les retraitements

Nous remarquons que la méthode retenue (Central) conduit à un CoV<sup>13</sup> de 6,7%, soit une hausse de 1,2%. Cette hausse de la volatilité est portée principalement par l'application durant cette année du Tail Factor qui entraîne une forte croissance de l'erreur de prédiction (67,0 M € vs 104,0 M € soit une hausse de 36%).

Nous notons aussi une stabilité globale du SCR en scénario identique (Stabilisation moyenne à partir de l'année de développement 20) entre N et N-1, d'où l'intérêt des méthodes de stabilisation. En effet, la sensibilité sans méthode de stabilisation donne un résultat de SCR en forte baisse par rapport à l'année passée. Enfin, l'utilisation supplémentaire d'un Tail Factor vient augmenter largement le SCR de la LoB qui passe de 222 M € à 315 M €.

Le premier test de la validation montre bien le côté prudent d'appliquer également la stabilisation sur le Tail Factor. En revanche, le second test qui donne une valeur de SCR de 507 M € par rapport à une valeur centrale de 422,3 M € montre bien la conséquence d'un décalage de la stabilisation. Nous précisons aussi que nous avons obtenu pour cette méthode un CoV de 9% par rapport à 6,7% de la méthode centrale.

Enfin, pour garder un bon niveau de SCR et une stabilité par rapport au SCR N-1, cette étude de sensibilité ne remet pas en cause la méthode de stabilisation actuelle pour la clôture 2020.

### 5.3.2 Tests de stabilité et convergence

<sup>13</sup> Coefficient de variation

Nous avons effectué des tests de stabilité à la graine aléatoire et des tests de convergence par rapport au nombre de simulations de la méthode de Merz-Wüthrich par Bootstrap. Les résultats sont reportés dans les tableaux suivants :

	Retraitements	Central - 100	Graine 1	Graine 2	Graine 3	Graine 4	Graine 5	Graine 6	Graine 7	Graine 8	Graine 9	Graine 10
04 - MTPL	TF sur 26 ans + Stabilisation moyenne àpd année de dév 20	6,8%	6,8%	6,7%	6,8%	6,8%	6,8%	6,8%	6,8%	6,9%	6,7%	6,8%

Figure 5-9: Test de stabilité

	Retraitements	Central - 20k	30k	50k	100k	200k
04 - MTPL	TF sur 26 ans + Stabilisation moyenne àpd année de dév 20	6,8%	6,8%	6,8%	6,8%	6,8%

Figure 5-10: Test de convergence

Les écarts constatés sont non significatifs. Le modèle est stable par rapport à la graine et converge rapidement. Les 20 000 simulations utilisées sont suffisantes pour obtenir une stabilité de la volatilité.

### 5.3.3 Tests de lois

L'équivalence entre les résultats de la méthode de Merz et Wüthrich et l'approche Bootstrap utilisée en modèle interne est assurée par deux aspects :

- La variance process : les coefficients de passage sont tirés de façon indépendante (étape 1 du Bootstrap)
- L'erreur d'estimation : à chaque itération Bootstrap, le même triangle supérieur bootstrappé sert à estimer les coefficients de passage  $\hat{f}_j$  qui rentrent ensuite dans le tirage de la diagonale N+1. (étape 3 du Bootstrap)

Ces deux aspects sont démontrés de façon analytique dans l'article [3].

Nous avons effectué des tests par rapport aux lois associées à ces deux aspects, les résultats obtenus sont les suivants :

Process Variance	Retraitements	Central Lognormal	None	Resampled	Normal	Gamma
04 - MTPL	TF sur 26 ans + Stabilisation moyenne àpd année de dév 20	6.8%	2.4%	6.8%	6.8%	6.9%

Figure 5-11: Sensibilité erreur de process

Estimation Variance	Retraitements	Central Resampled	None	Normal	Lognormal	Gamma
04 - MTPL	TF sur 26 ans + Stabilisation moyenne àpd année de dév 20	6.8%	6.4%	6.8%	6.8%	6.7%

Figure 5-12: Sensibilité erreur d'estimation

Nous remarquons des écarts non significatifs hormis le cas où nous n'intégrons pas l'erreur de processus. Les CoVs sont donc stables.

## 5.4 Tests d'hypothèses

Le modèle de Mack et la méthode de Merz-Wüthrich par Bootstrap reposent sur les hypothèses suivantes :

- Hypothèse de linéarité
- Hypothèse de variance
- Hypothèse d'indépendance des années de survénance
- L'hypothèse d'indépendance et l'identique distribution des résidus

Dans cette partie nous allons expliciter la théorie des tests statistiques effectués pour chaque hypothèse et ensuite les résultats obtenus.

### 5.4.1 Non-corrélation entre les coefficients de développement

L'hypothèse de linéarité est démontrée par un test de non-corrélation des coefficients de développement successifs. En effet, la linéarité implique la non-corrélation et la démonstration de cette relation est en annexe 2.

L'absence de corrélation est testée à deux niveaux :

- Sur le triangle dans son ensemble en utilisant le test statistique du  $\rho$  de Spearman
- Sur chaque couple de coefficients de développement successifs en utilisant les tests du  $\rho$  de Spearman, du  $\tau$  de Kendall et du coefficient de corrélation de Pearson.

Dans la suite, nous allons voir les différents tests statistiques permettant de vérifier la non-corrélation.

#### Test de significativité du coefficient de corrélation de Pearson

Le coefficient de corrélation de Pearson est construit à partir de la covariance entre deux variables  $X$  et  $Y$ . La covariance permet de mesurer si les deux variables sont du même côté de leur espérance. Sa formule est la suivante :

$$COV(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$$

L'estimateur de la covariance est pour un échantillon de taille  $n$  est :

$$COV_n(X, Y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

Avec

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Si la covariance est positive, quand la variable  $X$  est supérieure ou inférieure à sa moyenne,  $Y$  a tendance à l'être aussi, et l'inverse quand la covariance est négative.

Le coefficient de Pearson, noté  $\rho$  est une normalisation de la covariance pour qu'elle soit entre -1 et 1 et est donné par :



$$\rho = \frac{COV(X, Y)}{\sigma(X)\sigma(Y)}$$

Avec

$$\sigma(X) = \sqrt{E[(X - E[X])^2]}$$

L'estimateur du coefficient de Pearson est donné par :

$$\rho_n(X, Y) = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Soit  $n$  le nombre d'années de développement ou d'années de survénance (supposés être égaux)

Pour chaque année de développement  $j$  tel que  $n - j \geq 4$  :

Le coefficient de corrélation entre les données des deux colonnes successives est calculé de la façon suivante :

$$\rho_{j,j+1} = \frac{\sum_{i=1}^{n-j-1} (f_{i,j} - \bar{f}_j)(f_{i,j+1} - \bar{f}_{j+1})}{\sqrt{\sum_{i=1}^{n-j-1} (f_{i,j} - \bar{f}_j)^2} \sqrt{\sum_{i=1}^{n-j-1} (f_{i,j+1} - \bar{f}_{j+1})^2}}$$

Avec  $\bar{f}_j$  la moyenne des  $f_{i,j}$  pour  $1 \leq i \leq n - j - 1$

Ensuite, la statistique du test est la suivante :

$$t_{j,j+1}^* = \frac{|\hat{\rho}_{j,j+1}|}{\sqrt{\frac{1 - \hat{\rho}_{j,j+1}^2}{n - j - 3}}}$$

Sous l'hypothèse  $H_0 : \rho_{j,j+1} = 0$ ,  $t_{j,j+1}^*$  suit une loi de Student à  $n - j - 3$  degrés de liberté.

En choisissant le seuil  $\alpha = 5\%$  si  $t_{j,j+1}^* < t_{n-j-3}^{\frac{\alpha}{2}}$  l'hypothèse de la non-corrélation n'est pas rejetée.

Avec  $t_{n-j-3}^{\frac{\alpha}{2}}$  le quantile théorique d'ordre  $1 - \frac{\alpha}{2}$  d'une loi de Student à  $n - j - 3$  degrés de liberté.

Nous effectuons le test en calculant sa p-value qui est la probabilité que la valeur absolue d'une variable aléatoire suivant la loi de Student à  $n - j - 3$  degrés de liberté soit supérieure à la statistique du test.

La p-value doit être supérieur au seuil  $\alpha = 5\%$  pour ne pas rejeter l'hypothèse  $H_0$ .

### Test de signification du Rhô de Spearman

Le coefficient de corrélation de Spearman repose sur le même principe que le coefficient de Pearson, mais en transformant les données étudiées par la fonction rang qui a une observation  $x_i$  associe son rang dans l'échantillon  $X$ .

Pour chaque année de développement  $j$  tel que  $n - j \geq 4$  :

1. Nous calculons les rangs des coefficients de développement individuel  $f_{i,j}$  et on les note  $r_{i,j}$
2. Nous calculons les rangs des coefficients de développement individuel  $f_{i,j}$  privée du coefficient de passage de la dernière année de survénance observée  $f_{n-j,j}$  et nous notons  $s_{i,j}$  le rang de ces coefficients
3. Nous calculons le coefficient de corrélation entre la colonne des  $r_{i,j+1}$  et celle des  $s_{i,j}$  :

$$\rho_{j,j+1} = \frac{\sum_{i=1}^{n-j-1} (r_{i,j+1} - \bar{r}_{j+1})(s_{i,j} - \bar{s}_j)}{\sqrt{\sum_{i=1}^{n-j-1} (r_{i,j+1} - \bar{r}_{j+1})^2} \sqrt{\sum_{i=1}^{n-j-1} (s_{i,j} - \bar{s}_j)^2}}$$

La statistique du test est la même que pour le test de significativité de Pearson qui est :

$$T_{j,j+1}^* = \frac{|\rho_{j,j+1}|}{\sqrt{\frac{1 - \rho_{j,j+1}^2}{n - j - 3}}}$$

Et la décision de rejeter l'hypothèse  $H_0 : \rho_{j,j+1} = 0$  se fait de la même façon.

### Test de signification du Tau de Kendall

Le coefficient de Kendall noté  $\tau$  est aussi basé sur l'ordre des observations mais d'une manière différente que celle explicitée pour le coefficient de Spearman.

Notation :

Soient  $(x, y)$  et  $(u, v)$  deux couples d'observations indépendantes et identiquement distribuées :

- Les couples sont concordants si :  $(x < u \text{ et } y < v)$  ou  $(x > u \text{ et } y > v)$
- Les couples sont discordants si :  $(x < u \text{ et } y > v)$  ou  $(x > u \text{ et } y < v)$

Le coefficient est égal à la probabilité de concordance moins la probabilité de discordance.

Ainsi :

$$\tau(X, Y) = P[(X - U)(Y - V) > 0] - P[(X - U)(Y - V) < 0]$$

Avec  $(U, V)$  un couple de variables indépendantes de  $(X, Y)$  et identiquement distribué

L'estimateur du Tau de Kendall est pour un échantillon de taille  $n$  est :

$$\tau_n(X, Y) = \frac{\text{nombre de paires concordantes} - \text{nombre de paires discordantes}}{\text{nombre de paires total}}$$

Avec

$$\text{nombre de paires total} = \binom{n}{2} = \frac{n(n-1)}{2} \quad \text{avec} \quad \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

### Application sur un triangle de développement :

Soit  $n$  le nombre d'années de développement ou d'années de survénance (supposés être égaux)

La forme générale du Tau de Kendall entre les rangs des coefficients de développement entre deux années successives est donnée par :

$$\tau_{j,j+1} = \frac{C - D}{\binom{n-j-1}{2}}$$

Avec

C : nombre de couples concordants

D : nombre de couples discordants

Une modification à cette formule est appliquée dans le cas de coefficients nuls qui ont le même rang pour une année de développement

Le coefficient corrigé est le suivant :

$$\tau_{j,j+1} = \frac{C - D}{\frac{1}{2} \sqrt{((n-j-1)(n-j-2) - U_X)((n-j-1)(n-j-2) - U_Y)}}$$

Avec

$$U_X = \sum_x u_x(u_x - 1) \text{ et } U_Y = \sum_y u_y(u_y - 1)$$

$u_x$  et  $u_y$  sont respectivement le nombre d'apparitions du rang x dans l'année de développement j et le nombre d'apparitions du rang y dans l'année de développement j+1.

Cette correction modifie la variance du coefficient et donc le calcul de la statistique pour le test de significativité.

La variance du coefficient de Kendall est donnée par :

$$\sigma^2(\tau_{j,j+1}) = \frac{\sigma^2(S)}{Den_{j,j+1}^2}$$

Avec :

$$Den_{j,j+1} = \frac{1}{2} \sqrt{((n-j-1)(n-j-2) - U_X)((n-j-1)(n-j-2) - U_Y)}$$

$$\sigma^2(s) = \frac{(n-j-1)(n-j-2)(2(n-j-1) + 5) - V_X - V_Y}{18}$$

$$+ \frac{W_X W_Y}{9(n-j-1)(n-j-2)(n-j-3)} + \frac{U_X U_Y}{2(n-j-1)(n-j-3)}$$

Où :

$$V_X = \sum_x u_x(u_x - 1)(2u_x + 5), \quad V_Y = \sum_y u_y(u_y - 1)(2u_y + 5), \quad W_X = \sum_x u_x(u_x - 1)(u_x - 2)$$

$$W_Y = \sum_y u_y(u_y - 1)(u_y - 2)$$

Enfin,  $\tau_{j,j+1}$  suit une loi normale centrée, l'hypothèse  $H_0 : \tau_{j,j+1} = 0$  n'est pas rejetée au seuil  $\alpha$  si :

$$\frac{|\tau_{j,j+1}|}{\sigma(\tau_j)} < q_{1-\frac{\alpha}{2}}$$

avec  $q_{1-\frac{\alpha}{2}}$  le quantile d'ordre  $1 - \frac{\alpha}{2}$  de la loi normale centrée réduite.

### Test global de Spearman

Dans cette partie, nous testons la non-corrélation entre les coefficients de développement successifs sur le triangle dans son ensemble.

Soit  $n$  le nombre d'années de développement ou d'années de survénance (supposés être égaux)

Pour chaque colonne  $j$  tel que  $n - j \geq 4$ , nous déterminons les coefficients de corrélation de Spearman par paires de colonnes.

Pour tout  $j \leq n - 4$ ,  $E(\rho_{j,j+1}) = 0$  et  $V(\rho_{j,j+1}) = \frac{1}{n-j-2}$

Nous cherchons à calculer une moyenne pondérée des  $\rho_{j,j+1}$  avec une variance minimale.

Nous choisissons  $\omega_j = \frac{n-j-2}{\sum_{j=1}^{n-4} n-j-2}$

Le coefficient obtenu est :

$$\rho = \sum_{j=1}^{n-4} \frac{(n-j-2)\rho_{j,j+1}}{\sum_{j=1}^{n-4} n-j-2}$$

Avec :

$$E(\rho) = 0$$

Comme les  $\rho_{j,j+1}$  ne sont pas corrélés du fait de la non-corrélation entre les coefficients de développement successifs :

$$V(\rho) = \sum_{i=1}^{n-4} \frac{(n-j-2)V(\rho_{j,j+1})}{(\sum_{j=1}^{n-4} n-j-2)^2} = \frac{1}{\sum_{j=1}^{n-4} n-j-2}$$

Nous supposons que  $\rho$  suit approximativement une loi normale. Et la statistique est construite de la même façon que pour le test de significativité du Tau de Kendall.

#### 5.4.2 Analyse d'effets d'année calendaire

L'indépendance des différentes années de survénance peut être perturbée par la présence d'effets calendaires liés à des changements dans la gestion des sinistres.

Le test d'effets d'année calendaire permet de détecter de telles perturbations.

Soit les éléments d'une année calendaire qui se présente au sein d'un triangle de développement par une diagonale :

$$D_j = \{C_{j,1}, C_{j-1,2}, \dots, C_{1,j}\}$$

Les coefficients de développement qui dépendent des éléments de  $D_j$  sont regroupés en 2 ensembles :

$$A_j = \left\{ \frac{C_{j,2}}{C_{j,1}}, \frac{C_{j-1,3}}{C_{j-1,2}}, \dots, \frac{C_{1,j+1}}{C_{1,j}} \right\} = \{f_{j,1}, f_{j-1,2}, \dots, f_{1,j}\}$$

$$A_{j-1} = \left\{ \frac{C_{j-1,2}}{C_{j-1,1}}, \frac{C_{j-2,3}}{C_{j-2,2}}, \dots, \frac{C_{1,j}}{C_{1,j-1}} \right\} = \{f_{j-1,1}, f_{j-2,2}, \dots, f_{1,j-1}\}$$

Une perturbation des éléments de  $D_j$  par des effets d'année calendaire se répercute sur les éléments des deux ensembles. En effet, une hausse des éléments de  $D_j$  entraîne une hausse des éléments de  $A_{j-1}$  et une baisse des éléments de  $A_j$  et inversement.

L'effet d'année calendaire est testé en suivant les étapes ci-dessous :

1. Pour chaque colonne  $k$ , nous marquons les coefficients de développement par un  $G$  s'ils sont supérieurs à la médiane de la colonne et par un  $P$  s'ils y sont inférieurs. Nous éliminons les coefficients égaux à la médiane. Chaque coefficient non éliminé a donc 50% de chances d'appartenir à un des deux groupes.
2. Pour chaque diagonale  $A_j$  comportant au moins 2 coefficients marqués, nous comptons le nombre  $P_j$  de  $P$  et  $G_j$  de  $G$ .
3. Nous déterminons  $Z_j = \min(P_j; G_j)$
4. Sous l'hypothèse  $H_0$  : il n'y a pas d'effets d'année calendaire,  $P_j$  et  $G_j$  suivent une loi binomiale de paramètres  $N_j = P_j + G_j$  et  $p = 1/2$ . L'espérance et la variance de  $Z_j$  s'écrivent :

$$E(Z_j) = \frac{N_j}{2} - \left( \frac{N_j - 1}{m_j} \right) \frac{N_j}{2^{N_j}} \quad \text{avec } m_j = \frac{N_j - 1}{2}$$

$$Var(Z_j) = \frac{N_j(N_j - 1)}{4} - \left( \frac{N_j - 1}{m_j} \right) \frac{N_j(N_j - 1)}{2^{N_j}} + E(Z_j) - (E(Z_j))^2$$

5. La variable testée  $Z$  est la somme des  $Z_j$   
Sous  $H_0$ , les  $Z_j$  sont non corrélés et donc :  
$$E(Z) = \sum_j E(Z_j) \quad \text{et} \quad Var(Z) = \sum_j Var(Z_j)$$
6. Sous  $H_0$ ,  $Z$  suit approximativement une loi normale et le test se fait en comparant la valeur absolue de la statistique  $\frac{Z - E(Z)}{\sqrt{Var(Z)}}$  au quantile  $1 - \frac{\alpha}{2}$  de la loi normale centrée réduite.

#### 5.4.3 Test de Kolmogorov-Smirnov à deux échantillons

L'une des hypothèses de la méthode Bootstrap consiste à supposer que les résidus échantillonnés sont identiquement distribués.

Cette hypothèse est vérifiée, en séparant les résidus en deux groupes de quantité proche, selon trois axes :

- Entre les années de survénances récentes et anciennes ;
- Entre les années de développement récentes et anciennes ;
- Entre les années calendaires récentes et anciennes.

Ces deux groupes doivent avoir la même fonction de répartition.

Pour le vérifier, le test de Kolmogorov-Smirnov est utilisé. Ce dernier permet de tester la significativité de l'écart en valeur absolue entre les fonctions de répartition empiriques des deux groupes de résidus.

Les différentes étapes du test sont :

Soit  $n$  le nombre d'années de développement ou d'années de survénance (supposés être égaux)

1. Les résidus sont séparés par année de survénance, par année de développement et par année calendaire en deux groupes de quantité proche.

Pour les années de survénance, le moment à partir duquel les deux groupes de résidus sont séparés est donné par  $L$  qui est l'arrondi à l'unité de :

$$n - 1 - \frac{\sqrt{1 + 2 * n * (n - 1)} - 1}{2}$$

Pour les années de développement, le moment à partir duquel les deux groupes de résidus sont séparés est donné par  $C$  qui est l'arrondi à l'unité de :

$$n - 1 - \frac{\sqrt{1 + 2 * n * (n - 1)} - 1}{2}$$

Pour les années de calendaires, le moment à partir duquel les deux groupes de résidus sont séparés est donné par  $D$  qui est l'arrondi à l'unité de :

$$\frac{\sqrt{1 + 2 * n * (n - 1)} - 1}{2}$$

2. Pour chaque groupe  $k_i$  de chaque type de séparation  $k$ , nous comptons pour chaque résidu  $r_{i,j}$  du triangle, le nombre  $N_{k_i}(r_{i,j})$  de résidus du groupe inférieurs ou égaux à ce résidu. Nous le divisons ensuite par le nombre  $n_{k_i}$  de résidus du groupe pour obtenir la fonction de répartition empirique du groupe :

$$F_{k_i}(r_{i,j}) = \frac{N_{k_i}(r_{i,j})}{n_{k_i}}$$

3. Pour chaque type de séparation et pour chaque résidu du triangle, nous calculons :

$$d_k(r_{i,j}) = |F_{k_1}(r_{i,j}) - F_{k_2}(r_{i,j})|$$

4. La statistique de test de Kolmogorov-Smirnov est basée sur l'écart le plus grand écart en valeur absolue entre les fonctions de répartitions empiriques des deux groupes :

$$D_k = \sup_{i,j} d_k(r_{i,j})$$

5. Pour ne pas rejeter l'hypothèse  $H_0$  : les distributions sont identiques , à un seuil  $\alpha = 5\%$

$$|D_k| \leq D_{k,\alpha}$$

Avec

$$D_{k,\alpha} = c(\alpha) \sqrt{\frac{n_{k_1} + n_{k_2}}{n_{k_1} n_{k_2}}}$$

$c(\alpha)$  est tel que :

$$P \left[ \sup_x |F_n(x) - F(x)| > \frac{c(\alpha)}{\sqrt{n}} \right] \xrightarrow{n \rightarrow \infty} \alpha(c) = 2 \sum_{r=1}^{\infty} (-1)^{r-1} \exp(-2c^2 r^2)$$

Pour  $\alpha = 5\%$   $c(\alpha)$  vaut 1.36

## 5.4.4 Application

### 5.4.4.1 Hypothèse de linéarité

L'hypothèse de linéarité est vérifiée par des tests statistiques et graphiques.

