

**Mémoire présenté devant l'ENSAE Paris
pour l'obtention du diplôme de la filière Actuariat
et l'admission à l'Institut des Actuaires**

le 13/03/2023

Par : **Audrey Sentucq**

Titre: **Provisionnement en assurance non-vie : Extension
stochastique de méthodes usuelles et application au
calcul de l'ajustement pour risque sous IFRS 17**

Confidentialité : NON OUI (Durée : 1 an 2 ans)

Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus

Membres présents du jury de la filière

Pierre PICARD

Entreprise :

Nom : PwC

Signature :



*Membres présents du jury de l'Institut
des Actuaires*

Directeur de mémoire en entreprise :

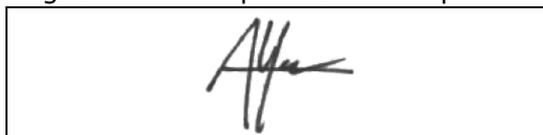
Nom : Alexandre YOU

Signature :



**Autorisation de publication et de
mise en ligne sur un site de
diffusion de documents actuariels**
(après expiration de l'éventuel délai de
confidentialité)

Signature du responsable entreprise



Secrétariat :

Signature du candidat

Bibliothèque :



Résumé

Contraints par un cycle de production inversé, les assureurs doivent évaluer au mieux le montant de leurs engagements envers les assurés et les tiers. En plus de la nécessité de calculer le montant des provisions en « Best Estimate », la nouvelle norme comptable IFRS 17 qui entrera en vigueur au 1^{er} janvier 2023 requerra le calcul d'un ajustement pour risque. Selon la définition de la norme IFRS 17, l'ajustement pour risque reflète la compensation requise par une entité d'assurance pour supporter l'incertitude des flux de trésorerie futurs résultants de risques non financiers. Dans l'approche par quantile, l'ajustement pour risque est calculé comme la différence entre le quantile au niveau choisi et le « Best Estimate ».

En assurance non-vie, de nombreuses méthodes existent pour estimer le montant « Best Estimate » des provisions. Cependant, pour un certain nombre de méthodes usuelles, il n'existe pas de mesure de l'incertitude associée à l'estimation. L'objectif de ce mémoire est de présenter des extensions stochastiques aux méthodes les plus couramment utilisées (Chain Ladder, Bornhuetter-Ferguson, Coût Moyen, Verbeek-Taylor) afin d'être en mesure d'obtenir non seulement une évaluation du « Best Estimate », mais aussi une estimation de la volatilité des provisions nécessaire au calcul de l'ajustement pour risque. Nous nous intéresserons également au calcul de l'erreur de prédiction au niveau d'un portefeuille ou d'une branche dans le cas où les méthodes de calcul du « Best Estimate » sont différentes selon l'exercice de survenance.

Les approches décrites dans ce mémoire seront d'une part illustrées avec leur application sur des triangles de données réelles issus de portefeuille d'assureurs IARD, et d'autre part comparées dans le cadre d'une étude par simulation.

Mots clés : Ajustement pour risque, Assurance Non-Vie, Best Estimate, Bornhuetter-Ferguson, Bootstrap, Chain Ladder, Coût moyen, Erreur de prédiction, IFRS 17, Provisionnement, Verbeek-Taylor.

Abstract

Constrained by an inverted production cycle, insurers must assess the liabilities to book in their balance as well as possible. In addition to the calculation of the “Best Estimate” reserve, under the new IFRS 17 accounting standard which is applicable on 1st January 2023, insurers will be required to calculate a risk adjustment. According to IFRS 17, the risk adjustment reflects the compensation that an entity requires for bearing the uncertainty about the amount and timing of the cash flows that arises from non-financial risk. With the so called “Quantile Approach”, the risk adjustment is obtained as the difference between the chosen quantile from distribution of reserves and the “Best Estimate”.

In non-life insurance, lot of methods are available to calculate the “Best Estimate”. Nonetheless, many of them, we cannot measure the uncertainty associated to our estimation. The objective of this thesis is to present a stochastic extension of the usual reserving methods (Chain Ladder, Bornhuetter-Ferguson, Average Cost, Verbeek-Taylor) in order to be able to obtain not only the « Best Estimate » reserve but also the volatility of reserves needed to calculate the risk adjustment. We will also be interested in the calculation of the prediction error at portfolio or line of business level when the methods for “Best Estimate” calculation will vary according to each accident year.

Approaches presented in this thesis will first be illustrated with real claims triangle coming from non-life insurance portfolios and then be compared in a simulation study.