#### Test graphique :

L'hypothèse est vérifiée à l'aide d'une étude graphique qui consiste à représenter pour chaque année de développement  $j$  l'ensemble des couples  $(C_{i,j}, C_{i,j+1})$ . Ces couples doivent être alignés sur une droite passant par l'origine et de pente  $f_j$ .

Sur l'annexe 3 nous avons les différents graphes pour chaque pair d'année de développements successives.

La linéarité est graphiquement acceptée chaque année de développement.

#### Test de régression linéaire :

Un test statistique de régression linéaire sur les couples  $(C_{i,j}, C_{i,j+1})$  permet de compléter l'analyse graphique à l'aide des différents coefficients  $R^2$  obtenus.

$R^2$  appelé coefficient de détermination est utilisé en statistiques pour juger la qualité d'une régression linéaire. Les résultats obtenus sont les suivants :

Couples des montants cumulés	Coefficient $R^2$ de la régression linéaire passant par l'origine
(Ci1,Ci2)	0,93
(Ci2,Ci3)	0,97
(Ci3,Ci4)	0,98
(Ci4,Ci5)	0,99
(Ci5,Ci6)	0,99
(Ci6,Ci7)	0,99
(Ci7,Ci8)	0,98
(Ci8,Ci9)	0,99
(Ci9,Ci10)	0,98
(Ci10,Ci11)	0,98
(Ci11,Ci12)	0,99
(Ci12,Ci13)	0,99
(Ci13,Ci14)	1,00
(Ci14,Ci15)	0,99
(Ci15,Ci16)	1,00
(Ci16,Ci17)	1,00
(Ci17,Ci18)	1,00
(Ci18,Ci19)	0,99
(Ci19,Ci20)	0,96
(Ci20,Ci21)	0,97
(Ci21,Ci22)	0,99

Figure 5-13: Coefficient  $R^2$  de la régression linéaire sans constante

Nous observons que tous les coefficients  $R^2$  sont proches de 1 avec une moyenne de 98% ce qui ne remet pas en cause l'hypothèse de linéarité.

### Test analytique :

L'hypothèse de linéarité est renforcée par un test de non-corrélation des coefficients de développement successifs.

Nous avons les résultats suivants :

T	E(T)	Var(T)	Intervalle à 95%	Hypothèse
0,065784905	0	0,0040	-0,123 0,123	Acceptée

Figure 5-14: Test global de Spearman

Le test de Spearman sur le triangle dans son ensemble accepte l'hypothèse de linéarité à un niveau de confiance de 95%.

Test de signification de rho de Spearman		Paires d'années de développement																			
		(1,2)	(2,3)	(6,7)	(7,8)	(8,9)	(9,10)	(10,11)	(11,12)	(12,13)	(13,14)	(14,15)	(15,16)	(16,17)	(17,18)	(18,19)	(19,20)	(20,21)	(21,22)	(22,23)	
rho de spearman		0,025	0,160	0,043	0,233	0,303	-0,079	-0,116	-0,106	-0,147	-0,045	-0,044	0,200	-0,230	0,179	0,057	-0,100	0,890	1,000		
Statistique		0,113	0,724	0,194	1,214	1,183	-0,284	-0,406	-0,423	-0,463	-0,137	-0,120	0,540	-0,600	0,406	1,144	0,174	1,885	---		
p value		0,923	0,473	0,843	0,217	0,241	0,779	0,653	0,668	0,643	0,883	0,883	0,443	0,533	0,603	0,173	0,863	0,003	0,003		
Intervalle (-/+)		2,080	2,086	2,120	2,131	2,145	2,160	2,173	2,201	2,228	2,262	2,296	2,355	2,447	2,511	2,576	3,188	4,393	12,705		
Hypothèse		Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Rejetée		

Test de signification de tau de Kendall		Paires d'années de développement																			
		(1,2)	(2,3)	(6,7)	(7,8)	(8,9)	(9,10)	(10,11)	(11,12)	(12,13)	(13,14)	(14,15)	(15,16)	(16,17)	(17,18)	(18,19)	(19,20)	(20,21)	(21,22)	(22,23)	
tau b de Kendall		0,036	0,100	0,020	0,067	0,143	-0,077	-0,103	0,093	-0,091	-0,019	-0,091	0,111	-0,143	0,143	0,000	0,667	1,000			
Statistique		0,237634101	0,648552385	0,151633103	0,30627181	0,44072054	0,742207483	0,353214451	0,488071101	0,411815831	0,077845834	0,2653321	0,4171023	0,4346073	0,4505644	1,6308061	0	1,323152	---		
p value		0,818	0,163	0,883	0,803	0,713	0,453	0,653	0,603	0,603	0,883	0,883	0,443	0,533	0,603	0,173	0,863	0,003	0,003		
Intervalle (-/+)		1,36	1,36	1,36	1,36	1,36	1,36	1,36	1,36	1,36	1,36	1,36	1,36	1,36	1,36	1,36	1,36	1,36	1,36		
Hypothèse		Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Rejetée		

Test de signification de coefficient de Pearson		Paires d'années de développement																			
		(1,2)	(2,3)	(6,7)	(7,8)	(8,9)	(9,10)	(10,11)	(11,12)	(12,13)	(13,14)	(14,15)	(15,16)	(16,17)	(17,18)	(18,19)	(19,20)	(20,21)	(21,22)	(22,23)	
coefficient de corrélation		-0,016163616	0,168223058	-0,04393281	0,144333148	0,12222144	0,020653461	-0,0210055	-0,14445312	0,21711011	0,07953852	-0,18323	0,230412	-0,33324	0,18453	-0,15513	-0,15713	0,314042	-0,290141		
Statistique		0,28084031	0,30727171	0,27735511	1,42077066	1,053407834	0,074473534	0,523331443	0,484195031	0,85283373	0,239517302	0,474031	0,626471	1,56475	0,443583	0,335663	0,236342	0,895311	7,297363		
p value		0,78163	0,715	0,78503	0,153	0,283	0,943	0,963	0,673	0,393	0,813	0,643	0,503	0,163	0,153	0,403	0,843	0,003	0,003		
Intervalle (-/+)		2,080	2,086	2,120	2,131	2,145	2,160	2,173	2,201	2,228	2,262	2,296	2,355	2,447	2,511	2,576	3,188	4,393	12,705		
Hypothèse		Acceptée	Acceptée	Acceptée	Acceptée	Rejetée	Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Acceptée	Rejetée		

Figure 5-15: Tests de corrélation pour la linéarité

L'hypothèse de non-corrélation entre facteurs de développement est rejetée pour un nombre restreint de couples à savoir (8,9), (20,21) et (21,22).

Les résultats des tests graphiques et du test global de Spearman sont toutefois privilégiés pour conclure que l'hypothèse de linéarité n'est pas remise en cause.

### 5.4.4.2 Hypothèse de variance

L'hypothèse de variance suppose que :

$\forall i = 1, \dots, N, \forall j = 1, \dots, N - 1$ , il existe un paramètre réel  $\sigma_j$  tel que :

$$Var(C_{i,j+1}|C_{i,j}) = \sigma_j^2 \cdot C_{i,j}$$

### Test graphique :

L'hypothèse de variance peut être vérifiée à l'aide d'un graphique représentant les résidus normalisés en fonction des montants cumulés  $C_{i,j}$ . L'hypothèse est rejetée si les résidus présentent une structure non-aléatoire.

Les graphiques sont disponibles en annexe 4.

À la lecture des tests graphiques, globalement, il n'y a pas de forte structure non-aléatoire. L'ensemble des observations graphiques sont valides et cohérentes avec l'acceptation de l'hypothèse de variance.



### Test analytique :

Une étude analytique de l'indépendance entre les résidus du modèle en utilisant les tests de Spearman, Kendall et coefficient de corrélation de Pearson. Le tableau ci-dessous présente les résultats des tests statistiques :

Tests par paires de colonnes		
Test de Spearman	Test de kendall	Coefficient de corrélation de Pearson
Rejetée pour la colonne 22	Rejetée pour les colonnes 17,22	Rejetée pour les colonnes 3 et 20

Figure 5-16: Tests analytiques pour la variance

Nous remarquons que le test est rejeté pour certaines années de développement, pour les années qui dépassent 20 cela est expliqué par le nombre de points faibles. Toutefois, ce résultat couplé au graphique des résidus ne révèle pas de tendances particulières qui remettraient en cause l'hypothèse de variance.

#### 5.4.4.3 Hypothèse d'indépendance des années de survénance

Cette hypothèse est vérifiée par un test graphique et ensuite un test analytique.

### Test graphique :

Ce test consiste à représenter la  $j^{\text{ème}}$  colonne du triangle, constituée des coefficients de développement qui permet le passage de l'année de développement  $j$  à l'année  $j+1$ , en fonction des années de survénance.

L'indépendance des années de survénance signifierait que pour une colonne  $j$  du triangle, les coefficients oscillent autour du coefficient de passage moyen.

Les graphiques sont disponibles en annexe 5.

À la lecture des graphiques, les coefficients oscillent autour de la moyenne.

### Test analytique :

L'hypothèse est également vérifiée par un test statistique d'absence d'effets d'années calendaires.

Le résultat du test est le suivant :

Z	E(Z)	Var(Z)	Intervalle de confiance		Hypothèse
111	112,3	25,8	102,3	122,2	Acceptée

Figure 5-17: Test d'effets d'année calendaire

Au vu du résultat du test statistique, l'hypothèse d'indépendance des années de survénance est acceptée.

#### 5.4.4.4 Hypothèse d'identique distribution des résidus

Cette hypothèse est vérifiée via le test de Kolmogorov-Smirnov. Les résidus sont classés en deux groupes selon les 3 axes :

- Années de survenance récentes et anciennes ;
- Années de développement récentes et anciennes ;
- Années calendaires récentes et anciennes.

Le test fournit les résultats suivants :

Axe de séparation des résidus		
Années d'accident	Années de développement	Années calendaire
Acceptée statistique : 0,142 Valeur critique: 0,157	Acceptée statistique : 0,09 Valeur critique: 0,157	Acceptée statistique : 0,132 Valeur critique: 0,157

Figure 5-18: Test de Kolmogorov-Smirnov

L'hypothèse d'identiques distributions des résidus est vérifiée.

#### 5.4.4.5 Hypothèse d'indépendance des résidus

L'indépendance des résidus est d'abord testée graphiquement à partir de 3 graphes représentant les résidus en fonction des années de survenance, de développement et d'années calendaires.

Ensuite, un test d'absence de corrélation est effectué de deux façons comme ce que nous avons fait pour l'hypothèse de linéarité.

##### Test graphique :

L'hypothèse n'est pas rejetée si les résidus ne présentent pas de structures et ne manifestent pas de tendance.

Les résultats du test graphique se trouvent en annexe 6.

Au vu des graphiques, l'hypothèse d'indépendance n'est pas rejetée.

##### Test analytique :

Les résultats des tests analytiques sont les suivants :

- Axes des années de survenance :

Test global	Tests par couples de résidus		
Test global de Spearman au seuil 5%	Années d'accident	Années de développement	Années calendaire
Acceptée statistique : 0,031 Intervalle : [-0,123;0,123 ]	Rejetée pour les colonnes 12,16	Rejetée pour les colonnes 3,12,16	Rejetée pour les clones 2,3,12,16

Figure 5-19: Indépendance des résidus par année de survenance

- Axes des années de développement :

Test global	Tests par couples de résidus		
Test global de Spearman au seuil 5%	Années d'accident	Années de développement	Années calendaire
Rejetée statistique : 0,059 Intervalle : [-0,042;0,042 ]	Rejetée pour la colonne 21	Rejetée pour la colonne 21	Rejetée pour les colonnes 8,20

Figure 5-20: Indépendance des résidus par année de développement

- Axes des années calendaires :

Test global	Tests par couples de résidus		
Test global de Spearman au seuil 5%	Années d'accident	Années de développement	Années calendaire
Acceptée statistique : 0,019 Intervalle : [-0,123;0,123 ]	Rejetée pour les colonnes 4,5,6	Rejetée pour les colonnes 4,6	Rejetée pour les colonnes 4,16,17,21

Figure 5-21: Indépendance des résidus par année calendaire

Les tests statistiques ci-dessus rejettent l'hypothèse d'indépendance des résidus sur certains axes.

Même si le caractère i.i.d des résidus est rejetée, nous allons voir dans la partie suivante si l'approche actuelle avec Bootstrap peut être maintenue en comparant ses résultats avec la formule fermée de Merz-Wüthrich.

## 5.5 Comparaison Bootstrap / Merz-Wüthrich

Dans cette partie, nous allons comparer les résultats de la méthode Bootstrap avec ceux issus de la formule fermée de Merz-Wüthrich.

Le tableau suivant présente les résultats de la formule fermée de Merz-Wüthrich et l'approche Bootstrap hors retraitement :

Lob Solvabilité 2	Best Estimate vision RA	L'approche Merz-Wüthrich		L'approche Bootstrap		Ecart relatif sur l'erreur de prédiction	Ecart des volatilités
		Erreur de prédiction à 1 an	Volatilité à 1 an	Erreur de prédiction à 1 an	Volatilité à 1 an		
MTPL	1535,5	60,5	3,94%	60,2	3,92%	0,6%	0,02%

Figure 5-22: Comparaison Bootstrap / Merz-Wüthrich

Nous remarquons un écart faible d'environ 0,02% pour la volatilité à 1 an entre les deux approches.

Cet écart ne remet pas en cause la pertinence du modèle et nous concluons que la méthode Bootstrap converge bien vers la formule fermée de Merz-Wüthrich.

## 5.6 Analyse de variation N/N-1

Nous avons retracé l'évolution du SCR du réserve pour la LoB MTPL.

L'évolution entre 2019 et 2020 a été rationalisée en isolant plusieurs effets :

- L'effet assiette du BE S2 net ;
- L'effet de la répartition des cédantes dans le BE actuariel total, qui pondère les coefficients de variation ;
- L'effet de la volatilité des CR, en isolant l'effet pur de l'écart-type et l'effet du rapport du BE implicite / BE actuariel ;

Plus précisément, le graphe en cascade ci-dessous est construit de la manière suivante :

- « Assiette BE S2 » : mise à jour de du BE S2 net total
- « Répartition assiette de BE » : en figeant le BE actuariel total et le coefficient de variation par cédante, mise à jour de la ventilation des cédantes.
- « Vol CR – EC brut » : mise à jour de l'écart-type en neutralisant l'effet de la variation du rapport BE implicite / BE actuariel. Le BE actuariel N-1 est, pour cette étape, corrigé du facteur (BE implicite N / BE implicite N-1). Le BE actuariel de GAN Assurances et de la construction est recalculé afin de conserver une répartition idoine.
- « Vol CR – Rapport BE implicite / BE actuariel » : mise à jour du BE actuariel.
- « Volatilité GAN » : mise à jour de la volatilité de GAN Assurance traitée en formule standard

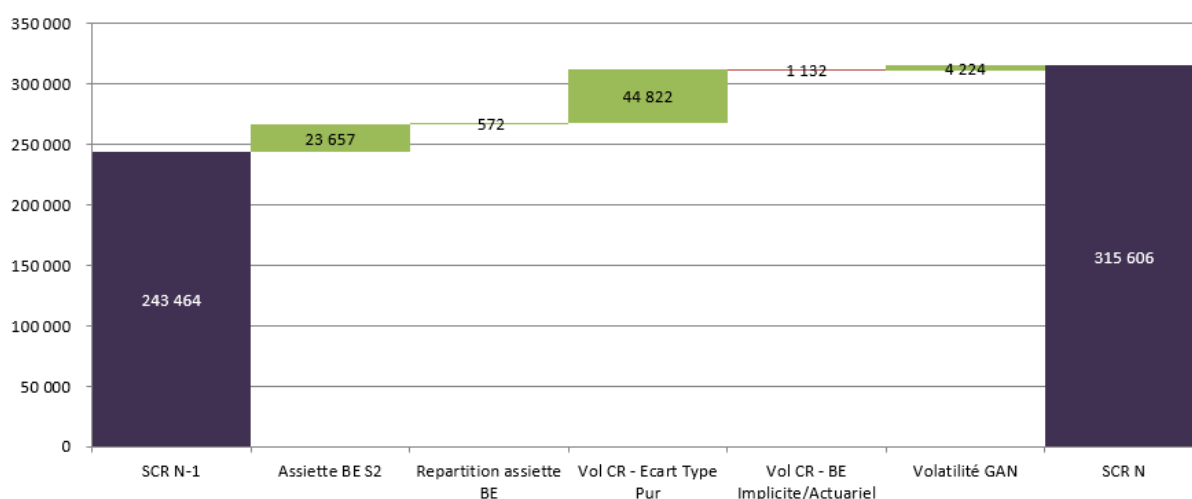


Figure 5-23: Analyse de variation N/N-1

Le SCR augmente de +72,1 M€ (+30%). Cette variation est liée d'une part à l'augmentation importante du BE S2 +150,7 M€ (+9,7 %) et d'autre part, à l'augmentation de (+1,1%) sur le coefficient de variation. Nous précisons que le montant du BE S2 intègre les cessions GAN Assurances traitées avec le coefficient de formule standard.

Cette croissance de la CoV résulte principalement de l'augmentation du coefficient de variation implicite des Caisses Régionales qui passe de 5,2% à 6,3% (due à l'écart-type qui passe de 76,1 M€ à 104 M€).

## 5.7 Analyse des boni/mali et Backtesting

Nous analysons les boni/mali pour la LoB MTPL entre fin 2019 et fin 2020.

Ces boni/mali se décomposent entre :

- Un écart d'expérience : différence entre la diagonale 2020 anticipée dans le modèle à fin 2019 et celle réellement constatée dans le triangle à fin 2020 ;
- Un écart de modèle/hypothèse : il s'agit des boni/mali générés mécaniquement par le modèle de projection ou par un changement de modèle (et non directement liés à la volatilité) ;
- Des boni mali comptables relatifs aux segments de revue de la ligne d'activité Solvabilité 2 non projetés par un modèle de Chain-Ladder et pour lesquels les provisions Best Estimate retenues correspondent aux provisions comptables.

Le boni/mali total est estimé sur la somme des BE (y compris segments non modélisés et donc repris à leur évaluation dossier/ dossier) évalués par les entités à fin 2019 et à fin 2020. Il est positionné graphiquement sur la distribution des boni-mali (distribution Lognormale centrée sur 0 et ayant comme écart-type l'erreur de prédiction de la LoB dans le cadre du risque de réserves 2019) afin de mesurer le quantile auquel il correspond.

Les résultats pour la LoB MTPL sont les suivants :

<b>Best Estimate vu fin 2019</b>	<b>1 342 440,9</b>
<b>Sinistres payés 2020</b>	<b>-126 681,3</b>
<b>Ecart modèle</b>	<b>146 446,5</b>
<b>Ecart expérience</b>	<b>0,0</b>
<b>Boni/Mali comptable</b>	<b>0,0</b>
<b>Best Estimate survenance 2020</b>	<b>194 983,9</b>
<b>Best Estimate vu fin 2020</b>	<b>1 557 190,1</b>

Figure 5-24: Analyse d'écart provisions Best Estimate

L'analyse de variation du Best Estimate entre 2019 et 2020 met en avant un mali de 146,4 M €.

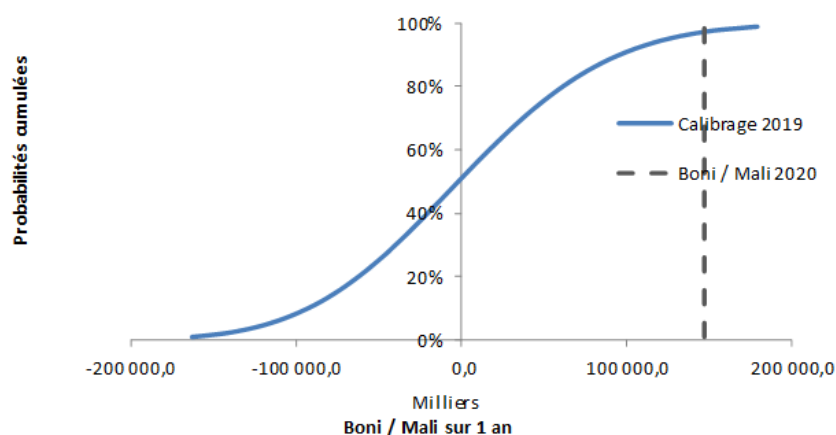


Figure 5-25: Distributions des réserves à horizon 1 an en 2019

Ce montant de mali se situe à un quantile d'ordre 65% sur la distribution des réserves à horizon 1 an telle qu'elle avait été estimée en 2019. Le quantile ne remet pas en cause le modèle d'estimation de la volatilité à 1 an.

Maintenant que nous avons étudié et validé la méthode actuelle de l'équipe modélisation pour le calcul de SCR du risque de réserve à 1 an, nous souhaitons la challenger avec un modèle permettant de modéliser la trajectoire de règlement d'un sinistre. Ce modèle sera appliqué au segment des graves corporelles de GMA pour la ligne d'activité responsabilité civile automobile.

## Chapitre 6

### 6 Modèle de projection individuelle par états

Nous avons vu qu'au sein du modèle interne, le calcul du SCR du risque de réserve dépend de deux paramètres, une volatilité à 1 an qui est estimée à partir d'une modélisation de la charge des sinistres de façon agrégée en utilisant des triangles de développement et un Best Estimate calculé par l'équipe provisionnement qui est utilisé dans le bilan économique.

Nous précisons que GMA construit ses triangles en deux parties selon la structure de réassurance appliquée, l'une pour la couverture en quote-part et l'autre pour la couverture en excédent de sinistre. La vision agrégée de ces données génère une volatilité dans les estimations de BE. En effet, les dossiers des corporels graves sont réévalués à chaque exercice et cette évaluation est fluctuante et dépend de plusieurs facteurs (la consolidation de l'état de santé des victimes, la composition des postes de préjudice du sinistre, l'évolution de la réglementation et la revalorisation des rentes) ce qui diffère des autres dossiers qui ont une évolution plus stable.

Pour toutes ces raisons un modèle individuel multi-états a été testé pour calculer le Best Estimate sur le segment des sinistres graves cédés à GMA.

Afin de réconcilier les modèles de risques et les modèles de provisionnement, nous allons tester le modèle individuel à états pour calculer le risque de réserve à 1 an comme une méthode alternative.

Dans ce chapitre, nous allons d'abord présenter la méthodologie d'utilisation du modèle individuel à états. Ensuite, nous allons voir une mise en pratique du modèle sur des données réelles afin de calculer le risque de réserve à 1 an.

#### 6.1 Modèles de provisionnement individuel

Dans la littérature, plusieurs modèles de provisionnement individuel ont été proposés pour modéliser la dynamique de règlement de chaque sinistre.

L'approche multi-états a d'abord été proposée par Hachemeister (1980) et considérée par Antonio et al. (2016). L'objectif est de construire un modèle stochastique pour les trajectoires des sinistres projetés depuis la survenance jusqu'à la clôture.

Nous rappelons la dynamique de règlement d'un sinistre : un dossier sinistre passe par plusieurs états avant d'être finalement réglé. Le sinistre commence par l'état survenu mais non encore déclaré « IBNR<sup>14</sup> ». Une fois notifié, l'assureur est au courant de la réclamation, mais le premier paiement peut prendre un certain temps avant d'être effectué. Le dossier sinistre est maintenant dans l'état déclaré mais non payé « RBNP<sup>15</sup> » puisqu'aucun paiement n'a encore été effectué. À moins que le dossier soit clôturé sans paiement, le dossier passe par l'état signalé mais pas définitivement réglé « RBNS<sup>16</sup> » lorsqu'il y a un premier paiement et reste dans cet état jusqu'au dernier règlement.

---

<sup>14</sup> Incurred but not reported

<sup>15</sup> Reported but not paid

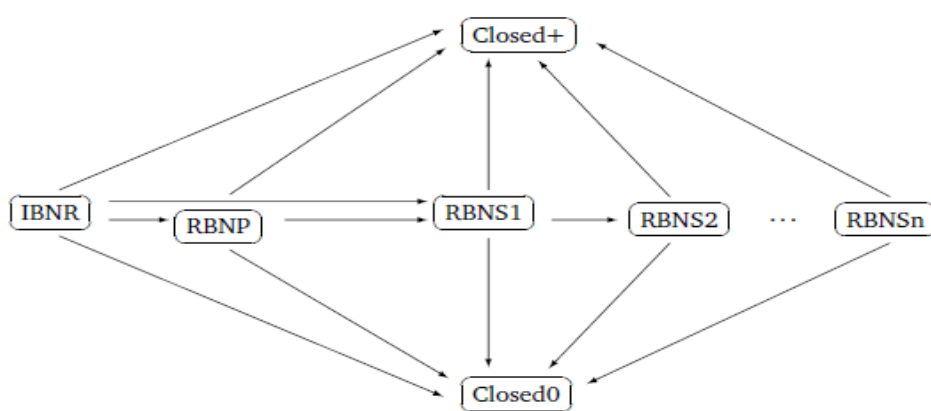
<sup>16</sup> Reported but not settled

Dans notre étude, une approche en temps discret est considérée afin de modéliser le délai de déclaration, l'état RBNP, et les temps entre les différents paiements lors du passage aux états RBNS. Cette approche est décrite dans l'article : Semi-Markov Multistate Individual Loss Reserving Model In General Insurance (2020). Elle sera détaillée dans le paragraphe suivant.

### 6.1.1 Approche multi-états en temps discret

Considérons un sinistre survenu au cours d'une année de survenance donnée, prise comme origine de l'axe de temps. Le processus stochastique  $\{S_t, t = 0, 1, \dots\}$  décrit la dynamique de règlement du sinistre avec comme état initial  $S_0 = IBNR$ .

Le graphique suivant montre les différents états et les transitions possibles entre deux états :



Extrait de l'article Detralytics micro-reserving

Figure 6-1: dynamique de règlement d'un sinistre selon le modèle multi-états

Les différents états du modèle sont les suivants :

#### Etat IBNR

L'état IBNR est l'état initial et le dossier sinistre reste dans cet état jusqu'à ce qu'il soit déclaré à l'assureur. Les nombres d'événements pour chaque année de survenance sont supposés mutuellement indépendants et sont distribués selon une loi Poisson dont les paramètres varient en fonction de l'année de survenance.

Une fois déclaré à l'assureur, le dossier quitte l'état IBNR et plusieurs possibilités sont envisageables :

- soit le sinistre est déclaré mais l'assureur n'effectue aucun paiement de sorte qu'il se déplace à l'état RBNP ;
- soit le sinistre est déclaré et l'assureur effectue un premier paiement et donc le dossier passe directement vers l'état RBNS1 ;
- soit le dossier est clôturé durant l'année de déclaration et quitte l'état IBNR vers :
  - o l'état terminal avec paiement Closed+
  - o ou l'état terminal sans paiement Closed0

## Etat RBNP

Le dossier sinistre reste dans cet état tant que l'assureur n'effectue aucun paiement. Quand il quitte l'état IBNR le dossier peut soit passer aux états terminaux soit aux états RBNS lorsqu'il y'a des paiements.

## Etats RBNS

Si l'assureur effectue au moins un paiement qui n'est pas final, le dossier passe vers la série d'états RBNS1, RBNS2, ... jusqu'à la clôture. Chaque transition est identifiée par un paiement effectué.

Nous supposons que tous les dossiers sont réglés après un nombre maximum de paiements  $n$ , afin d'avoir un nombre fini d'états RBNS.

## Etats de clôture Closed+ et Closed0

Ce sont les états terminaux possibles d'un dossier sinistre.

Dans la partie suivante, nous allons voir la méthodologie de modélisation des transitions et des paiements effectués relatifs au processus stochastique défini.

## 6.2 Modélisation

Un processus stochastique  $\{S_t, t = 0, 1, \dots\}$  est entièrement déterminé par une distribution de probabilité décrivant la trajectoire jusqu'au règlement final du sinistre. Cette distribution est décrite à l'aide des probabilités de transition.

Dans cette partie, nous allons voir la méthodologie d'estimation des probabilités de transition, ainsi que la modélisation des paiements et enfin le calcul de la charge ultime et le risque à 1 an.

### 6.2.1 Probabilités de transition

Afin de modéliser la distribution de probabilité du processus stochastique correspondant au règlement du sinistre, nous allons construire des variables binaires de transition qui seront nécessaires pour estimer les probabilités de transition.

Les variables créées sont :

- Transition : variable binaire indiquant si le sinistre a fait une transition durant l'année ou pas.
- Transition vers l'état clôture : variable binaire indiquant si le sinistre a fait une transition vers un état terminal ou pas.
- Transition vers l'état clôture avec paiement : variable binaire indiquant si le sinistre a fait une transition vers un état terminal avec paiement ou pas.
- Transition vers l'état RBNP : variable binaire indiquant si le sinistre a fait une transition de l'état IBNR vers l'état RBNP ou pas.

La modélisation des probabilités de transition se fait en plusieurs étapes : d'abord la probabilité de sortie d'un état puis la probabilité de transition d'un état à un autre sachant que le sinistre quitte l'état actuel.

Chaque probabilité est modélisée par une régression logistique prenant comme variable réponse une indicatrice indiquant la transition concernée. Les variables explicatives sont l'état actuel dans lequel le sinistre se trouve et le temps déjà passé dans cet état ce qui est cohérent avec un environnement semi-markovien.



## 6.2.2 Survenance des sinistres et délai de déclaration

Puisque l'assureur n'observe les sinistres que lorsqu'ils sont déclarés, il faudrait penser à un modèle permettant d'avoir le nombre de sinistres déclarés chaque année et leur durée de déclaration.

Pour estimer le nombre de sinistres survenus chaque année et le temps qu'il faut pour les déclarer, nous allons créer un triangle avec tous les sinistres, regroupés par année de survenance et délai de déclaration.

Plus précisément, notons  $N_{tj}$  le nombre de sinistres relatifs à une année de survenance  $t$ , et qui ont été déclarés à l'assureur lors du développement  $j$ .

Nous utilisons une spécification multiplicative telle que :

$$E[N_{tj}] = \alpha_t \beta_j \text{ soumis à la contrainte habituelle d'identifiabilité } \sum_{k \geq 1} \beta_k = 1$$

Ensuite, en faisant l'hypothèse que les variables aléatoires  $N_{tj}$  sont indépendantes et suivent une loi de Poisson, les paramètres  $\alpha_t$  et  $\beta_j$  peuvent être estimés du triangle à l'aide d'une régression Poisson.

Ces hypothèses assurent que le nombre total de sinistres  $N_t = \sum_{k \geq 1} N_{tk}$  relatives à une année de survenance  $t$  suit une loi de Poisson de moyenne  $E[N_t] = \alpha_t$  et que  $\beta_j$  est la probabilité qu'un sinistre soit déclaré  $j - 1$  périodes après sa survenance.

## 6.2.3 Modélisation des paiements

Les paiements sont estimés par la modélisation d'un paiement initial et d'une série éventuelle de link ratios décrivant l'évolution des paiements cumulés jusqu'au coût final.

Les paiements accompagnent les transitions entre certains états. En effet, l'assureur effectue un premier paiement non final lorsque le sinistre passe de l'état IBNR ou RBNP au premier état RBNS. Ensuite, à chaque fois que le sinistre passe à un état RBNS suivant, l'assureur effectue un paiement au cours de la période jusqu'au règlement final.

Le montant cumulé payé jusqu'à présent pour un dossier dans l'état  $RBNS_j$  est donné par :

$$C_j = \sum_{k=1}^j P_k$$

Lorsqu'il y'a des recours, certains paiements peuvent être négatifs. Cette particularité peut être corrigée dans les modèles actuariels en utilisant des link ratios  $\lambda_j$  définies de la manière suivante :

$$\lambda_j = \frac{C_{j+1}}{C_j}, j = 1, 2, \dots$$

Dans le cas d'un paiement négatif  $\lambda_j$  est inférieur à 1.

Enfin, nous avons les formules suivantes :

$$C_1 = P_1 \text{ et } C_j = P_1 \prod_{k=1}^{j-1} \lambda_k \text{ pour } j \geq 2$$

Ces link ratios ont une distribution propre à chaque état RBNS.

## 6.2.4 Calcul de la charge ultime et des réserves

La charge ultime et par la suite les réserves peuvent être calculées soit à partir des simulations Monte Carlo en utilisant des lois binomiales pour les transitions entre états et des lois continues pour les paiements et la série des link ratios, soit à partir d'une méthode analytique en utilisant une matrice composée des probabilités de transition et du coût moyen. La méthode analytique est décrite dans l'article Bettonville et al. (2020).

Puisque les réserves sont calculées en termes de moments du portefeuille, nous allons utiliser la méthode analytique. De plus, dans un environnement Markovien, les calculs sont rapidement effectués en utilisant le produit de matrices contenant les probabilités de transitions entre chaque paires d'états.

Puisque nous travaillons sur un modèle Semi-Markovien et la méthode analytique est adaptée au modèles Markoviens, nous allons transformer le processus semi-Markovien en un processus Markovien à l'aide de la propriété suivante :

Un processus stochastique  $\{S_t, t = 0, 1, \dots\}$  est semi-Markovien si et seulement si le processus  $\{(S_t, D_t), t = 0, 1, \dots\}$  est Markovien, où  $D_t$  est la durée de séjour dans l'état  $S_t$ .

Soit  $M$  la matrice contenant les probabilités de transitions dans l'espace Markovien.

Nous posons  $w$  un vecteur de taille égale au nombre d'états Markoviens avec des valeurs nulles sauf celle correspondant à l'état actuel occupé qui est égale à 1.

La probabilité d'atteindre n'importe quel état à l'horizon  $k$  est obtenue à partir de  $w^T M^k$ .

Pour obtenir les coûts correspondants, nous définissons la matrice  $\tilde{M}$  où chaque valeur de  $M$  correspondant à une probabilité de transition entre états est soit multipliée par la valeur moyenne des premiers paiements  $E[P_1]$ , soit la valeur moyenne du link ratio  $E[\lambda_j]$  selon le type de transition.

Nous précisons que pour un sinistre occupant un état RBNS, le premier paiement ou les links ratios observés sont utilisés à la place de leurs valeurs moyennes correspondantes.

Enfin, en sommant les éléments de  $w^T \tilde{M}^k$  on obtient la charge ultime moyenne avec  $k$  le temps restant jusqu'à la clôture de tous les dossiers sinistres du portefeuille. Précisément on a :

- Pour un sinistre dans l'état  $RBNS_j$  en temps  $t$  avec un paiement initial  $P_1$  et des link ratios  $\lambda_1, \dots, \lambda_{j-1}$  son coût moyen ultime est :

$$P_1 \left( \prod_{k=1}^{j-1} \lambda_k \right) \left( w_{RBNS_j}^T \tilde{M}^k \mathbf{1} \right) \text{ où } \mathbf{1} \text{ est le vecteur unitaire}$$

- Pour un sinistre dans l'état RBNP, son coût moyen ultime est :

$$w_{RBNP}^T \tilde{M}^k \mathbf{1}$$

- pour toutes les réclamations IBNR relatives à l'année de survenance  $i$ , leur coût moyen ultime :

$$\alpha_i \sum_{k \geq k_{0+1}} \beta_k (w_{IBNR}^T \tilde{M}^k \mathbf{1})$$

Avec  $i + k_0$  la date d'évaluation,  $k_0 \geq 1$ .

## 6.2.5 Modélisation du risque à un an

Le risque à un an est basé sur le résultat d'évolution des sinistres (CDR) défini comme la somme des paiements effectués l'année suivante et la différence entre les réserves à la fin et au début de l'année calendaire.

La matrice analytique exposée dans la sous partie précédente permet de calculer les réserves au début et à la fin de l'année et donc d'éviter le problème des simulations dans une simulation.

La simulation du CDR se fait selon les étapes suivantes :

1. Simulation des paiements effectués durant l'année calendaire suivante à l'aide du modèle à états.
2. En fin d'année, la perte ultime attendue est calculée en utilisant la matrice analytique dans l'espace Markovien en commençant par l'état occupé par chaque sinistre et les paiements cumulés en fin d'année.
3. Le CDR est obtenu en soustrayant les valeurs obtenues dans les deux étapes précédentes des réserves disponibles au début de l'année.

Répéter ces simulations un grand nombre de fois permet d'obtenir une distribution des CDR et enfin calculer le capital requis correspondant au quantile 99,5% des CDR.

La simulation des paiements durant l'année est divisée en deux étapes :

- Simulation des nouveaux IBNR à l'aide du modèle régression Poisson et calcul du montant total payé au cours de l'année pour ces nouveaux sinistres déclarés.
- Simulation des évolutions de sinistres de l'année pour tout le portefeuille et calcul du montant total payé au cours de l'année.

Le graphique suivant décrit les différentes trajectoires qu'un dossier sinistre peut suivre durant une année :

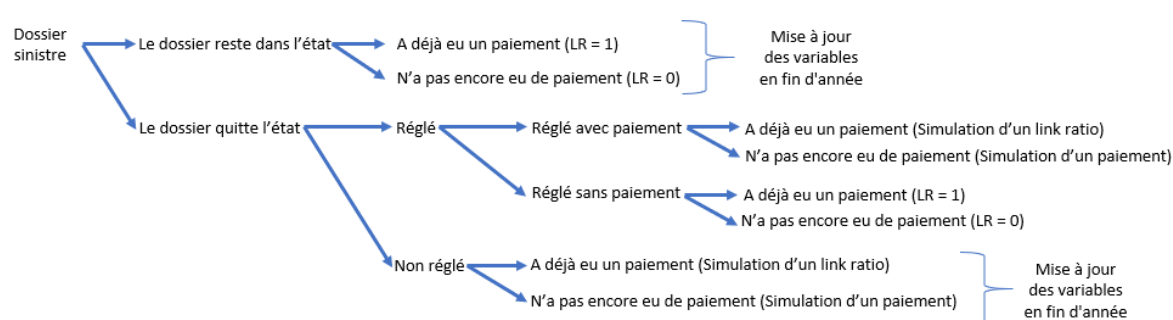


Figure 6-2: Toutes les trajectoires possibles pour un sinistre durant une année

## 6.3 Application

Dans cette partie, nous allons appliquer le modèle sur des données réelles de la ligne d'activité responsabilité civile automobile, en commençant par décrire la base de données utilisées et ensuite nous allons détailler les résultats de la partie modélisation permettant de calculer la charge ultime et le risque de réserves à 1 an.

### 6.3.1 Analyse descriptive des données utilisées dans le modèle

Les données correspondent aux sinistres graves corporels des caisses régionales de Groupama qui dépassent le seuil des 500 k€ à la date du 31/12/2019. La base de données comprend 3112 sinistres et 6158 sous-dossiers victimes liés à ces sinistres. Un sinistre est constitué de 1 à 31 sous dossiers victimes.

Les années de survenances des sinistres encore ouverts sont situées entre 1966 et 2019.

Pour chaque sous-dossier victime nous disposons des paiements et la réserve dossier ainsi que leur date de la façon suivante :

Années d'accident	Paiement			Reserve Dossier	Charge
	Rente	Reglement et Autre	Total		
1966-1997	123,4	375,9	499,3	57,4	556,6
1998	13,2	66,2	79,4	16,9	96,3
1999	9,0	64,2	73,1	9,5	82,6
2000	18,8	67,3	86,1	21,3	107,4
2001	26,9	64,0	90,9	13,6	104,5
2002	18,9	71,9	90,8	17,6	108,4
2003	14,3	54,0	68,3	4,0	72,3
2004	25,7	71,8	97,5	6,2	103,8
2005	6,3	84,0	90,3	9,3	99,6
2006	12,3	62,3	74,7	27,0	101,7
2007	10,1	78,5	88,6	10,7	99,4
2008	5,4	77,5	82,9	7,7	90,6
2009	10,3	77,1	87,4	45,4	132,8
2010	13,3	90,0	103,3	50,6	153,9
2011	7,9	86,2	94,1	38,2	132,2
2012	8,4	89,3	97,7	76,5	174,2
2013	14,5	58,5	73,0	45,8	118,8
2014	1,8	82,6	84,4	316,4	400,8
2015	0,5	43,6	44,1	106,4	150,5
2016	0,0	40,6	40,6	267,3	307,9
2017	0,0	21,0	21,0	205,3	226,3
2018	0,0	11,3	11,3	227,1	238,4
2019	0,0	2,1	2,1	186,6	188,7
<b>TOTAL</b>	<b>341,2</b>	<b>1 739,8</b>	<b>2 081,0</b>	<b>1 766,8</b>	<b>3 847,7</b>

Figure 6-3: Base de données sinistres (M €)

Les gestionnaires-sinistres prévoient une charge future de 1,76 milliard € pour les sinistres déjà survenus. Nous remarquons aussi pour les sinistres survenus après 2016 que les rentes ont un délai de paiement plus long que les paiements classiques.

Les paiements sont composés de 16,4% paiements en rentes et les 83,6% restants sont soit des paiements classiques, soit des paiements en capitaux. Il est à noter qu'il n'est pas possible de distinguer entre les deux types de paiements composant les 83,6% dans la base de données.