Keywords: Average Cost, Best Estimate, Bootstrap, Bornhuetter-Ferguson, Chain Ladder, IFRS 17, Non-life insurance, Prediction error, Reserving, Risk Adjustment, Verbeek-Taylor.

Note de synthèse

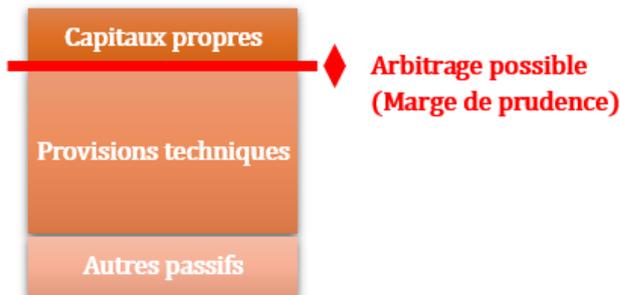
Contexte

Contraints par un cycle de production inversé, les assureurs doivent vendre des produits avant d'en connaître le coût définitif. Cette spécificité a deux conséquences pour les entités d'assurance :

- Elles doivent mettre en place des outils mathématiques pour évaluer le montant de la prime à demander à l'assuré (la tarification),
- Elles doivent mettre en place des outils mathématiques pour garantir qu'il y aura toujours au passif du bilan les sommes nécessaires au respect des engagements de l'assureur (le provisionnement).

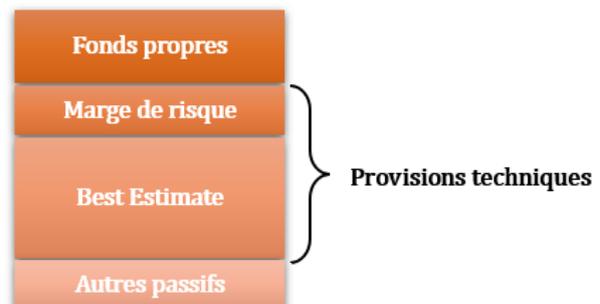
Dans ce mémoire, nous allons nous focaliser sur le provisionnement en assurance non-vie. L'une des provisions principales est la provision pour sinistre à payer (PSAP) qui correspond à une estimation des dépenses nécessaires pour régler tous les sinistres survenus et non payés. Elle se compose des provisions dossier/dossier calculées par les gestionnaires, des IBNER (Incurred But Not Enough Reported) sur les sinistres déclarés et des IBNYR (Incurred But Not Yet Reported) qui correspondent aux sinistres tardifs.

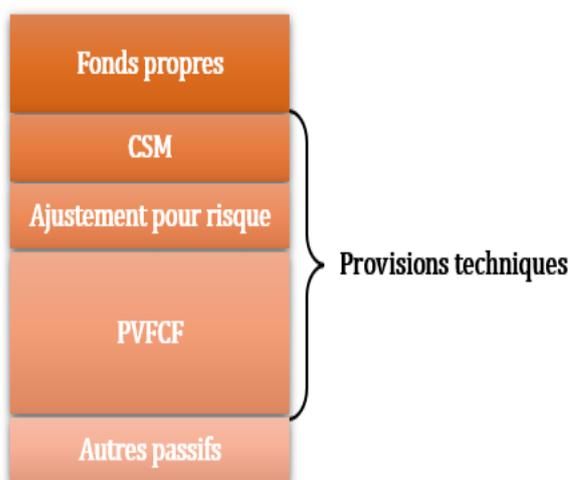
En fonction des normes, les provisions ne sont pas calculées de la même manière :



Selon la norme française, les provisions doivent être « *suffisantes pour le règlement intégral de leurs engagements vis-à-vis des assurés, des souscripteurs et bénéficiaires de contrats et des entreprises réassurées* » (Article 141-1).

Selon la directive 2009/138/CE (Solvabilité 2), les provisions techniques correspondent à « *la somme de la meilleure estimation (Best Estimate ou BE) et de la marge de risque (Risk Margin ou RM)* » (Article 77).