Les graphiques suivants décrivent les variables que nous allons utiliser dans la modélisation :

#### Nombre de paiements par sous-dossier victime clôturé :

	Minimum	1 <sup>er</sup> Quantile	Médiane	Moyenne	3 <sup>ème</sup> Quantile	Maximum
Nb de paiements	0	2	3	4,32	6	31

Figure 6-4: Nombre de paiements sous-dossier clôturé

#### Délai de déclaration :

Délai de déclaration	Minimum	1 <sup>er</sup> Quantile	Médiane	Moyenne	3 <sup>ème</sup> Quantile	Maximum
Portefeuille complet	1	1	1	1,21	1	42
Sinistres survenus avant 2000	1	1	1	1,79	1	42
Sinistres survenus après 2000	1	1	1	1,06	1	7

Figure 6-5: Délai de déclaration

Nous remarquons que les sinistres survenus avant 2000 ont un délai de déclaration plus long par rapport aux sinistres survenus après 2000.

#### Délai de clôture :

	Minimum	1 <sup>er</sup> Quantile	Médiane	Moyenne	3 <sup>ème</sup> Quantile	Maximum
Temps	1	4	6	7,9	10	45

Figure 6-6: Temps jusqu'à la clôture

#### Nombre de sous-dossiers victime avec rente :

Nb de rentes	0	1	2	3	4	5
Nb de victimes	5196	560	114	42	6	4
Proportion (%)	87,74	9,46	1,93	0,71	0,1	0,07

Figure 6-7: Nombre de sous-dossiers victime avec rente

Sur les 5922 sous-dossiers victimes dans le portefeuille 87,74 % n'ont pas de rentes.

#### Distribution des valeurs des rentes :

	Minimum	1 <sup>er</sup> quantile	Médiane	Moyenne	3 <sup>ème</sup> Quantile	Maximum	Ecart type
Valeur de rente	841	211.629	455.346	908.586	1.022.677	8.752.550	1.277.763

Figure 6-8: Valeur moyenne rente

#### Distribution des premiers paiements :

PREMIER PAIEMENT	Minimum	1 <sup>er</sup> quantile	Médiane	Moyenne	3 <sup>ème</sup> Quantile	Maximum	Ecart type
Pas de distinction entre les paiements	0	1.000	7.222	42.011	41.543	3.242.038	132.489
Paiements classiques seulement	0	1.000	7.067	38.669	40.918	2.606.944	106.326

Figure 6-9: Distribution des premiers paiements

Dans le tableau ci-dessus se trouve la distribution des premiers paiements avec et sans séparation entre les paiements de rentes représentés par les capitaux constitutifs et les paiements classiques.

Nous remarquons que la distribution est presque semblable dans les deux cas, en revanche nous avons une plus petite volatilité des données lorsque les rentes sont enlevées des paiements.

Etant donné la différence de méthodes de provisionnement entre les rentes et les autres paiements, les deux types de paiements ne doivent pas être traité de la même manière au sein du modèle individuel à états. En effet, les rentes sont calculées à l'aide des méthodes de provisionnement vie avec des tables de mortalité, alors que les autres paiements sont estimés dans un cadre de provisionnement non-vie.

Les informations nécessaires pour avoir les cash-flows du portefeuille pour la partie rente sont les suivantes :

- Age et sexe de la victime
- Montant de la rente
- La table de mortalité et les différents taux utilisés pour les calculs associés au capital constitutif

Vu le manque d'information par rapport à ces variables et le fait que lorsqu'un capital est payé directement il est enregistré comme un paiement quelconque et non comme une rente, un seuil de séparation entre les paiements classiques et les paiements en capital a été choisi par jugement d'expert et la partie rente sera traitée à l'aide d'un modèle binaire indiquant s'il y'a paiement de rente ou pas.

Le modèle que nous allons appliquer consiste à :

- Projeter les sous-dossiers victime sans rente à l'aide d'un modèle multi-état classique
- Projeter les sous-dossiers victime avec rente à l'aide de deux modèles :
  - Un modèle multi-état pour les paiements au-dessous du seuil déterminé
  - Un modèle binaire indiquant s'il y'a au moins une rente pour le sous-dossier victime et le montant correspondant au coût total de la partie rente liée au sous-dossier victime

### 6.3.2 Modélisation des probabilités de transitions

Chaque probabilité est modélisée par une régression logistique prenant comme variable d'intérêt une indicatrice indiquant la transition concernée. Les variables explicatives sont quant à elles les variables reprenant l'état dans lequel le sinistre se trouve et le temps déjà passé dans cet état comme dans l'environnement semi-Markovien.

#### **Depuis l'état IBNR**

- Probabilité de sortie

La variable réponse utilisée égale à 1 s'il y'a une transition vers un autre état, 0 sinon.

Variable explicative : variable temps\_precedent qui correspond au temps passé dans l'état IBNR jusqu'à la date d'évaluation. Comme nous faisons l'hypothèse qu'un sinistre ne peut pas dépasser 2 ans dans l'état IBNR après la survenance, les modalités de cette variable sont groupées de la manière suivante : 0,1,2 et plus.

Lorsque le temps vaut plus de 2, la probabilité de sortir de l'état IBNR est égale à 1.

Les résultats sont les suivants :

```

Coefficients:
              Estimate Std. Error
(Intercept)      2.86918    0.05764
as.factor(Temps_precedent_ExitIBNR)1 -2.07523    0.13413
as.factor(Temps_precedent_ExitIBNR)2 -4.96242    0.12112
              z value Pr(>|z|)
(Intercept)      49.77   <2e-16 ***
as.factor(Temps_precedent_ExitIBNR)1 -15.47   <2e-16 ***
as.factor(Temps_precedent_ExitIBNR)2 -40.97   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 6530.4  on 7141  degrees of freedom
Residual deviance: 3497.1  on 7139  degrees of freedom
AIC: 3503.1

Number of Fisher Scoring iterations: 5

```

- Probabilité de transition vers la clôture

La variable réponse vaut 1 lorsque le sinistre transite vers la clôture, 0 sinon.

Aucune variable explicative significative, nous utilisons un modèle avec constante seulement.

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -4.370      0.117  -37.35   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 795.66  on 5921  degrees of freedom
Residual deviance: 795.66  on 5921  degrees of freedom
AIC: 797.66

Number of Fisher Scoring iterations: 7

```

- Probabilité de transition vers la clôture avec paiement

La variable réponse vaut 1 lorsque le sinistre se déplace vers l'état clôture avec paiement dans l'année, 0 sinon.

De même pas de variable explicative significatives.

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    1.856      0.340    5.459 4.79e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 58.613  on 73  degrees of freedom
Residual deviance: 58.613  on 73  degrees of freedom
AIC: 60.613

Number of Fisher Scoring iterations: 4

```

- Probabilité de transition vers l'état RBNP

Variable réponse vaut 1 si le sinistre transite vers l'état RBNP dans l'année, 0 sinon ?

Variable explicative : temps\_precedent qui correspond au temps passé dans l'état IBNR. Les modalités sont groupées de la façon suivante : 0,1 et plus

```

Coefficients:
                                Estimate Std. Error z value Pr(>|z|)
(Intercept)                   0.63648    0.02826  22.524 < 2e-16
as.factor(Temps_precedent_IBNRRBNP)1 -0.40617    0.11708  -3.469 0.000522

(Intercept)                    ***
as.factor(Temps_precedent_IBNRRBNP)1 ***
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 7581.4  on 5847  degrees of freedom
Residual deviance: 7569.6  on 5846  degrees of freedom
AIC: 7573.6

Number of Fisher Scoring iterations: 4

```

### Depuis les états RBNP et RBNS

- Probabilité de sortie

Variable réponse : vaut 1 s'il y'a une transition vers un autre état, 0 sinon.

Variables explicatives : Etat\_precedent correspondant à l'état occupé par le sinistre avant la transition et la variable Temps\_precedent qui correspond au temps passé dans cet état.

Les modalités sont groupées de la façon suivante pour la première variable : 0,1,2,3,4,5,6 et plus.

Pour le temps les modalités sont les suivantes : 1,2,3,4,5,6 et plus.

```

Coefficients:
                                Estimate Std. Error z value Pr(>|z|)
(Intercept)                   0.17934    0.02904   6.176 6.58e-10 ***
Etat_precedent_ExitRBNPS1     0.42288    0.03606  11.727 < 2e-16 ***
Etat_precedent_ExitRBNPS2     0.55797    0.04014  13.900 < 2e-16 ***
Etat_precedent_ExitRBNPS3     0.39405    0.04293   9.179 < 2e-16 ***
Etat_precedent_ExitRBNPS4     0.29068    0.04723   6.155 7.53e-10 ***
Etat_precedent_ExitRBNPS5     0.24904    0.05401   4.611 4.00e-06 ***
Etat_precedent_ExitRBNPS6     0.37630    0.04123   9.126 < 2e-16 ***
as.factor(Temps_precedent_ExitRBNPS)2 -0.42297    0.03064 -13.807 < 2e-16 ***
as.factor(Temps_precedent_ExitRBNPS)3 -0.89702    0.04179 -21.465 < 2e-16 ***
as.factor(Temps_precedent_ExitRBNPS)4 -1.06727    0.05366 -19.890 < 2e-16 ***
as.factor(Temps_precedent_ExitRBNPS)5 -1.36193    0.05449 -24.996 < 2e-16 ***
as.factor(Temps_precedent_ExitRBNPS)7 -1.85717    0.05127 -36.226 < 2e-16 ***
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 43698  on 31536  degrees of freedom
Residual deviance: 40648  on 31525  degrees of freedom
AIC: 40672

Number of Fisher Scoring iterations: 4

```

- Probabilité de transition vers la clôture

Variable réponse vaut 1 si la transition se fait vers la clôture, 0 sinon.

Variables explicatives : Etat\_precedent et Temps\_precedent.

Les modalités pour l'état sont : 0,1,2,3,4,5 et plus.

Les modalités pour le temps : 1,2,3,4,5 et plus.



```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -1.88935    0.05432  -34.782 < 2e-16 ***
Etat_precedent_c1  0.17665    0.06396   2.762 0.005747 **
Etat_precedent_c2  0.25067    0.06840   3.665 0.000248 ***
Etat_precedent_c3  0.29947    0.07305   4.099 4.14e-05 ***
Etat_precedent_c4  0.33102    0.08039   4.118 3.83e-05 ***
Etat_precedent_c5  0.21635    0.06556   3.300 0.000967 ***
as.factor(Temps_precedent_c)2  0.63163    0.04937  12.793 < 2e-16 ***
as.factor(Temps_precedent_c)3  0.98256    0.06889  14.263 < 2e-16 ***
as.factor(Temps_precedent_c)4  1.38632    0.08711  15.915 < 2e-16 ***
as.factor(Temps_precedent_c)5  1.72721    0.06784  25.462 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 17076  on 16182  degrees of freedom
Residual deviance: 16163  on 16173  degrees of freedom
AIC: 16183

Number of Fisher Scoring iterations: 4

```

- Probabilité de transition vers la clôture avec paiement

La variable réponse vaut 1 lorsque la transition se fait vers la clôture avec un paiement dans l'année, 0 sinon.

Variables explicatives : Etat\_precedent et Temps\_precedent.

Les modalités pour l'état sont : 0, 1 et plus.

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    1.38127    0.10655  12.964 < 2e-16 ***
Etat_precedent_c11 -1.62162    0.10387 -15.612 < 2e-16 ***
as.factor(Temps_precedent_c1)2 -0.55744    0.09264  -6.017 1.78e-09 ***
as.factor(Temps_precedent_c1)3 -0.64453    0.12439  -5.182 2.20e-07 ***
as.factor(Temps_precedent_c1)4 -0.83820    0.15069  -5.563 2.66e-08 ***
as.factor(Temps_precedent_c1)5 -1.17581    0.19223  -6.117 9.56e-10 ***
as.factor(Temps_precedent_c1)6 -2.15280    0.31672  -6.797 1.07e-11 ***
as.factor(Temps_precedent_c1)7 -1.49828    0.16798  -8.919 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 4795.2  on 3568  degrees of freedom
Residual deviance: 4357.6  on 3561  degrees of freedom
AIC: 4373.6

Number of Fisher Scoring iterations: 4

```

### 6.3.3 Modélisation des paiements

Le seuil choisi pour séparer entre les paiements classiques et paiements entrant dans la partie rente est de 75 k€.

#### Premiers paiements

L'équipe modélisation a testé plusieurs lois et le meilleur fit correspond à une loi de mélange avec deux composantes lognormale.

Une loi de mélange fini est définie de la façon suivante :

Supposons un ensemble fini de fonctions de densité de probabilité  $f_1, \dots, f_n$ , ou des fonctions de répartition  $F_1, \dots, F_n$  et des probabilités  $w_1, \dots, w_n$  tel que  $w_i \geq 0$  et  $\sum_{i=1}^n w_i = 1$ .

La loi de mélange est déterminée soit par sa densité  $f$ , soit par sa fonction de répartition  $F$  de la manière suivante :

$$f = \sum_{i=1}^n w_i * f_i$$

$$F = \sum_{i=1}^n w_i * F_i$$

Sur R, le package gamlss.mx avec sa fonction gamlssMXfits permet d'ajuster des modèles de mixture avec nos données.

Les variables explicatives utilisées dans la régression sont la variable Etat\_precedent et Temps\_precedent.

Les modalités pour l'état sont groupées en deux : -1 et 0 (état IBNR et RBNP)

Les modalités pour le temps sont groupées de la façon suivante : 0, 1 et plus.

Les paramètres de la loi ajustées sont détaillés dans le tableau suivant :

	Composante lognormale			Composante lognormale		
Poids	$p_1 = 0,34$			$p_2 = 0,66$		
Paramètres	$\hat{\mu}_1$		$\hat{\sigma}_1$	$\hat{\mu}_2$		$\hat{\sigma}_2$
Coefficients	Constante	10,0882		Constante	6.9346	
	Temps_precedent 1	0,2757	0,6262806	Etat_precedent 0	0,6362	1,574885

Figure 6-10: Loi mélange des premiers paiements

Pour nous assurer du bon ajustement, nous allons appliquer cette distribution aux premiers paiements observés. Ensuite, nous allons voir si la courbe de la fonction de répartition empirique des valeurs calculées est proche de la diagonale.

L'idée de cette approche graphique vient du fait que pour toute fonction de répartition continue  $F_X$ ,

$$\text{Si } X \sim F_X, \text{ alors } F_X(X) \sim Unif[0,1]$$

Nous obtenons la courbe suivante :

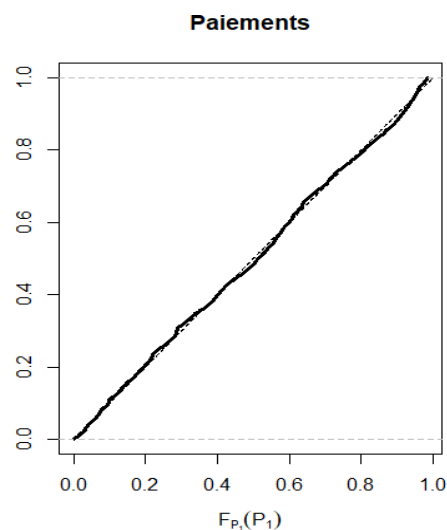


Figure 6-11: Uniformité de la fonction de répartition des premiers paiements

Nous remarquons que la droite de la fonction de répartition empirique s'approche de la diagonale et donc le modèle qui est un mélange de deux lois lognormales ajuste bien les premiers paiements.

## link ratios

En regardant la valeur des premiers paiements en fonction des link ratios qui les suivent, nous obtenons le graphe suivant :

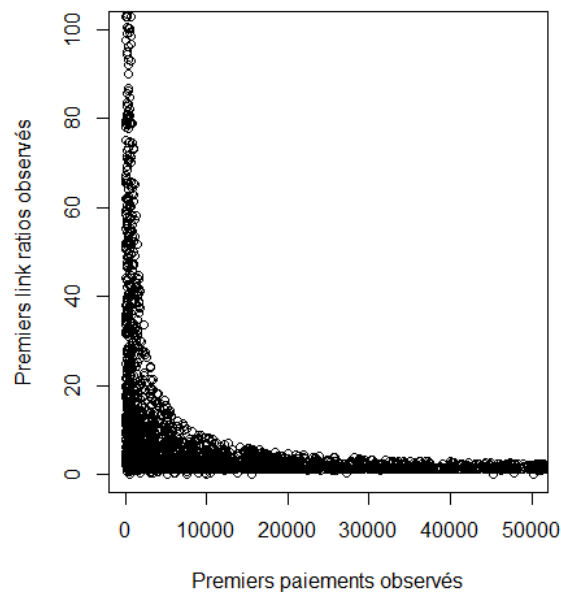


Figure 6-12: Premier paiement en fonction du premier link ratio

Nous observons que les paiements plus petits sont suivis par de grands link ratios alors que les paiements déjà élevés sont suivis par de plus petits link ratios. Afin d'éviter la possibilité d'avoir un grand paiement suivi par un grand link ratio, nous déterminons un seuil qui interviendra dans la modélisation des link ratios.

D'après le graphique ci-dessus le seuil de 30000 est retenu.

- Link ratio 1

Le modèle retenu est le mélange de 3 composantes lognormales et a comme variables explicatives :

- Variable seuil qui vaut 1 lorsque le premier paiement est en dessous de 30000, 0 sinon.
- Variable Temps\_precedent avec les modalités : 1,2,3, 4 et plus.

- Link ratio 2

Le modèle retenu est le mélange de 2 distributions lognormales et a comme variables explicatives :

- Variable seuil qui vaut 1 lorsque le premier paiement est en dessous de 30000, 0 sinon.
- Variable Temps\_precedent avec les modalités : 1,2, 3 et plus.

- Link ratio 3

Le modèle retenu est le mélange de 2 distributions lognormales et a comme variables explicatives :

- Variable seuil qui vaut 1 lorsque le premier paiement est en dessous de 30000, 0 sinon.
- Variable Temps\_precedent avec les modalités : 1,2, 3 et plus.

- Link ratio 4

Le modèle retenu est le mélange de 2 distributions lognormales et a comme variables explicatives :

- Variable seuil qui vaut 1 lorsque le premier paiement est en dessous de 30000, 0 sinon.

- Link ratio 5

Le modèle retenu est le mélange de 2 distributions lognormales et a comme variables explicatives :

- Variable seuil qui vaut 1 lorsque le premier paiement est en dessous de 30000, 0 sinon.

### 6.3.4 Estimation du nombre de sinistres déclarés chaque année

La modélisation du nombre d'IBNR survenu chaque année et leur durée de déclaration nécessite la création d'un triangle de développement du nombre de sinistres avec en lignes les années de survenance et en colonnes les durées de déclaration.

Nous faisons l'hypothèse que la durée maximale avant déclaration est de deux ans.

Le triangle obtenu est en annexe 7.

Nous observons que le nombre des sinistres déclarés n'est pas stable entre certaines années de survenance, pour cette raison nous allons grouper certaines modalités.

Le modèle utilisé est une régression GLM poisson avec :

Variable réponse : nombre de sinistres déclarés

Variable explicative : année de survenance groupées et délai de déclaration.

Les résultats de la régression Poisson sont les suivants :

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    2.53615    0.06469   39.206 < 2e-16 ***
as.factor(AYG)1991  0.99996    0.18095    5.526 3.27e-08 ***
as.factor(AYG)1992  1.44503    0.11415   12.659 < 2e-16 ***
as.factor(AYG)1994  1.85275    0.09025   20.530 < 2e-16 ***
as.factor(AYG)1997  2.84424    0.06607   43.048 < 2e-16 ***
as.factor(RepDelG)1 -3.10602    0.06892  -45.068 < 2e-16 ***
as.factor(RepDelG)2 -3.94117    0.10146  -38.843 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 15789.70 on 140 degrees of freedom
Residual deviance: 784.31 on 134 degrees of freedom
AIC: 1313.2

Number of Fisher Scoring iterations: 5

```

Les nombres moyens ajustés sont les suivants :

	0	1	2
1990	12,63	0,57	0,25
1991	34,33	1,54	0,67
1992	53,58	2,40	1,04
1994	80,55	3,61	1,56
1997	217,11	9,72	4,22
nombre d'IBNR moyen		9,72	5,78

Figure 6-13: nombre moyen d'IBNR déclaré par année de survenance et délai de déclaration

Il reste en moyenne 10 sinistres dans l'état IBNR à être déclaré après 1 an de survenance et 6 sinistres après 2 ans de survenance.

### 6.3.5 Charge ultime et réserve

La charge ultime totale et les réserves sont calculées selon 2 modèles, le premier est un modèle multi-états pour les paiements classiques au-dessous du seuil de 75 k€, le deuxième est un modèle binaire pour les paiements en capital.

#### **Modèle multi-états**

lié au paiements classiques et par la suite les réserves sont calculées à l'aide de la matrice analytique à un horizon  $k$ .

L'horizon choisi est tel que la plupart des dossiers sinistres sont réglés à ce moment-là.

Pour cela nous allons travailler avec la matrice  $M$  contenant les probabilités de transitions dans l'espace markovien.

D'abord, nous allons créer une liste  $L$  contenant un nombre de vecteurs égal au nombre d'états Markovien, chaque vecteur est de taille égale au nombre d'états Markoviens avec des valeurs nulles sauf celle correspondant à l'état actuel occupé qui est égale à 1. Ensuite, nous calculons  $L^T M^k$  afin d'obtenir les probabilités de transition à un horizon de  $k$  années.

L'horizon optimal est choisi de façon à avoir une grande probabilité de clôture à un moment  $k$ .

L'horizon de 20 ans donne en moyenne une probabilité de clôture égale à 94%.

Les résultats du modèle multi-états sont les suivants :

Années d'accident	Paielement	Reserve	Charge
1966-1997	5,8	9,0	14,7
1998	1,2	1,6	2,8
1999	0,8	1,7	2,5
2000	1,5	2,7	4,3
2001	1,3	1,6	2,9
2002	1,2	2,1	3,3
2003	1,2	1,9	3,1
2004	0,8	1,4	2,2
2005	1,7	2,8	4,6
2006	2,3	3,3	5,7
2007	1,4	2,4	3,8
2008	1,1	3,4	4,5
2009	1,8	3,6	5,4
2010	2,7	6,8	9,5
2011	3,5	8,6	12,1
2012	4,0	11,0	14,9
2013	3,7	8,5	12,2
2014	5,9	20,0	25,9
2015	4,2	22,6	26,8
2016	5,7	39,7	45,4
2017	3,7	61,7	65,4
2018	2,1	94,2	96,3
2019	0,7	99,9	100,6
TOTAL	58,3	410,6	468,8

Figure 6-14: Charge ultime et réserve du modèle multi-états (en M €)

Nous obtenons une charge future de 469 M € et un BE de 410 M €.

### Modèle binaire pour la partie rente

La partie rente à rajouter aux résultats du modèle multi-états consiste à simuler s'il y'a un paiement de capital à l'aide d'une loi binomiale ayant comme paramètre la probabilité d'avoir au moins une rente pour le sous-dossier victime. Nous avons essayé d'ajuster la probabilité à l'aide d'une régression logistique et nous retenons un modèle avec constante.

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.05436    0.03315    1.64    0.101

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 5047.6 on 3642 degrees of freedom
Residual deviance: 5047.6 on 3642 degrees of freedom
AIC: 5049.6