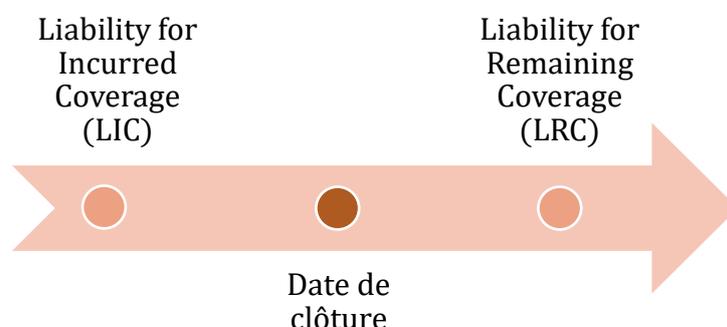
Selon IFRS 17, les provisions techniques doivent être composées de :

- la PVFCF (qui correspond à la meilleure estimation des engagements de l'assureur et sera dans la suite appelée provision en « Best Estimate »),
- l'ajustement pour risque (qui reflète la compensation que l'assureur requiert pour supporter l'incertitude des flux de trésorerie futurs résultants de risques non financiers, tant sur les montants que les dates d'occurrence),
- la CSM (qui représente les profits futurs attendus du contrat).

Dans ce mémoire la norme d'intérêt est la nouvelle norme comptable IFRS 17 qui entrera en vigueur au 1^{er} janvier 2023. Dans ce modèle, il existe trois modèles d'évaluation des éléments du passif :

- Un modèle général (BBA) applicable à tous les contrats,
- Un modèle pour les contrats ayant une durée courte (inférieure à 1 an) ou ceux dont l'estimation par ce modèle n'est pas fondamentalement différente du modèle classique (PAA) et,
- Un modèle pour les contrats à participation directe (VFA).

Dans le référentiel IFRS 17, les provisions se décomposent en deux parties :



Le LIC fait référence aux événements passés (par exemple les sinistres survenus mais non clôturés) et le LRC aux événements futurs (par exemple les sinistres futurs rattachés aux périodes de couverture pour lesquelles l'entité d'assurance est déjà engagée).

Dans ce mémoire, nous nous plaçons dans le cadre de l'assurance non-vie. La majorité des acteurs du marché va retenir le modèle simplifié (PAA) pour la comptabilisation des portefeuilles non-vie.

Dans ce modèle, le LRC se calcule comme les Primes Emises Non Acquisées diminuées des frais d'acquisition. Le LIC est lui égal à la somme d'un « Best Estimate » (BE LIC) et de l'ajustement pour risque (RA LIC).

Ainsi, dans le cadre du modèle PAA, le passif vaut :

$$Passif_{PAA} = PPNA_{LRC} - frais_{LRC} + BE_{LIC} + RA_{LIC}$$

L'évaluation va donc être centrée sur l'estimation du BE et du RA où le BE correspond à la PSAP en vision économique et le RA correspond au quantile de niveau α de la distribution des PSAP économiques. Il est donc nécessaire de connaître la distribution des provisions pour pouvoir calculer le RA.

Nous disposons pour notre étude de deux bases de données réelles de responsabilité civile automobile.

Notre premier jeu de données est composé uniquement de sinistres en affaires directes sur des polices françaises survenus entre 1996 et 2021. Afin d'avoir des groupes homogènes, les données ont été séparées par tranche de coût pour différencier :

- Les sinistres « Attritional » qui se caractérisent par une fréquence élevée mais une sévérité faible,
- Les sinistres « Large » qui se caractérisent par une fréquence faible et une sévérité élevée,
- Les sinistres « Exceptionnels ».

Le deuxième jeu de données concerne des sinistres en acceptation de réassurance sur des polices anglaises survenus entre 2002 et 2021.

Objectifs

L'objectif de ce mémoire est de présenter des extensions stochastiques aux méthodes les plus couramment utilisées (Chain Ladder, Bornhuetter-Ferguson, Coût Moyen, Verbeek-Taylor) afin de pouvoir estimer rigoureusement les erreurs de prédiction associées.

En effet, pour calculer les provisions « Best Estimate », il existe une vaste palette de modèles mathématiques plus ou moins complexes. En revanche, pour estimer la volatilité des provisions (*i.e.* l'erreur de prédiction) nécessaire au calcul de l'ajustement pour risque, le panel des méthodes à disposition de l'actuaire est plus restreint.

Dans la majorité des cas, les assureurs ont recours à la méthode de Mack qui fournit une formule fermée de l'erreur de prédiction associée au modèle Chain Ladder. Cependant, en pratique, les assureurs n'utilisent pas uniquement la méthode Chain Ladder pour calculer leurs provisions « Best Estimate ». Pour calculer l'erreur de prédiction dans le cas où ils utilisent une autre méthode, les assureurs font généralement appel à une approche de *scaling* : en calculant le montant des provisions avec la méthode choisie puis avec la méthode Chain Ladder dont l'erreur de prédiction est calculable via la formule de Mack, on approxime l'erreur de prédiction de la méthode choisie en multipliant l'erreur de prédiction du modèle Chain Ladder par le ratio « Best Estimate / Provision Chain Ladder ». Une des problématiques est alors de savoir si cette approximation est cohérente.

Les approches décrites dans ce mémoire seront d'une part illustrées avec leur application sur les triangles de données réelles présentées ci-dessus, et d'autre part comparées dans le cadre d'une étude par simulation.

Méthodes déterministes

Les méthodes déterministes nous permettent de calculer les provisions « Best Estimate ». Dans la majorité des cas, pour pouvoir utiliser les méthodes, il est nécessaire de valider des hypothèses et une façon d'y parvenir, est d'utiliser des pondérations afin de sélectionner les données que nous utilisons.

Une fois les hypothèses validées, le montant des provisions varie en fonction des méthodes utilisées.

Ainsi, pour nos données, nous obtenons les résultats suivants :

	Classe « Attritional »	Classe « Large »	Classe « Exceptionnelle »	Portefeuille Anglais
Chain Ladder		402		
Chain Ladder pondéré	-28	509	241	388
Bornhuetter-Ferguson pondéré	-6	556	375	336
Coût Moyen			336	
Coût Moyen pondéré			280	
Verbeek-Taylor			246	

Tableau 1 : Récapitulatif des IBNR calculés pour chaque méthode

Méthodes stochastiques

Les méthodes stochastiques vont permettre d'estimer l'erreur de prédiction associée aux « Best Estimate » calculés précédemment. La mesure utilisée pour quantifier l'erreur qui est commise est la MSEP (Mean Square Error of Prediction) définie comme :

$MSEP_{\theta}(\hat{\theta}) = E \left[(\theta - \hat{\theta})^2 \middle| D \right]$ avec θ la variable aléatoire, $\hat{\theta}$ le prédicteur et D les données.

En appliquant les méthodes stochastiques sur nos données, nous obtenons le tableau ci-dessous.

	Classe « Attritional »	Classe « Large »	Classe « Exceptionnelle »	Portefeuille Anglais
Chain Ladder		8,3%		
Chain Ladder pondéré	15,0%	9,8%	13,3%	9,1%
Bornhuetter-Ferguson pondéré	10,7%	7,8%	11,7%	7,6%
Coût Moyen			10,4%	
Coût Moyen pondéré			13,2%	
Verbeek-Taylor			11,1%	

Tableau 2 : Récapitulatif des erreurs de prédiction en fonction des méthodes (en % des PSAP)

En l'absence d'informations, nous pouvons choisir de garder la méthode qui donne l'erreur de prédiction la plus faible.

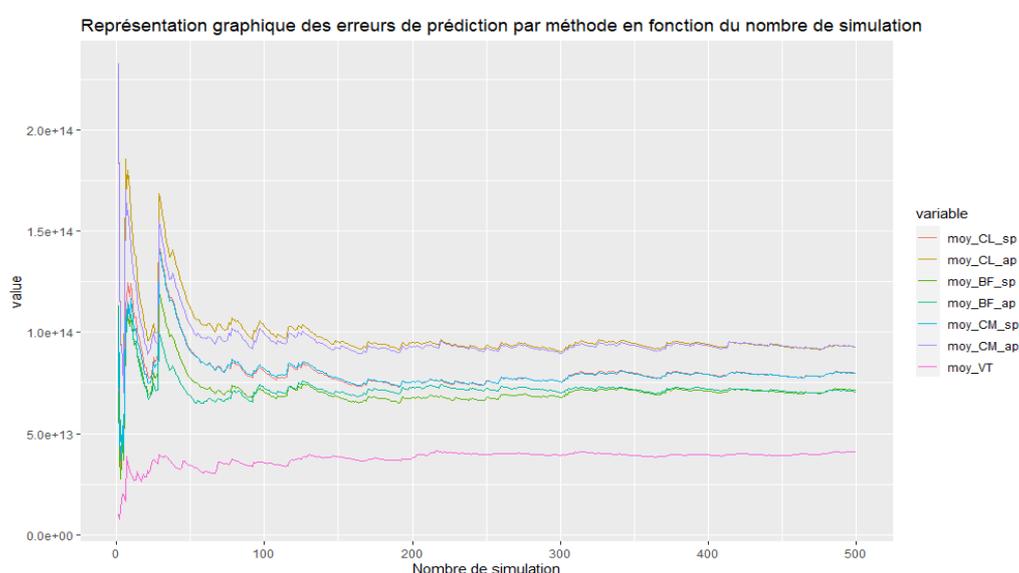
Cela ne veut pas dire que l'estimation de la méthode avec l'erreur la plus faible est meilleure mais simplement que la volatilité est plus faible et donc le risque d'erreur s'en retrouve amoindri.