Number of Fisher Scoring iterations: 3

```

Chaque sous-dossier victime à une probabilité de 51 % d'avoir au moins une rente.

Enfin, nous avons simulé s'il y'a un paiement en capital pour chaque sous-dossier victime et le montant total de ce paiement est égal à la moyenne des paiements en rente totaux observés pour les sous-dossiers victimes clôturés, car le but est de calculer une charge ultime moyenne.

Afin de comparer entre le modèle que nous avons développé et ce qui est fait actuellement, un modèle Chain Ladder a été construit et va servir comme point de référence. Les résultats des deux modèles sont les suivants :

Années d'accident	Réserve Chain Ladder	Reserve modèle multi-états + binaire	Ecart
1966-1997	57,4	23,1	34,2
1998	15,5	4,1	11,4
1999	9,6	4,2	5,4
2000	23,7	7,2	16,5
2001	18,9	3,1	15,9
2002	24,2	6,3	17,9
2003	8,4	6,8	1,6
2004	13,2	4,1	9,2
2005	16,0	8,1	7,8
2006	33,5	10,9	22,6
2007	17,4	8,0	9,5
2008	13,2	10,7	2,5
2009	53,0	14,0	38,9
2010	54,7	25,5	29,2
2011	43,5	31,4	12,1
2012	78,3	42,5	35,8
2013	46,7	33,0	13,6
2014	317,2	65,0	252,2
2015	98,0	66,1	31,9
2016	248,5	105,7	142,8
2017	194,2	141,9	52,3
2018	209,4	177,7	31,6
2019	195,6	186,7	8,9
<b>TOTAL</b>	<b>1 789,9</b>	<b>986,1</b>	<b>803,8</b>

Figure 6-15: Comparaison des réserves du modèle multi-états + binaire avec Chain Ladder

Nous remarquons que les réserves du modèle que nous avons développé sont plus faibles que les résultats du Chain Ladder. Cela vient du fait d'un manque d'information précise sur les montants qui correspondent à un capital. Nous avons essayé de remédier à cette difficulté en introduisant le seuil de 75 k€. Mais il se peut que certains paiements sélectionnés au-dessus du seuil correspondent encore à des paiements classiques plus élevés, de ce fait le modèle multi-états ne va pas générer en moyenne des paiements plus élevés que le seuil.

Nous constatons aussi que plus l'année de survenance est récente plus le modèle que nous avons développé génère des paiements plus élevés. Cependant, le modèle ne produit pas des montants aussi importants que pour les années récentes ayant une charge dossier extrême par exemple les années de survenance 2014 et 2016 ayant respectivement une charge dossier de 400 M € et 308 M € (cf figure 6-3)

Enfin, nous concluons que modéliser à une maille plus fine à savoir le sous-dossier implique un impact sur les résultats du modèle à cause de manque d'informations précises même si tous les modèles statistiques sont bien ajustés.

### 6.3.6 Calcul du risque à 1 an

Le calcul du risque à 1 an nécessite d'abord de simuler les paiements effectués durant l'année suivante. Ces paiements sont soit liés au sous-dossier déjà présent dans le portefeuille, soit au IBNR seront déclarés pendant l'année.

Pour la partie IBNR le nombre d'IBNR qui seront déclarés pendant l'année est simulé à partir d'une loi de Poisson avec comme paramètre la moyenne obtenue avec le modèle GLM Poisson que nous avons ajusté dans les parties précédentes.

Les lois retenues sont les suivantes :

	Délai de déclaration	
Poisson	1	2
Paramètre	9,7	1,6

Figure 6-16: Nombre moyen d'IBNR déclaré pour l'année calendaire 2020

Ces IBNR et les sous-dossiers victimes déjà présents dans le portefeuille sont projetés selon l'arbre de possibilités défini dans la partie théorique de la modélisation du risque à un an.

Les paiements et link ratios correspondant aux possibilités d'avoir un paiement durant l'année sont simulés à l'aide de la fonction `rlnormMix` du package `EnvStats` de R.

Les paiements liés à la partie rente sont simulés à l'aide d'un modèle fréquence coût moyen ajustés sur les paiements effectués chaque année calendaire.

Enfin, les CDR sont simulés et le quantile à 1 an du risque de réserve avec le modèle multi-états est obtenu.

De plus nous avons effectué un calcul du risque de réserve à 1 an avec la méthode bootstrap à 1 an en appliquant la volatilité à 1 an au niveau du BE calculé par le modèle alternatif (multi-états + binaire) qui servira comme référence, les résultats sont affichés dans le tableau suivant :

	Risque à 1 an
Bootstrap à 1 an avec un CoV de 5,8%	162,57
Modèles à états + binaire	80,7

Figure 6-17: risque à 1 an en M €.

Nous remarquons que le risque à 1 an calculé par la méthode alternative est inférieur à celui calculé par le Bootstrap à 1 an à un même niveau de BE.

Dans cette partie nous avons proposé un modèle alternatif permettant de calculer les réserves et le risque à 1 an des sinistres graves corporelles. Nous remarquons que l'application des modèles multi-états nécessite une bonne qualité de données et des informations complètes sur les sous-dossiers victimes afin d'avoir des résultats plus précis et proches de la réalité. Malgré un modèle multi-états répliquant la vie réelle d'un sinistre grave, l'insuffisance des données constitue une limite à sa pertinence, notamment au niveau de son calibrage.



Dans la partie suivante, nous allons voir comment les sinistres graves de la ligne d'activité responsabilité civile automobile sont traités au sein du risque de prime.

## 7 Validation du calibrage de la charge des graves

Nous avons décrit dans la première partie la méthodologie de calibrage de la charge des graves au sein du modèle interne, elle suit le processus suivant :

1. Détermination du seuil des graves
2. Mise en as-if des graves
3. Calibrage des lois de fréquence et sévérité pour calculer la charge des graves

Dans ce chapitre, nous allons présenter les travaux de validation du calibrage des graves sur la LoB MTPL pour les CR.

### 7.1 Choix du seuil de séparation des sinistres

Le seuil de séparation est clé pour la modélisation du risque de primes car il détermine à la fois les données individuelles du module « sinistres graves » et les données en entrée du module « sinistres attritionnels » qui sont modélisées de façon agrégée.

#### 7.1.1 Méthodologie de l'équipe de modélisation

Les données servant à l'étude correspondent aux données au Q4 2019. Nous nous intéressons aux sinistres graves à date avec un seuil d'extraction déterminé (500 k€ pour les caisses)

L'équipe de modélisation dispose de deux sources de données :

- Un fichier où se trouvent les données de sinistres initiales.
- Les maquettes de mise en as-if des sinistres du modèle interne

La variable étudiée pour le coût du sinistre est la charge nette de recours la plus récente. Cette charge est ultime en considérant que l'effet de la projection est non significatif sur notre étude. Ensuite, cette charge est mise en vision as-if pour l'inflation sectorielle (5% par année).

#### Maille d'étude :

Les données relatives à la LoB 04\_MTPL comprennent deux segments : SRI AUTO RC CORP et SRI AUTO RC MAT.

Lob	Review_segment	Risk_segment	Nombre de sinistres
04_D_MTPL	04_auto_rc_corp 04_auto_rc_mat	SRI AUTO RC CORP SRI AUTO RC MAT	2685

Figure 7-1: Nombre de sinistres LoB MTPL

A partir de ces données de sinistres individuels, nous avons déterminé un nombre de 2558 sinistres individuels qui dépassent le seuil d'extraction comme input de notre étude de seuils.

## Sélection du seuil des graves

Afin de déterminer le seuil des sinistres graves pour la LoB MTPL, l'équipe de modélisation a implémenté des méthodes de la théorie des valeurs extrêmes explicitées dans la première partie du mémoire.

Deux méthodes à interprétation graphique ont été utilisées :

- Mean Excess plot
- Hill plot

Les résultats obtenus sont les suivants :

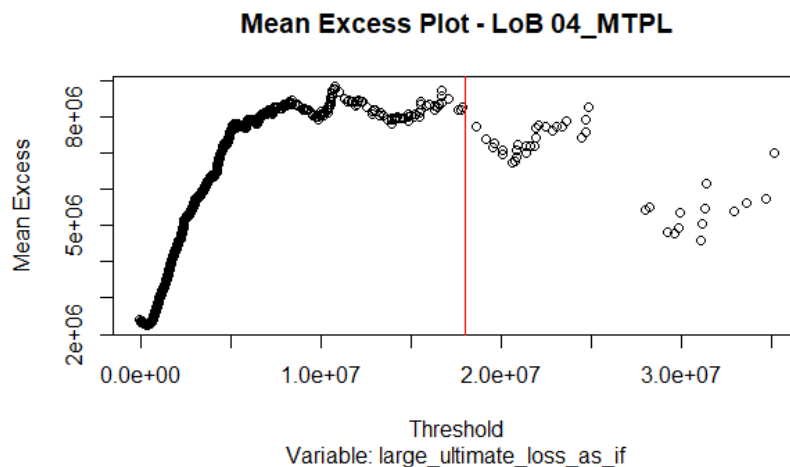


Figure 7-2: Graphique Mean Excess

Nous remarquons que les derniers points du graphique sont très variables et rendent l'interprétation du graphique difficile pour des fins de détermination d'un seuil.

L'équipe de modélisation a donc uniquement conservé les valeurs à gauche de la droite verticale pour analyser le graphique.

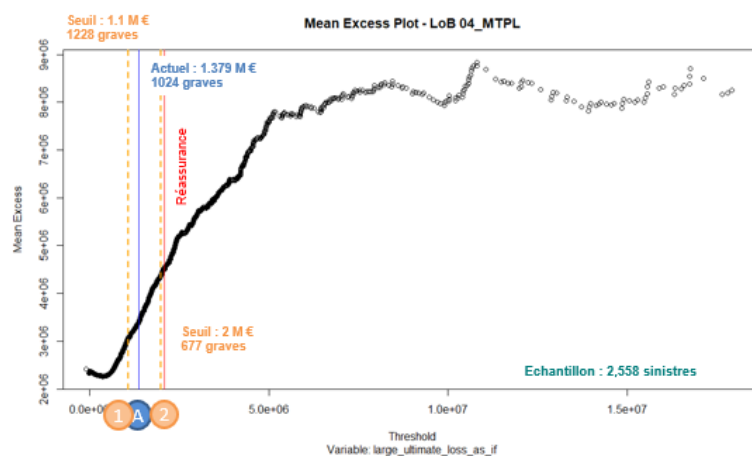


Figure 7-3: Nouveau Graphique Mean Excess

Nous remarquons que les deux seuils proposés par l'équipe de modélisation sont proches au seuil actuel de 1.379 M €.

Le retrait des dernières valeurs dépend de la personne qui réalise l'étude, et est donc susceptible de différer suivant les personnes. Cela peut avoir un impact sur la forme du graphique, son interprétation visuelle et donc sur le choix du seuil.

Donc, de nombreux seuils entre les 2 envisagés pourraient être justifiables. D'où la limite de cette méthode de détermination visuelle.

Nous obtenons aussi les résultats du Hill plot :

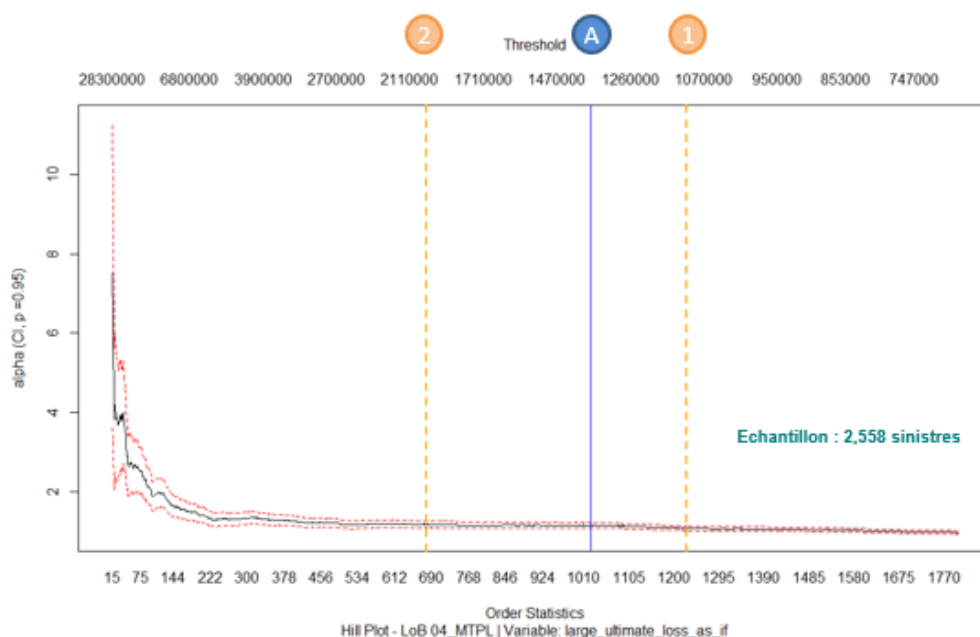


Figure 7-4: Graphique de l'estimateur de Hill

Le seuil optimal est déterminé quand le graphique se stabilise.

Les deux droites 1 et 2 représentent les deux seuils identifiés avec le Mean Excess Plot. Les deux sont acceptables d'après le Hill plot.

D'après ces deux indicateurs (MEP et Hill), le seuil actuel semble pertinent et est conservé.

Lob	Sinistres totaux	Seuil choisi	Nombre de graves	Ancien seuil
04 - MTPL	2558	inchangé	1024	1 379 000 €

Figure 7-5: Graphique Mean Excess

### 7.1.2 Méthodes alternatives pour la détermination du seuil

Les méthodes utilisées par l'équipe de modélisation sont des méthodes « visuelles » qui peuvent conduire à des interprétations différentes selon les experts en charge de leur interprétation.

Dans cette partie, nous allons proposer d'autres méthodes permettant de déterminer le seuil de séparation des graves afin de challenger les résultats obtenus par l'équipe de modélisation.

### 7.1.2.1 Gerstengarbe and Werner plot

Le graphique de Gerstengarbe (Gerstengarbe et Werner 1989) est une méthode non paramétrique qui permet d'identifier pour une série de données le seuil à partir duquel commence la région des valeurs extrêmes.

Concrètement, avec cette méthode, nous nous intéressons à la série des différences :

$$\Delta_i = X_{i+1} - X_i$$

Avec  $X_i$  Les valeurs de notre série ordonnées dans l'ordre croissant.

L'idée de base est qu'il est raisonnable de s'attendre à ce que le comportement des  $\Delta_i$  d'un ensemble de données est différent entre les observations extrêmes et normales. Donc il doit y'avoir un changement dans la série des  $\Delta_i$ , et ce point de changement significatif  $i$  est le point de départ de la région des extrêmes

La statistique utilisée pour tracer le graphique est la série :

$$U_i = \frac{U_i^* - \frac{i(i-1)}{4}}{\sqrt{\frac{i(i-1)(2i+5)}{72}}}, \quad \forall i \in \{1, \dots, n-1\}$$

Avec  $U_i^* = \sum_{k=2}^i n_k$  et  $n_k$  le rang de  $\Delta_k$  dans la série  $\{\Delta_1, \dots, \Delta_k\}$ .

Une deuxième courbe est tracée basée sur le même principe mais cette fois ci sur la série décroissante  $\{\Delta_{n-1}, \dots, \Delta_1\}$ .

Le seuil d'entrée de la zone extrême est le point d'intersection des deux courbes.

En appliquant la méthode sur nos données, nous obtenons le graphique suivant :

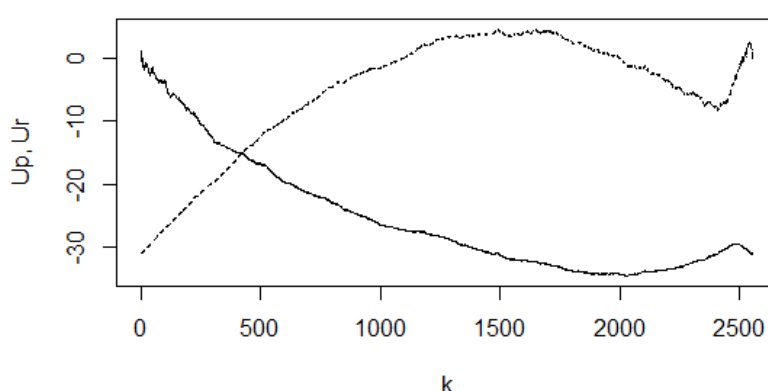


Figure 7-6: Graphique Gerstengarbe and Werner

Nous obtenons un seuil de séparation de 3 043 737 M € .

Nous observons que le seuil obtenu avec le graphique de Gerstengarbe et Werner dépasse celui choisi par l'équipe de modélisation et le seuil de réassurance.

Cela est dû à la différence d'hypothèse des méthodes. En effet, la méthode de Gerstengarbe et Werner est non paramétrique et ne fait pas d'hypothèse sur la distribution des sinistres extrêmes, contrairement aux méthodes issues de la théorie des valeurs extrêmes qui restent les mieux adaptées à notre problématique.

### 7.1.2.2 Estimateur de Hill en minimisant l'erreur quadratique moyenne :

Cette méthode est basée sur l'estimateur de Hill du paramètre de queue  $\xi$ . Cet estimateur est construit en faisant l'hypothèse d'une loi Pareto indexée par le paramètre de forme  $\alpha$  sur les valeurs qui dépassent un certain seuil  $b$ . Donc  $\ln(X) - \ln(b)$  suit une loi exponentielle de moyenne  $\xi$ .

D'où l'estimateur connu :

$$\hat{\xi}_{k,n}^{Hill} = \frac{1}{k} \sum_{i=1}^k \ln(X_{n-i+1:n}) - \ln(X_{n-k:n})$$

Où  $\ln(X_{n-k:n})$  est l'estimateur du seuil  $b$

L'idée proposée par Caeiro and Gomes (2015) est de choisir comme estimateur du seuil  $\hat{b} = X_{n-\hat{k}_o:n}$  où  $\hat{k}_o$  est l'estimateur de  $k_o$  qui minimisent l'erreur quadratique moyenne asymptotique (AMSE) de  $\hat{\xi}_{k,n}^{Hill}$  qui est égale à la somme du biais<sup>17</sup> au carré et la variance asymptotique de la distribution de  $\hat{\xi}_{k,n}^{Hill}$ .

L'estimateur de  $k_o$  est donné par :

$$\hat{k}_o = \left\lceil \left( \frac{(1-\hat{\rho})^2 n^{-2\hat{\rho}}}{-2\hat{\rho}\hat{\lambda}^2} \right)^{\frac{1}{1-2\hat{\rho}}} \right\rceil$$

La démonstration et l'estimation des différents paramètres de la formule se trouvent dans l'article Caeiro and Gomes (2015).

Toujours dans cet article<sup>18</sup>, une méthode utilisant des techniques de Bootstrapping pour minimiser le critère AMSE est introduite.

En appliquant les deux méthodes à l'aide du package R « tea » sur nos données, nous obtenons les résultats suivants :

	DAG	DAMSE	GOMES
Seuil reassurance (M Euro)	2,1	2,1	2,1
Seuil (M Euro)	1,38	2,00	1,44
Nombre de graves	1024	678	964

Figure 7-7: Résultats méthodes DAMSE et GOMES

<sup>17</sup> Soit  $\hat{k}$  un estimateur de  $k$ , le biais de  $\hat{k}$  est :  $E(\hat{k} - k)$

<sup>18</sup> Caeiro, F. and Gomes, I. (2016). Threshold selection in extreme value analysis. *Extreme Value Modeling and Risk Analysis: Methods and Applications*, 69–86.

Les deux fonctions donnent généralement des seuils inférieurs au seuil de réassurance de 2.1 M € et restent proches de celui choisi par l'équipe de modélisation.

### 7.1.2.3 QQ estimateur

Comme l'estimateur de Hill, l'estimateur QQ introduit par Kratz & Resnick dans leur article<sup>19</sup> permet d'estimer le paramètre de queue d'une distribution de valeurs extrêmes.

Cet estimateur a été construit de la manière suivante :

Soient  $U_{1:n} \leq U_{2:n} \leq \dots \leq U_{n:n}$  les statistiques d'ordre d'un échantillon de  $n$  observations i.i.d uniformes sur  $[0,1]$ .

Donc 
$$E(U_{i:n}) = \frac{i}{n+1}$$

Et puisque les  $U_{i:n}$  doivent être proche de la moyenne  $\frac{i}{n+1}$  la courbe de  $\left\{ \left( U_{i:n}, \frac{i}{n+1} \right), 1 \leq i \leq n \right\}$  doit former une droite.

Maintenant, soient les  $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$  les statistiques d'ordre de notre échantillon i.i.d. Nous supposons que notre échantillon suit une loi de fonction de répartition  $G$ .

Si notre hypothèse est vraie alors la courbe  $\left\{ \left( G(X_{i:n}), \frac{i}{n+1} \right), 1 \leq i \leq n \right\}$  doit être approximativement linéaire. Et de même pour la courbe  $\left\{ \left( X_{i:n}, G^{-1}\left(\frac{i}{n+1}\right) \right), 1 \leq i \leq n \right\}$ .

Avec :

$G^{-1}\left(\frac{i}{n+1}\right)$  le quantile théorique et  $X_{i:n}$  le quantile empirique, d'où le nom QQ-plot.

Enfin, nous supposons que nos données viennent d'une distribution  $G$  avec les paramètres de location et d'échelle  $\mu$  et  $\sigma$ . Tel que :

$$G_{\mu,\sigma}(x) = G_{0,1}\left(\frac{x-\mu}{\sigma}\right)$$

En appliquant cette technique sur des distributions à queue lourde comme la famille de lois Pareto, nous obtenons le QQ-estimateur de la façon suivante :

Nous faisons l'hypothèse que  $Z \leq Z_{2:n} \leq \dots \leq Z_{n:n}$  sont les statistiques d'ordre d'un échantillon aléatoire d'une famille de lois Pareto avec un paramètre  $\alpha > 0$  : tel que  $F_\alpha(x) = 1 - x^{-\alpha}$ ,  $x \geq 1$ .

Donc :

$$\forall y > 0, G_{0,\alpha}(y) = P[\ln(Z_1) > y] = \exp(-\alpha y)$$

Et la courbe  $\left\{ \left( \ln(Z_{i:n}), -\ln\left(1 - \frac{i}{n+1}\right) \right), 1 \leq i \leq n \right\}$  doit être approximativement une droite qui passe par l'origine et de pente  $\alpha^{-1}$ .

Enfin, l'estimateur proposé pour la pente est :

<sup>19</sup> Kratz, M. and Resnick, S.I. (1996). The QQ-estimator and heavy tails. Stochastic Models, 12(4), 699–724.

$$\hat{\alpha}^{-1} = \frac{\sum_{i=1}^n -\ln\left(\frac{i}{n+1}\right) \{n * \ln(Z_{n-i+1:n}) - \sum_{j=1}^n \ln(Z_{n-j+1:n})\}}{n \sum_{i=1}^n \left(-\ln\left(\frac{i}{n+1}\right)\right)^2 - \left(\sum_{i=1}^n -\ln\left(\frac{i}{n+1}\right)\right)^2}$$

En appliquant cette méthode sur nos données, nous obtenons le graphique suivant :

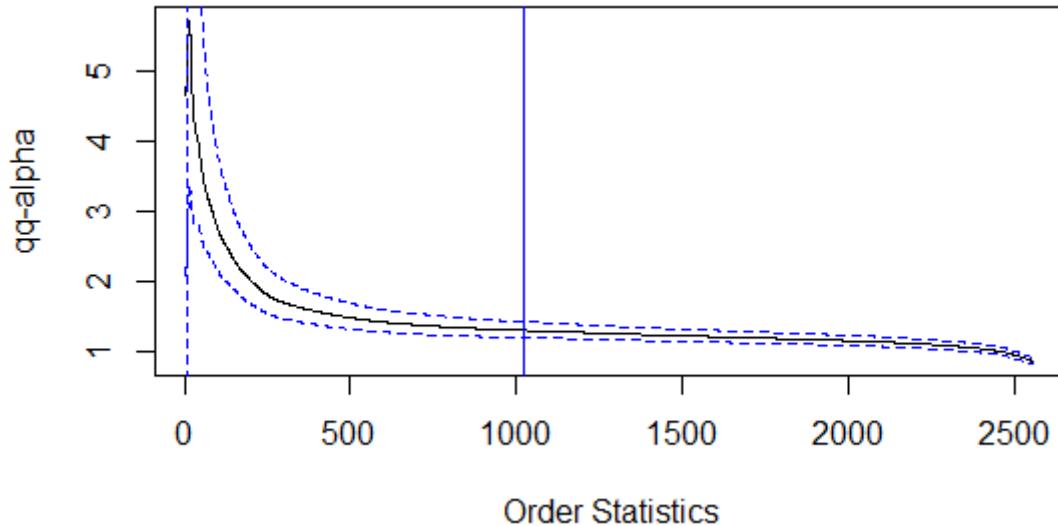


Figure 7-8: Graphique estimateur QQ

Nous remarquons que la courbe est moins volatile par rapport à la courbe obtenue par l'estimateur de Hill.

Le seuil actuel est dans une région où la courbe est stable et il est donc pertinent.

### 7.1.3 Sensibilité du seuil

Dans la partie suivante, nous allons réaliser une sensibilité par rapport au seuil retenu en choisissant le seuil de 1.1 M €.

Cette étude de sensibilité suit le processus suivant :

1. Calibrage de la charge des sinistres graves as if avec un modèle fréquence coût moyen
2. Calcul de la charge attritionnelle à partir de la charge totale moins la charge des graves
3. Modélisation de la charge attritionnelle à l'aide des S/P attritionnels en choisissant un modèle exponentiel pour la LoB MTPL.
4. Simulation du résultat technique et calcul du SCR

Les résultats de cette sensibilité sont les suivants :

	Primes acquises brutes	Charge courante ultime brute	Solde réassurance (hors pool)	Solde pool rétrocession	Frais généraux techniques	Comm de réassurance (hors pool)	Participation aux bénéfices	Résultat tech courant hors escompte	Escompte Taux sans risque	Résultat technique éco courant
mean	656,6	-587,7	27,7	0,0	-174,6	0,0	0,0	-78,0	-6,4	-84,5
standard_deviation	0,0	116,6	52,8	0,0	0,0	0,0	0,0	84,5		85,4
return_period_1_10	656,6	-736,5	82,6	0,0	-174,6	0,0	0,0	-188,8	-7,6	-196,4
return_period_1_20	656,6	-800,5	121,8	0,0	-174,6	0,0	0,0	-229,1	-8,0	-237,1
return_period_1_50	656,6	-888,9	179,3	0,0	-174,6	0,0	0,0	-279,8	-8,4	-288,2
return_period_1_100	656,6	-954,6	226,5	0,0	-174,6	0,0	0,0	-315,7	-8,8	-324,5
return_period_1_200	656,6	-1 021,6	278,4	0,0	-174,6	0,0	0,0	-352,1	-9,0	-361,1
Mean / Return period 99,5%		57,5%	10,0%	0,0%				22,2%		23,4%

Figure 7-9: Résultat technique économique avec sensibilité seuil



Nous avons une charge moyenne de -587,7 M € et un résultat technique éco de -84 M €, le SCR risque de prime est de -361,1 M €

	Primes acquises brutes	Charge courante ultime brute	Solde réassurance (hors pool)	Solde pool rétrocession	Frais généraux techniques	Comm de réassurance (hors pool)	Participation aux bénéfices	Résultat tech courant hors escompte	Escompte Taux sans risque	Résultat technique éco courant
mean	656,6	-583,3	31,0	0,0	-174,6	0,0	0,0	-70,3	-6,3	-76,7
standard_deviation	0,0	120,5	56,3	0,0	0,0	0,0	0,0	85,2		86,1
return_period_1_10	656,6	-738,1	91,3	0,0	-174,6	0,0	0,0	-181,9	-7,5	-189,4
return_period_1_20	656,6	-805,7	133,9	0,0	-174,6	0,0	0,0	-223,1	-7,9	-231,1
return_period_1_50	656,6	-894,6	194,8	0,0	-174,6	0,0	0,0	-273,7	-8,5	-282,2
return_period_1_100	656,6	-962,8	242,3	0,0	-174,6	0,0	0,0	-309,9	-8,5	-318,4
return_period_1_200	656,6	-1 033,0	296,3	0,0	-174,6	0,0	0,0	-344,9	-8,7	-353,6
Mean / Return period 99,5%		56,5%	10,5%	0,0%				20,4%		21,7%

Figure 7-10: Résultat technique économique

Le tableau de variation est le suivant :

	Primes acquises brutes	Charge courante ultime brute	Solde réassurance (hors pool)	Solde pool rétrocession	Frais généraux techniques	Comm de réassurance (hors pool)	Participation aux bénéfices	Résultat tech courant hors escompte	Escompte Taux sans risque	Résultat technique éco courant
mean	0,0	-4,4	-3,3	0,0	0,0	0,0	0,0	-7,7	-0,1	-7,8
standard_deviation	0,0	-3,9	-3,5	0,0	0,0	0,0	0,0	-0,7		-0,7
return_period_1_10	0,0	1,7	-8,7	0,0	0,0	0,0	0,0	-6,9	0,0	-6,9
return_period_1_20	0,0	5,1	-12,1	0,0	0,0	0,0	0,0	-6,0	0,0	-6,0
return_period_1_50	0,0	5,7	-15,5	0,0	0,0	0,0	0,0	-6,1	0,1	-6,0
return_period_1_100	0,0	8,1	-15,8	0,0	0,0	0,0	0,0	-5,8	-0,3	-6,2
return_period_1_200	0,0	11,5	-17,8	0,0	0,0	0,0	0,0	-7,2	-0,3	-7,5
Mean / Return period 99,5%	0,0%	1,1%	0,5%	0,0%	0	0	0,0%	-1,8%		-1,7%

Figure 7-11: Variation du résultat technique économique

Nous remarquons une augmentation de la charge moyenne de 4,4 M € lorsque nous avons baissé le seuil de séparation cela est expliqué par l'augmentation de la fréquence moyenne des graves.

Nous constatons aussi une baisse du solde réassurance (3,3 M €) qui est expliquée par la baisse du coût moyen des sinistres graves : les cessions des traités excédents de sinistre seront plus faibles.

Le résultat technique économique moyen baisse donc de 7,8 M €.

Enfin, la sensibilité du seuil a un impact négatif sur le quantile à 200 ans du résultat technique avec une baisse 7,5 M €, soit 2% du SCR de prime de la LoB RC Automobile. L'impact sur le SCR Modèle interne sera moindre grâce aux effets de diversification avec le risque de réserve et les autres LoBs S2.

## 7.2 Analyse de l'indépendance

Après l'étude sur le choix du seuil de séparation attritionnels et graves, nous avons réalisé des tests d'indépendance concernant les deux hypothèses du modèle :

- Indépendance entre sévérité et fréquence des graves
- Indépendance entre sinistres attritionnels et sinistres graves

### 7.2.1 Indépendance Fréquence / Sévérité

Les tests d'indépendance entre fréquence et sévérité des sinistres graves reposent sur la comparaison des distributions des montants de sinistres  $X$  conditionnées par le nombre de sinistres graves ( $N$ ) c'est à la dire les distributions de  $X/N$ .

Les données utilisées sont les couples  $(X_i, N_i)$  avec :

- $X_i$  le coût d'un sinistre grave individuel
- $N_i$  le nombre de sinistres ultime ajusté par l'exposition de l'année de survenance du sinistre  $X_i$  (arrondi à l'entier)

Les tests utilisés sont le test de Kruskal-Wallis et le test de Jonckheere-Terpstra qui sont détaillés en annexe 8. Ils permettent de comparer des distributions empiriques quand leur nombre est supérieur à 2. Ce sont des tests d'hypothèse avec comme hypothèse nulle : l'indépendance.

Si la p-value obtenue est inférieure à 5% cette hypothèse est rejetée.

Les résultats de ces deux tests sur les données de calibrage de la clôture 2020 sont les suivants :

Lob	Seuil(M Euro)	Jonckheere-Terpstra	Kruskal-Wallis	Hypothèse d'indépendance
MTPL	1,45	11%	5%	Non rejetée

Figure 7-12: Hypothèse indépendance fréquence sévérité

L'hypothèse d'indépendance entre fréquence et sévérité des sinistres graves n'est pas rejetée.

### 7.2.2 Indépendance sinistres attritionnels / graves

La validation de l'hypothèse d'indépendance entre attritionnels et graves s'est faite en deux étapes :

- Validation de l'indépendance attritionnels/Sévérité des graves
- Validation de l'indépendance attritionnels/Fréquence des graves

En premier temps on va tester les variables sévérité des graves et S/P attritionnels, on obtient le résultat suivant :

Lob	Seuil(M Euro)	Jonckheere-Terpstra	Kruskal-Wallis	Hypothèse d'indépendance
MTPL	1,45	12%	24%	Non rejetée

Figure 7-13: Hypothèse indépendance S/P attritionnels et sévérité

L'hypothèse d'indépendance entre attritionnels et sévérité des graves n'est pas rejetée.

Ensuite nous appliquons les tests d'indépendance sur les deux séries de données :

- Les nombres de sinistres graves annuels ultimes ajustés par rapport à l'exposition
- Les S/P attritionnels

Lob	Seuil(M Euro)	Jonckheere-Terpstra	Kruskal-Wallis	Hypothèse d'indépendance
MTPL	1,45	11%	50%	Non rejetée

Figure 7-14: Hypothèse indépendance S/P attritionnels et fréquence

L'hypothèse d'indépendance entre attritionnels et fréquence des sinistres graves n'est pas rejetée.

Maintenant que ces deux hypothèses d'indépendance sont vérifiées, nous pouvons commencer la validation du calibrage du modèle fréquence/sévérité des graves.

## 7.3 Calibrage de la charge des graves

La modélisation des risques graves se base sur un modèle fréquence/sévérité impliquant donc la calibration de deux lois. Celles-ci sont calibrées à partir des sinistres graves historiques observés au-delà du seuil des graves qui a été revu.

Les travaux de validations concernent le calcul miroir et le calibrage indépendant des lois de fréquence et sévérité, des analyses d'évolution N/N-1 et enfin des analyses de sensibilités.

### 7.3.1 Données

Les données correspondent aux sinistres graves des CR au Q4 2020, qui dépassent le seuil de modélisation de l'année 2020 qui est le seuil de séparation entre attritionnels et graves calculé de la façon suivante :

$$\text{Seuil 2020} = \text{Seuil 2019} * (1 + 5\%)$$

Notre base de données se compose de 827 sinistres graves.

Pour des raisons d'application de la réassurance, la modélisation des sinistres graves MTPL nécessite deux approches distinctes : avec ou sans prise en compte des revalorisations FGAO.

Etant donné le très faible volume de données sur les survenances inférieures à 2000, il a été décidé de ne pas en tenir compte dans le calibrage du risque de Primes.

### 7.3.2 Calibrage de la loi de fréquence

Nous avons effectué un calibrage indépendant de la loi de fréquence des graves sur les nombres suivants :

Surv	As at 31/12/2020	Nb sinistres ultimes	Indice d'exposition	Ultimes ajustés
1996	45,0	45,0		
1997	56,0	56,0		
1998	70,0	70,0		
1999	54,0	54,3		
2000	68,0	69,3	0,603	55,7
2001	52,0	53,6	0,793	42,5
2002	61,0	63,0	0,788	49,7
2003	41,0	42,9	0,783	33,6
2004	31,0	32,9	0,784	25,8
2005	44,0	47,2	0,799	37,7
2006	36,0	39,2	0,833	32,6
2007	25,0	26,9	0,856	23,1
2008	40,0	43,0	0,869	37,3
2009	38,0	40,9	0,865	35,4
2010	41,0	44,0	0,849	37,3
2011	41,0	43,7	0,847	37,0
2012	48,0	51,8	0,848	43,9
2013	32,0	34,3	0,873	30,0
2014	51,0	54,7	0,903	49,4
2015	32,0	34,0	0,934	31,8
2016	44,0	46,9	0,961	45,0
2017	27,0	29,5	0,982	29,0
2018	33,0	36,6	0,997	36,4
2019	28,0	32,6	1,011	33,0
2020	10,0	30,3	1,012	30,6
Moyenne	41,92	44,91		37,00
Ecart-type	13,83	11,34		8,23

Figure 7-15: Nombre de sinistres ultimes ajustés

Les coefficients utilisés pour projeter les nombres à l'ultime sont ceux issus d'un Chain Ladder classique de notre triangle de nombre de sinistres graves.

L'indice d'exposition considéré est la taille du portefeuille en terme de volume de primes.

Des tests d'adéquation et de comparaison des fonctions de répartition empiriques et théoriques ont été réalisés, les résultats sont les suivants :

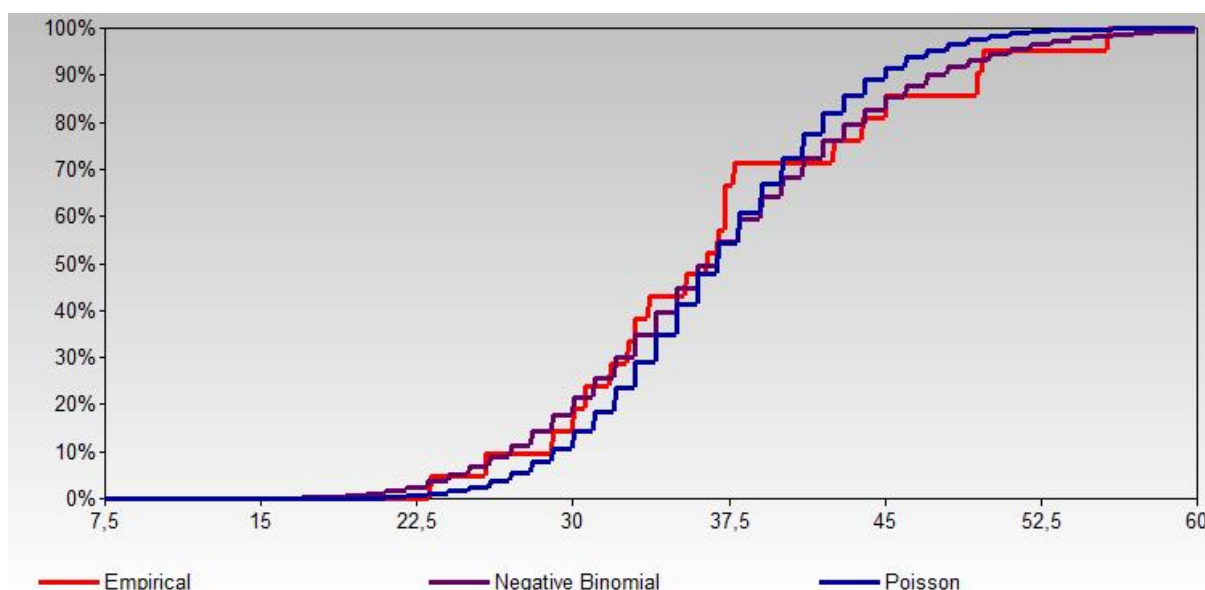


Figure 7-16: Comparaison de fonctions de répartition

Nous constatons que les quantiles de la loi binomiale négative sont proches des quantiles empiriques.

	Least Squares	Kolmogorov	Kuiper	Anderson
Negative Binomial	0,06	0,17	0,20	0,34
Poisson	0,13	0,17	0,24	0,47

Figure 7-17: Tests d'adéquation

La loi binomiale négative donne un meilleur ajustement d'après les tests d'adéquation.

Enfin, compte tenu du fait que, pour l'échantillon étudié, la variance est supérieure à la moyenne, nous pouvons maintenir le choix fait par l'équipe de modélisation de la loi binomiale négative.

L'estimation des paramètres se fait par la méthode des moments et les résultats obtenus sont les suivants :

Segment	Choix de la loi	Paramètres	Calibrage (DAG)	Calcul miroir	Ecart %
Fréquence	Binomiale négative	E	37	37	0%
		V/E	1,83	1,83	0%

Figure 4.10 : Comparaison calibrage équipe modélisation et validation

Le calcul miroir est cohérent avec le calcul obtenu par l'équipe de modélisation.

### 7.3.3 Calibrage de la loi du coût des graves

Le choix de la loi du coût des sinistres graves MTPL a été réalisé avec et sans revalorisation FGAO par l'équipe modélisation. Ce choix est justifié par l'obligation de prendre en compte les spécificités du programme de réassurance en lien avec les revalorisations FGAO.

Des tests d'adéquation et de comparaison des fonctions de répartition empiriques et théoriques ont été réalisés, les résultats sont les suivants :

- Avec FGAO

Les résultats des tests d'adéquations réalisés sur les différentes lois testées sont les suivants :

	Akaike	Least Squares	Kolmogorov	Kuiper	Anderson
Pareto Simple	-16,94	0,14	0,03	0,05	0,17
Pareto	-17,94	0,14	0,03	0,05	0,17
Lognormal	-17,94	0,27	0,04	0,06	0,18
Weibull	-17,94	0,59	0,05	0,08	0,19
Gamma	-18,02	8,81	0,17	0,20	0,36
Normal	-18,22	23,44	0,27	0,32	Unlimited

Figure 4.11 : Tests d'adéquation

Nous remarquons que la loi Pareto donne un meilleur ajustement, mais elle ne présente pas un avantage significatif du point de vue des critères d'adéquation.

Ensuite, nous avons calculé la moyenne, l'écart-type et le quantile à 99,5% pour chaque distribution et nous obtenons les résultats suivants :

	Gamma	Lognormal	Normal	Weibull	Pareto	Pareto simple	Empirique
Mean	5 466 018,57	5 994 115,89	7 142 206,58	5 587 410,72	6 302 856,48	6 335 179,79	5 465 654,47
SD	5 234 219,66	12 844 700,23	#NOMBRE!	10 403 819,54	14 532 855,19	14 619 139,49	7 257 528,11
Quantile 99,5%	31 536 518,61	90 857 604,13	1 448 000,00	69 036 615,89	107 896 403,50	108 711 997,33	39 070 548,51

Figure 4.12 : moments et quantiles des distributions testées

Les lois Pareto simple et généralisée ont des moments plus élevés que les moments empiriques et ceux dans une mesure plus importante que la surestimation des moments obtenue via la loi lognormale qui a des résultats moins bons mais confortables d'adéquation.

C'est pour cette raison que l'équipe de modélisation a choisi la loi lognormale pour la modélisation de la sévérité.

- Hors FGAO

Les résultats des tests d'adéquations réalisés sur les différentes lois testées sont les suivants :

	Akaike	Least Squares	Kolmogorov	Kuiper	Anderson
Pareto Simple	-16,84	0,12	0,03	0,05	0,17
Pareto	-17,83	0,17	0,03	0,06	0,18
Lognormal	-17,83	0,22	0,03	0,06	0,18
Weibull	-17,83	0,34	0,04	0,07	0,18
Gamma	-17,87	5,24	0,13	0,17	0,27
Normal	-18,00	16,33	0,23	0,29	Unlimited

Figure 4.13 : Tests d'adéquation

	Gamma	Lognormal	Normal	Weibull	Pareto	Pareto simple	Empirique
Mean	4 712 114,60	5 027 992,28	7 401 944,05	4 700 375,45	5 334 422,41	5 726 321,31	4 711 309,05
SD	4 115 860,57	8 952 276,43	#NOMBRE!	6 960 015,80	11 054 366,28	12 992 122,83	5 239 813,54
Quantile 99,5%	24 939 556,36	57 469 173,57	1 448 000,00	44 263 384,29	74 627 586,58	93 245 960,09	30 463 015,34

Figure 4.14 : moments et quantiles des distributions testées

Pour les mêmes raisons que celles décrites pour le périmètre y compris revalorisations FGAO. La loi lognormale a été choisie pour la modélisation de la sévérité hors revalorisations FGAO.

L'estimation des paramètres de la loi retenue se fait par la méthode de maximum de vraisemblance et les résultats du calibrage indépendant sont les suivants :

Segment	Choix de la loi	Paramètres	Calibrage (DAG)	Calcul miroir	Ecart %
Coût hors FGAO	Lognormale	mu	1,96	1,40	40,62%
		sigma	3,52	3,59	-1,93%
Coût avec FGAO	Lognormale	mu	10,45	10,45	0,04%
		sigma	2,09	2,09	-0,04%

Figure 4.15 : estimation des paramètres de la loi de sévérité

Nous constatons qu'il y'a des écarts sur les paramètres de la loi lognormale pour la sévérité hors revalorisation FGAO qui sont dues à la différence d'outils d'optimisation entre R et le logiciel utilisé par l'équipe de modélisation.

Nous avons choisi de réaliser une sensibilité en nous reposant sur une loi de Pareto simple (qui est la loi avec le meilleur résultat aux tests d'adéquation). Les résultats de cette sensibilité sont les suivants :

Sinistralité Grave				CR	
				Avec FGAO	Sans FGAO
Calibrage des sinistres Graves	Nombre	Distribution	DAG	Binomiale négative	
			Validation	Binomiale négative	
		Esp	DAG	37,00	
			Validation	37,00	
			écart	0%	
			DAG	1,83	
		Var/Esp	Validation	1,83	
			écart	0%	
	Sévérité	Distribution	DAG	LogNormale	LogNormale
			Validation	Pareto Simple	Pareto Simple
		Min	DAG	1448000	
			Validation	1448000	
		Max	DAG	250000000	
			Validation	250000000	
		Paramètres	DAG	$\mu = 1,96 ; \sigma = 3,52$	$\mu = 10,45 ; \sigma = 2,091$
			Validation	1,111	1,183
		Charge espérée (M€)	DAG	222,0	185,8
			Validation	234,5	212,0
			écart	-5%	-12%
	Quantile à 99,5% des charges (M€)	DAG	558,6	436,3	
			Validation	607,5	552,8
			écart	-8,0%	-21,1%

Figure 4.15 : Sensibilité de la charge par rapport à loi de sévérité

Le quantile à 99,5% obtenu via cette sensibilité est significativement plus élevé mais ce choix de loi de Pareto reste trop prudent par rapport à la modélisation par une loi lognormale, d'autant plus que le quantile 99,5% théorique n'a pas déjà été observé par Groupama.

Nous avons réalisé un recalcul miroir du quantile 99,5% associé à ce risque en reprenant la calibration retenue par l'équipe de modélisation. Les résultats sont les suivants :

Sinistralité Grave				CR	
				Avec FGAO	Sans FGAO
Calibrage des sinistres Graves	Nombre	Distribution	DAG	Binomiale négative	
			Validation	Binomiale négative	
		Esp	DAG	37,00	
			Validation	37,00	
			écart	0%	
		Var/Esp	DAG	1,83	
			Validation	1,83	
			écart	0%	
	Sévérité	Distribution	DAG	LogNormale	LogNormale
			Validation	LogNormale	LogNormale
		Min	DAG	1448000	
			Validation	1448000	
		Max	DAG	250000000	
			Validation	250000000	
		Paramètres	DAG	$\mu = 1,96 ; \sigma = 3,52$	$\mu = 10,45 ; \sigma = 2,091$
			Validation	$\mu = 1,39 ; \sigma = 3,59$	$\mu = 10,44 ; \sigma = 2,091$
	Charge espérée (M€)		DAG	221,5	186,1
			Validation	221,4	186,0
			écart	0,04%	0,05%
	Quantile à 99,5% des charges (M€)		DAG	557,8	438,1
			Validation	556,1	440,3
			écart	0,3%	-0,5%

Figure 7-18: Calibrage de la charge des graves

Nous observons des écarts non significatifs liés au caractère simulateur du calcul et à la différence d'outils utilisés.

#### 7.3.4 Comparaison de la calibration N/N-1

Le tableau suivant présente une comparaison de la calibration et de la charge résultante pour l'exercice 2019 et 2020 :

Sinistralité Grave				CR	
				Avec FGAO	
Calibrage des sinistres Graves	Nombre	Distribution	N	Binomiale négative	
			N-1	Binomiale négative	
		Esp	N	36,998	
			N-1	37,754	
		Var/Esp	N	1,833	
			N-1	2,436	
		Sévérité	Distribution	N	LogNormale
	N-1			Pareto Simple	
	Min		N	1 448 000	
			N-1	1 379 000	
	Max		N	250 000 000	
			N-1	50 065 000	
	Paramètres		N	$\mu = 1,96 ; \sigma = 3.52$	
			N-1	0,901	
	Charge espérée (M€)		N	222,0	
			N-1	210,8	
			écart	5,31%	
	Quantile à 99,5% des charges (M€)		N	559,1	
			N-1	426,2	
			écart	31.18%	

Figure 7-19: Comparaison calibrage N/N-1

Nous remarquons une augmentation significative de la charge à 99,5% (+31%) qui résulte principalement par la suite de l'évolution du seuil maximum de modélisation qui est passé de 50 M€ à 250M€. Dans une autre mesure, cette hausse est aussi liée à la hausse du seuil minimum des graves et est légèrement compensée par la baisse légère de la fréquence.



## Conclusion

Notre étude consiste à analyser et valider les méthodes actuelles de modélisation de sinistres graves de la ligne d'activité responsabilité civile automobile. Au sein du modèle interne ces sinistres sont traités et identifiés de manière différente entre les deux modules : risque de primes et risque de réserves.

Pour la partie risque de réserves, nous avons étudié la méthode actuelle de Merz-Wüthrich par Bootstrap qui traite les sinistres graves de façon agrégé. L'étude consiste à :

- Valider et analyser les hypothèses du modèle
- Implémenter le modèle et réaliser des tests de sensibilité

Les hypothèses ont été vérifiées au global et testées par plusieurs méthodes et les tests de sensibilité ne remettent pas en cause le niveau du SCR calculé.

Ensuite, nous avons proposé une méthode alternative qui traite les sinistres graves de façon individuelle à l'aide d'un modèle multi-états. Cette méthode nécessite l'estimation des probabilités de transition d'un processus stochastique semi-Markovien et l'estimation des montants de paiements lorsque le sinistre transite vers un état de paiement. Nous avons vu que l'implémentation de ce modèle nécessite des données plus précises et plus détaillées au niveau de chaque sous-dossier victime. Enfin à l'aide des hypothèses permettant de refléter la réalité des données observées, nous avons pu calculer la réserve et le risque à 1 an. Le montant de capital obtenu reste largement inférieur à celui calculé par la méthode actuelle ce qui nous donne un confort de prudence sur cette estimation.

Pour la partie risque de primes, la méthode actuelle consiste à identifier un seuil de séparation pour les sinistres graves des caisses régionales et par la suite modéliser la charge future de chaque type de sinistres. L'étude donc consiste à :

- Analyser et valider le seuil de séparation entre sinistres attritionnels et sinistres graves
- Implémenter le calcul miroir et réaliser des tests de sensibilité

Nous avons testé les méthodes actuelles d'identification du seuil et étendu notre analyse à plusieurs méthodes alternatives telles que l'estimateur QQ-plot et le graphique de Gerstengarbe and Werner.

Nous précisons que les méthodes issues de la théorie des valeurs extrêmes offrent des intervalles donnant la possibilité de plusieurs choix de seuil de séparation. Le choix du seuil doit prendre en considération aussi les contraintes de modélisation du risque de prime, telles que la priorité des traités de réassurance et l'indépendance entre le nombre et coût des sinistres du modèle collectif.

Les travaux de ce mémoire s'inscrivent et enrichissent le processus de validation du modèle interne au sein de Groupama. Ils ont permis de valider les résultats obtenus du modèle.

Enfin, pour donner des perspectives de développement du modèle multi-états, une première possibilité est d'appliquer le modèle individuel à états sur des portefeuilles classiques avec des typologies de sinistres moins complexes. Une autre possibilité est de l'étendre à la modélisation de la charge ultime utilisée pour le calibrage du modèle de risque de primes.

## Lexique des abréviations

DAG : Direction Actuariat Groupe

CR : Caisses Régionales

GMA : Groupama Assurances Mutuelles

CGMI : Comité Groupe du Modèle Interne

LoB : *Line of Business*

MTPL : *Motor Third Party Liability*

BE : *Best Estimate*

SCR : *Solvency Capital Requirement*

VaR : *Value at Risk*

IBNR : *Incurred but Not Reported*

RBNP : *Reported but Not Paid*

RBNS : *Reported but Not Settled*

## Annexes

### Annexe 1 : Les statistiques d'adéquation

Les mesures de statistiques utilisées dans la modélisation du risque de primes sont toutes calculées lors de la maximisation de la vraisemblance.

Soit  $x_1 \dots x_n$  un échantillon de données de taille  $n$ . Soit  $F_{empirique}(x_i)$  et  $F_{ajustée}(x_i)$  respectivement les fonctions de répartition empirique et ajustée valorisées au point  $x_i$ .

#### Statistique d'Akaike (critère AIC) :

Le critère AIC d'un modèle est calculé comme la log-vraisemblance maximisée retranchée du nombre de paramètres  $p$  à estimer. Il s'écrit :

$$AIC = \sum_{i=1}^n \ln(F_{ajustée}(x_i)) - p.$$

#### Statistique des moindres carrés (critère MC) :

Le critère MC d'un modèle est calculé comme la somme des carrés des différences entre la valeur estimée et la valeur empirique de chaque donnée. Il s'écrit :

$$MC = \sum_{i=1}^n (F_{empirique}(x_i) - F_{ajustée}(x_i))^2.$$

#### Statistique de Kolmogorov-Smirnov (critère KS) :

Le critère KS d'un modèle est calculé comme le plus grand écart en valeur absolue entre les valeurs estimées et les valeurs empiriques des données. Il s'écrit :

$$KS = \max_i |F_{empirique}(x_i) - F_{ajustée}(x_i)|$$

#### Statistique de Kuiper (critère KU) :

Le critère KU d'un modèle est calculé comme la somme des distances maximales des valeurs estimées en dessous et au-dessus des valeurs empiriques. Il s'écrit :

$$KU = \max_i [F_{empirique}(x_i) - F_{ajustée}(x_i)] + \max_i [F_{ajustée}(x_i) - F_{empirique}(x_i)]$$

#### Statistique d'Anderson-Darling (critère AD) :

Le critère AD d'un modèle est le critère KS pondéré de façon à donner plus d'importance aux points extrêmes à gauche et à droite. Il s'écrit :

$$KS = \max_i \frac{|F_{empirique}(x_i) - F_{ajustée}(x_i)|}{\sqrt{(F_{ajustée}(x_i)) \cdot (1 - F_{ajustée}(x_i))}}$$

## Annexe 2 : lien entre la linéarité et non corrélation des facteurs de développement

L'hypothèse de linéarité de Chain-Ladder implique la non-corrélation entre les facteurs de développement de colonnes différentes. Ci-dessous la démonstration de cette implication.

Proposition :

Soit  $I$  le nombre d'années d'accident et  $J$  le nombre d'années de développement.

Sous l'hypothèse de linéarité :

$\exists f_1, \dots, f_{J-1}$  positifs tels que  $\forall k \in [1, J-1], \forall i \in [1, I], E(C_{i,k+1} | C_{i,1}, \dots, C_{i,k}) = C_{i,k} f_k$   
les coefficients de développement successifs  $f_k$  et  $f_{k+1}$  ne sont pas corrélés et nous avons :

$$\forall k \in [2, J-1], \forall i \in [1, I], \quad E\left(\frac{C_{i,k}}{C_{i,k-1}} \frac{C_{i,k+1}}{C_{i,k}}\right) = E\left(\frac{C_{i,k}}{C_{i,k-1}}\right) E\left(\frac{C_{i,k+1}}{C_{i,k}}\right)$$

Preuve :

$\forall i \in [1, I], \forall k \in [1, J-1], \forall j \leq k$ , nous avons :

$$(1) \quad E\left(\frac{C_{i,k+1}}{C_{i,j}}\right) = E\left(E\left(\frac{C_{i,k+1}}{C_{i,j}} \middle| C_{i,1}, \dots, C_{i,k}\right)\right) \quad (a)$$

$$= E\left(\frac{E(C_{i,k+1} | C_{i,1}, \dots, C_{i,k})}{C_{i,j}}\right) \quad (b)$$

$$= E\left(\frac{f_k C_{i,k}}{C_{i,j}}\right) \quad (c)$$

$$= f_k E\left(\frac{C_{i,k}}{C_{i,j}}\right) \quad (d)$$

L'égalité (a) vient d'une propriété de l'espérance conditionnelle :

$$E(E(X|Y)) = E(X)$$

(b) vient du fait qu'étant donnés  $C_{i,1}, \dots, C_{i,k}$ ,  $C_{i,j}$  est un scalaire (car  $j$  est inférieur à  $k$ ), (c) vient de l'hypothèse de linéarité et on obtient finalement (d) car  $f_k$  est un scalaire.

En particulier pour  $j = k$ , nous avons :

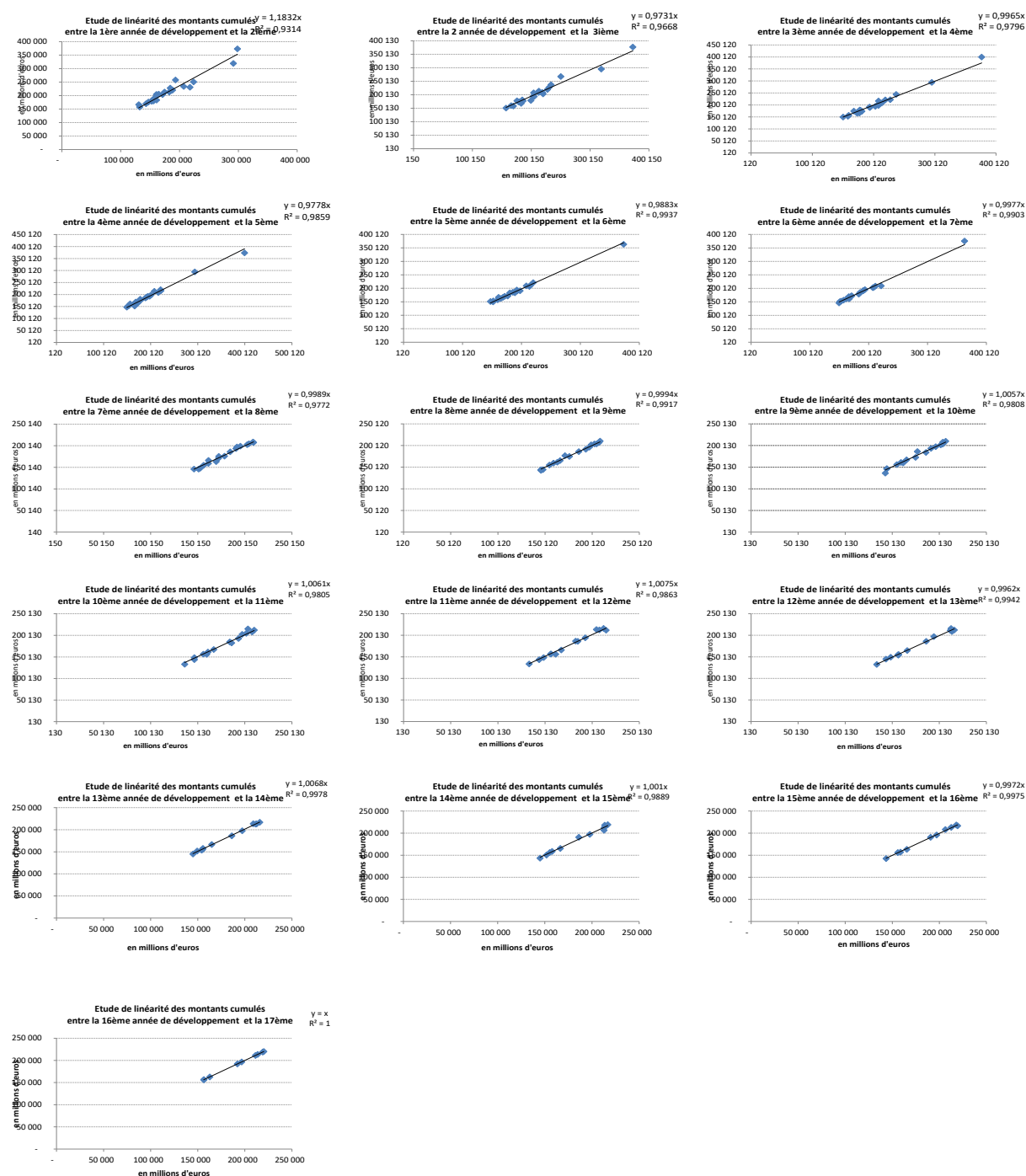
$$(2) \quad E\left(\frac{C_{i,k+1}}{C_{i,k}}\right) = f_k E\left(\frac{C_{i,k}}{C_{i,k}}\right) = f_k$$

Et pour  $j = k-1$  :

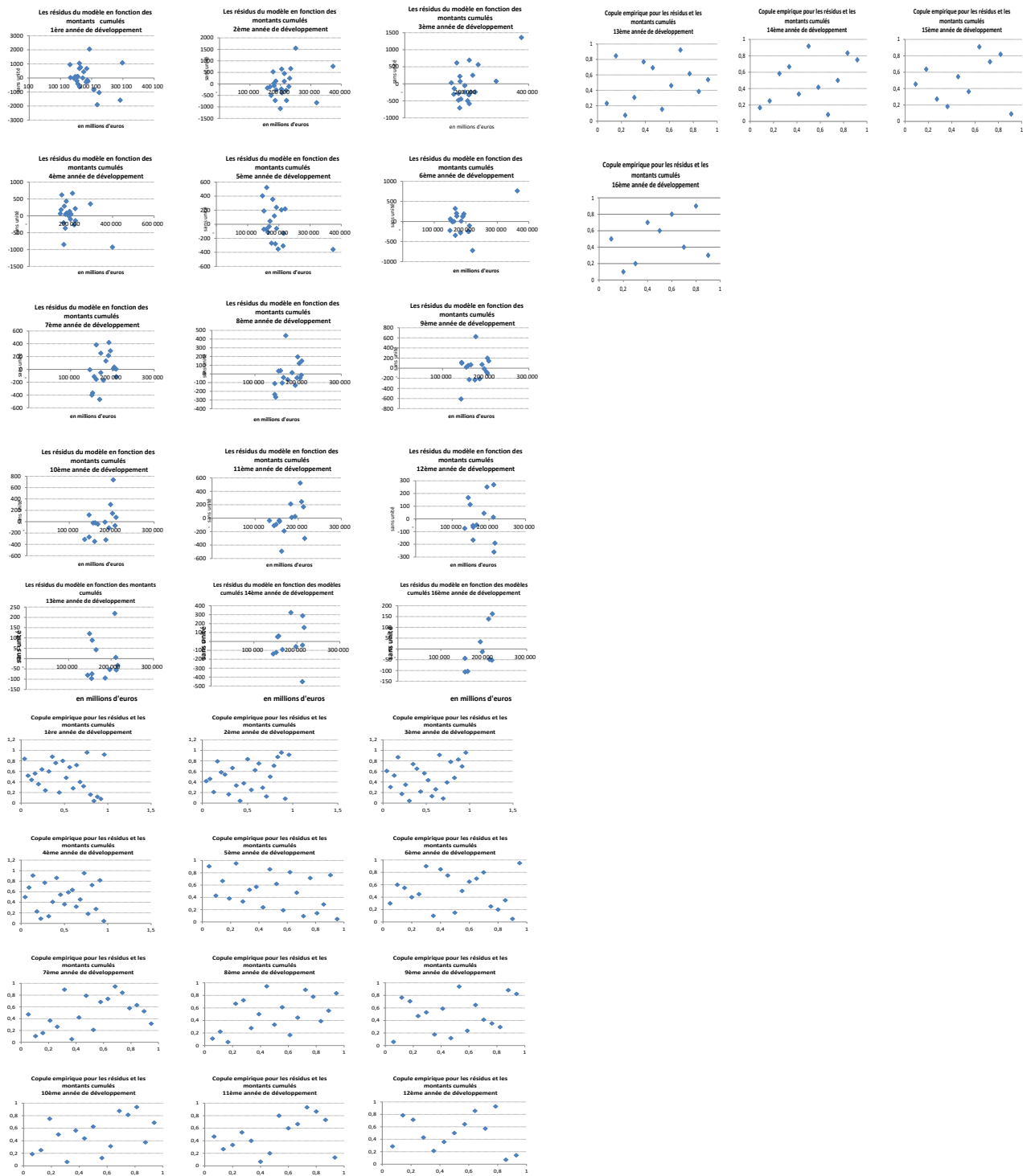
$$(3) \quad E\left(\frac{C_{i,k+1}}{C_{i,k-1}}\right) = E\left(\frac{C_{i,k+1}}{C_{i,k}} \frac{C_{i,k}}{C_{i,k-1}}\right) = f_k E\left(\frac{C_{i,k}}{C_{i,k-1}}\right)$$

En remplaçant dans (3)  $f_k$  par l'expression trouvée en (2), nous obtenons le résultat cherché.

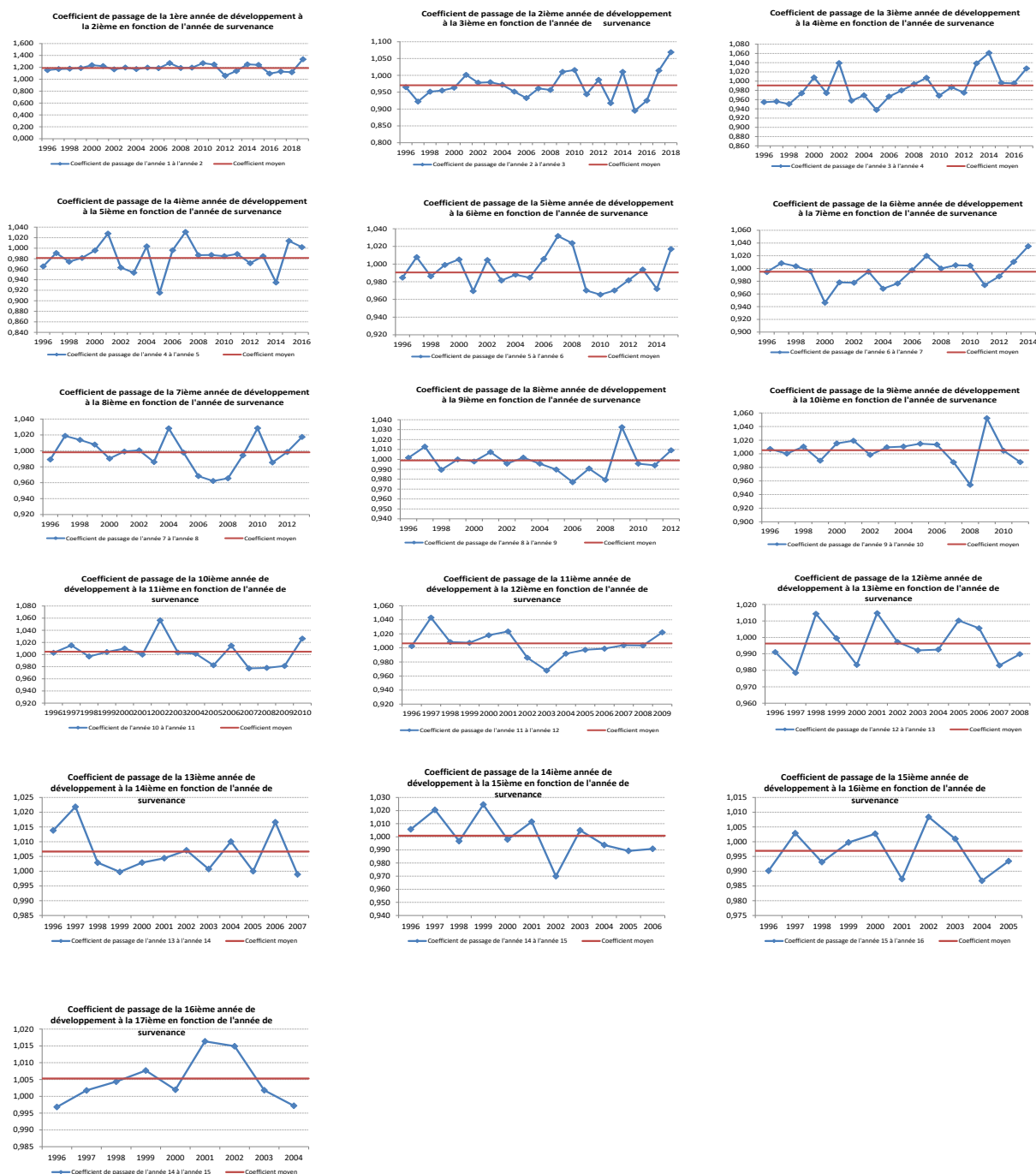
## Annexe 3 : Test graphique de l'hypothèse de linéarité



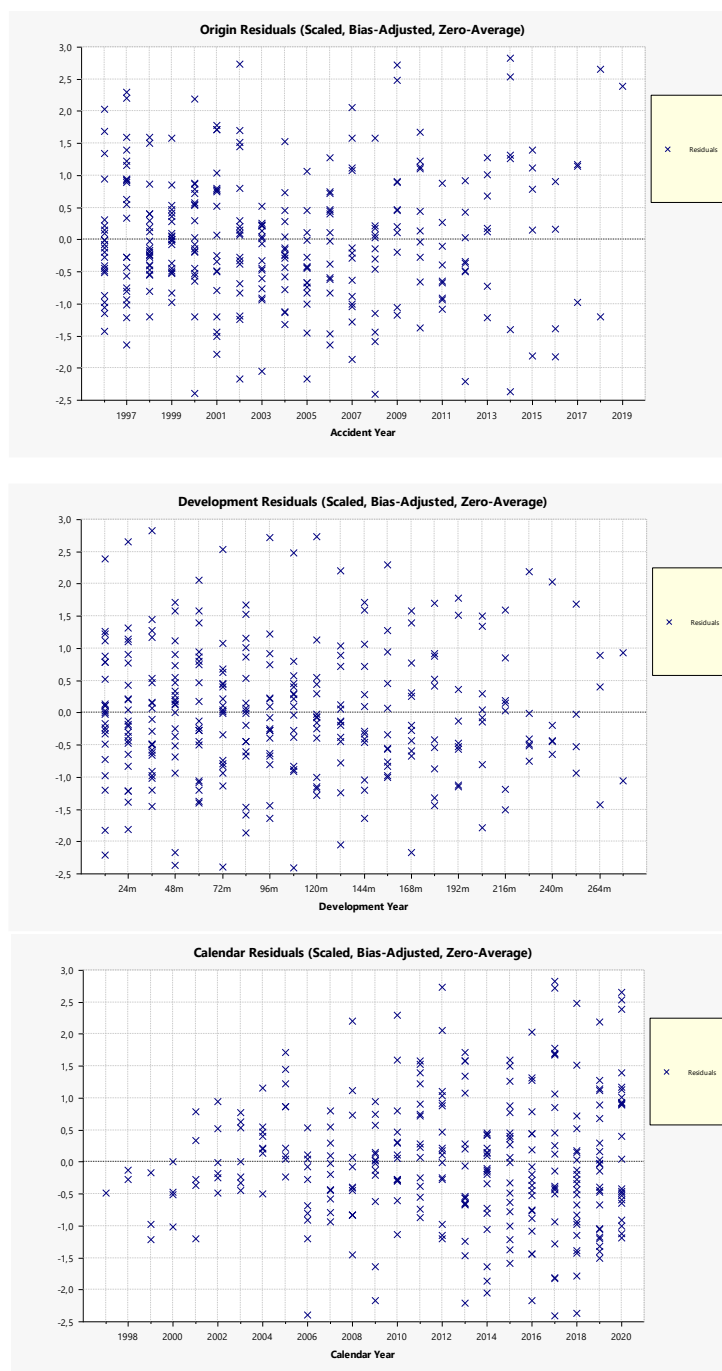
## Annexe 4 : Test graphique de l'hypothèse de variance



## Annexe 5 : Test graphique de l'hypothèse d'indépendance entre années de survénance



## Annexe 6 : Test graphique de l'indépendance des résidus





# Annexe 7 : Triangle de nombres de sinistre déclaré par année de survenance et délai de déclaration

	0	1	2		0	1	2
1966	0	0	1	1998	177	0	1
1968	0	0	1	1999	213	4	0
1969	1	0	0	2000	170	14	3
1971	1	0	1	2001	183	8	5
1974	0	0	2	2002	182	7	1
1975	1	0	2	2003	182	20	1
1976	1	0	0	2004	184	3	9
1977	2	0	2	2005	223	5	0
1978	1	0	2	2006	211	9	0
1979	1	0	4	2007	191	7	3
1980	6	0	1	2008	214	5	1
1981	6	0	5	2009	198	12	1
1982	9	1	3	2010	253	18	1
1983	16	1	1	2011	269	13	0
1984	14	0	1	2012	343	11	3
1985	15	0	2	2013	207	4	1
1986	22	5	5	2014	324	3	0
1987	16	0	7	2015	264	4	4
1988	24	4	6	2016	292	17	0
1989	16	0	6	2017	270	19	1
1990	20	5	0	2018	277	2	
1991	34	0	1	2019	88		
1992	55	2	0				
1993	54	1	1				
1994	79	0	1				
1995	89	4	0				
1996	76	2	1				
1997	130	9	8				

## Annexe 8 : Tests d'homogénéité

### Kruskal-Wallis

Le test de Kruskal-Wallis permet de comparer les distributions de  $N$  échantillons avec  $N \geq 2$

La statistique du test est :

$$KW = (n_{tot} - 1) \frac{\sum_{j=1}^N n_j (\bar{r}_j - \bar{r})^2}{\sum_{i=1}^N \sum_{j=1}^{n_i} (\bar{r}_{ij} - \bar{r}_i)^2}$$

Avec :

- $N$  : le nombre d'échantillons
- $n_j$  : le nombre d'observations au sein de l'échantillon  $j$
- $\bar{r}_{ij}$  : le rang (tous échantillons confondus) de l'observations  $j$  de l'échantillon  $i$
- $n_{tot}$  : le nombre total d'observations
- $\bar{r}_i = \frac{\sum_{j=1}^{n_i} \bar{r}_{ij}}{n_i}$  est le rang moyen des observations de l'échantillon  $i$
- $\bar{r} = \frac{1+n_{tot}}{2}$  la moyenne des  $\bar{r}_{ij}$

Sous l'hypothèse nulle,  $KW$  suit une loi du khi-deux à  $N - 1$  degrés de liberté.

### Jonckheere-Terpstra

Ce test permet aussi de comparer  $N$  échantillons avec  $N \geq 2$ . La statistique du test est :

$$JT = 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \left( U_{ij} - \frac{n_i n_j}{2} \right)$$

Tel que :

- $N$  : le nombre d'échantillons
- $n_i$  : le nombre d'observations au sein de l'échantillon  $i$
- $U_{ij} = \sum_{l=1}^{n_i} \sum_{m=1}^{n_j} 1_{R^+}(X_{j,l} - X_{i,m})$ ,  $1 \leq i, j \leq N$ , et  $X_{i,m}$  la  $m^{ème}$  observation du  $i^{ème}$  échantillon ordonné.

Sous l'hypothèse nulle, la distribution asymptotique de  $JT$  est une loi normale centrée et de variance :

$$V(JT) = \frac{1}{18} \left( n_{tot}^2 (2n_{tot} + 3) - \sum_{i=1}^N n_i^2 (2n_i + 3) \right)$$

Où  $n_{tot}$  est le nombre total d'observations tous échantillons confondus.

## Table des figures

Figure 1: Résultats du calcul miroir .....	6
Figure 2: Comparaison des stabilisations .....	6
Figure 3: Test de stabilité .....	6
Figure 4: Test de convergence .....	6
Figure 5: Comparaison des réserves entre modèle multi états et Chain Ladder classique .....	7
Figure 6: Risque à 1 an en M € .....	7
Figure 7: Comparaison des méthodes.....	7
Figure 8: Variation du résultat technique économique .....	8
Figure 9: Hypothèse d'indépendance.....	8
Figure 10: Calibrage de la charge des graves .....	8
Figure 1-1: Piliers de Solvabilité 2 .....	19
Figure 1-2: Modules de risques du modèle interne partiel.....	20
Figure 1-3: Processus de modélisation du risque de primes.....	24
Figure 2-1: Extrait du corpus méthodologique - Groupama .....	26
Figure 2-2: Extrait du corpus méthodologique - Groupama .....	31
Figure 2-3: Estimation du facteur de queue.....	32
Figure 3-1: les densités des 3 lois de valeurs extrêmes .....	35
Figure 5-1: Retraitements entre 2019 et 2020.....	43
Figure 5-2: Sigma de Mack les plus récentes des 3 dernières clôtures.....	44
Figure 5-3: Effet de stabilisation sur les sigma de Mack .....	44
Figure 5-4: Résidus du Bootstrap .....	44
Figure 5-5: Rejeu du calcul de volatilité à 1 an.....	45
Figure 5-6: Calcul miroir de volatilité à 1 an.....	45
Figure 5-7: Sigma de Mack stabilisés.....	46
Figure 5-8: Sensibilité sur les retraitements.....	46
Figure 5-9: Test de stabilité .....	47
Figure 5-10: Test de convergence .....	47
Figure 5-11: Sensibilité erreur de process.....	47
Figure 5-12: Sensibilité erreur d'estimation.....	47
Figure 5-13: Coefficient $R^2$ de la régression linéaire sans constante .....	55
Figure 5-14: Test global de Spearman.....	56
Figure 5-15: Tests de corrélation pour la linéarité.....	56
Figure 5-16: Tests analytiques pour la variance .....	57
Figure 5-17: Test d'effets d'année calendaire .....	57
Figure 5-18: Test de Kolmogorov-Smirnov.....	58
Figure 5-19: Indépendance des résidus par année de survenance.....	58
Figure 5-20: Indépendance des résidus par année de développement.....	59
Figure 5-21: Indépendance des résidus par année calendaire .....	59
Figure 5-22: Comparaison Bootstrap / Merz-Wüthrich .....	59
Figure 5-23: Analyse de variation N/N-1 .....	60
Figure 5-24: Analyse d'écart provisions Best Estimate .....	61

Figure 5-25: Distributions des réserves à horizon 1 an en 2019 .....	61
Figure 6-1: dynamique de règlement d'un sinistre selon le modèle multi-états .....	63
Figure 6-2: Toutes les trajectoires possibles pour un sinistre durant une année .....	67
Figure 6-3: Base de données sinistres (M €) .....	68
Figure 6-4: Nombre de paiements sous-dossier clôturé .....	69
Figure 6-5: Délai de déclaration .....	69
Figure 6-6: Temps jusqu'à la clôture .....	69
Figure 6-7: Nombre de sous-dossiers victime avec rente .....	69
Figure 6-8: Valeur moyenne rente .....	69
Figure 6-9: Distribution des premiers paiements.....	69
Figure 6-10: Loi mélange des premiers paiements .....	74
Figure 6-11: Uniformité de la fonction de répartition des premiers paiements.....	74
Figure 6-12: Premier paiement en fonction du premier link ratio.....	75
Figure 6-13: nombre moyen d'IBNR déclaré par année de survenance et délai de déclaration .....	77
Figure 6-14: Charge ultime et réserve du modèle multi-états (en M €.) .....	78
Figure 6-15: Comparaison des réserves du modèle multi-états + binaire avec Chain Ladder .....	79
Figure 6-16: Nombre moyen d'IBNR déclaré pour l'année calendaire 2020 .....	80
Figure 6-17: risque à 1 an en M €. .....	80
Figure 7-1: Nombre de sinistres LoB MTPL .....	82
Figure 7-2: Graphique Mean Excess .....	83
Figure 7-3: Nouveau Graphique Mean Excess .....	83
Figure 7-4: Graphique de l'estimateur de Hill .....	84
Figure 7-5: Graphique Mean Excess .....	84
Figure 7-6: Graphique Gerstengarbe and Werner .....	85
Figure 7-7: Résultats méthodes DAMSE et GOMES .....	86
Figure 7-8: Graphique estimateur QQ.....	88
Figure 7-9: Résultat technique économique avec sensibilité seuil .....	88
Figure 7-10: Résultat technique économique .....	89
Figure 7-11: Variation du résultat technique économique .....	89
Figure 7-12: Hypothèse indépendance fréquence sévérité .....	90
Figure 7-13: Hypothèse indépendance S/P attritionnels et sévérité .....	90
Figure 7-14: Hypothèse indépendance S/P attritionnels et fréquence .....	90
Figure 7-15: Nombre de sinistres ultimes ajustés .....	91
Figure 7-16: Comparaison de fonctions de répartition .....	92
Figure 7-17: Tests d'adéquation.....	92
Figure 7-18: Calibrage de la charge des graves .....	95
Figure 7-19: Comparaison calibrage N/N-1.....	95

## Bibliographie

- [1] ACPR, *Notice Solvabilité 2 – Modèles Internes*
- [2] Carole BETTONVILLE, Robin VAN OIRBEEK, Michel DENUIT, Julien TRUFIN, *Semi-Markov multistate individual loss reserving model in general insurance*, 2020
- [3] Alexandre BOUMEZOUED, Yoboua ANGOUA, Laurent DEVINEAU, Jean-Philippe BOISSEAU, *One-year reserve risk including a tail factor: closed formula and bootstrap approaches*, 2011
- [4] CAEIRO and GOMES, *Threshold selection in extreme value analysis. Extreme Value Modeling and Risk Analysis: Methods and Applications*, 69–86, 2016
- [5] Directive 2009/138/CE
- [6] P. D. ENGLAND and R. J. VERRALL, *Predictive distributions of outstanding liabilities in general insurance*, *Annals of Actuarial Science*, 2006
- [7] P. D. ENGLAND and R. J. VERRALL, *Stochastic Claims Reserving in General Insurance*, *British Actuarial Journal* 8, III, 2002
- [8] Eva FRAISSE, *Cours de réglementation des assurances*, ISUP, 2019-2020
- [9] GERSTENGARBE, F. WERNER, *A method for the statistical definition of extreme-value regions and their application to meteorological time series*, *ZMeteorol*, 1989
- [10] GROUPAMA, *Corpus méthodologique : calcul du SCR de souscription non-vie risque de réserves et de prime*
- [11] GROUPAMA, *Politique du modèle interne partiel non-vie*
- [12] GROUPAMA, *Rapport de modélisation du risque de réserve GMA*, 2020
- [13] GROUPAMA, *Rapport de modélisation du risque de prime : 04 – MTPL*, 2020
- [14] KRATZ and RESNICK, *The QQ-estimator and heavy tails*. *Stochastic Models*, 12(4), 699–724, 1996
- [15] Thomas MACK, *Distribution-free calculation of the standard error of the Chain-Ladder reserve estimates*, *Astin Bulletin*, Vol. 23, n°2, 1993
- [16] Michael MERZ and Mario V. WÜTHRICH, *Modelling the claims development result for solvency purposes*, *Casualty Actuarial Society E-Forum*, Fall, 2008
- [17] Michael MERZ et Mario V. WÜTHRICH, *Uncertainty of the Claims Development Result in the Chain-Ladder Method*, *Scandinavian Actuarial Journal*, 2008
- [18] Règlement délégué, 2015-035 CE