Il semblerait que la méthode Bornhuetter-Ferguson donne les volatilités les plus faibles pour les données du portefeuille anglais, tout comme pour les classes « Attritional » et « Large » des données issues du portefeuille français.

Pour la classe « Exceptionnelle », c'est la méthode du coût moyen, suivie de très près par la méthode Verbeek-Taylor, qui donne l'erreur la plus faible.

Comparaison des méthodes

Les observations faites précédemment s'appliquent sur nos données seulement. Il est intéressant de regarder s'il est possible de généraliser les observations précédentes. En utilisant des données simulées, nous avons observé l'erreur de prédiction des méthodes :



Les observations faites précédemment semblent se vérifier et en testant plusieurs sensibilités, nous remarquons que les hiérarchies se confirment.

Agrégation des méthodes

Dans la pratique, il est parfois possible d'utiliser plusieurs méthodes sur une même base de données en différenciant les années de survenance ou en segmentant nos données en classes. C'est ce qui s'est passé pour le portefeuille français. S'il est possible d'additionner les montants des provisions calculés à partir de plusieurs méthodes différentes, les erreurs de prédiction n'étant pas additives, l'erreur de prédiction globale ne peut pas se calculer comme une somme des erreurs de chacune des méthodes. Pour approximer l'erreur de prédiction globale, il est possible d'utiliser la formule suivante :

$$\sum_{m=1}^M MSEP(R_m) + 2 \sum_{1 \leq m < l \leq M} \rho_{l,m} \left(\sqrt{Var(\hat{R}_m|D)} \sqrt{Var(\hat{R}_l|D)} + \sqrt{Var(R_m|D)} \sqrt{Var(R_l|D)} \right)$$

Dans cette formule, M représente le nombre de méthodes différentes, m est l'indice de la méthode utilisée, R est le montant des provisions et \hat{R} le montant estimé de ces provisions.

En appliquant cette formule au portefeuille français, nous obtenons un pourcentage d'erreur global à 7,6% et profitons de l'effet de diversification.

Calcul de l'ajustement pour risque (RA)

Aucune méthode standard n'est prescrite par la norme IFRS 17. Le calcul doit simplement vérifier les principes généraux suivants (paragraphe B91) :

- **Cohérence avec le niveau de risque** : si la distribution des réserves de la branche 1 a une queue épaisse (niveau de risque élevé) et si la distribution des réserves de la branche 2 a une queue fine (niveau de risque plus faible) alors $RA_1 > RA_2$;
- **Cohérence avec la durée des engagements** : l'ajustement pour risque augmente avec la durée de projection des cash flows ;
- **Cohérence avec les données** : s'il y a beaucoup d'incertitude, l'ajustement pour risque doit être plus élevé que si l'incertitude est faible ;
- **Cohérence avec la sévérité** : si les risques sont peu fréquents mais graves, l'ajustement pour risque sera plus élevé que si les risques sont fréquents mais peu graves.

Les deux méthodes les plus utilisées sont l'approche par coût du capital dérivée du calcul de Risk Margin sous Solvabilité 2 et l'approche par quantile. En assurance non-vie, on constate qu'une majorité d'acteurs ont adopté la seconde approche, celle-ci étant déjà fréquemment mise en œuvre pour fixer un niveau de prudence explicite dans le provisionnement en comptes statutaires. Dans cette approche, l'ajustement pour risque se calcule comme la différence entre une mesure de risque d'un niveau de confiance α et la moyenne (Best Estimate).

Nous avons choisi d'utiliser la VaR de niveau 80% pour calculer la valeur de l'ajustement pour risque car cette mesure de risque a l'avantage d'être facilement interprétable.

Les modèles stochastiques que nous avons utilisés donnent une estimation de l'espérance et de la variance mais ne fournissent pas la distribution complète des provisions (sauf pour le cas de la méthode par séparation qui utilise le Bootstrap). Cependant, pour utiliser les mesures de risque définies précédemment, nous avons besoin de formuler une hypothèse sur cette distribution. Nous avons choisi de regarder les lois lognormale, normale et gamma, qui sont les distributions usuellement retenues en pratique par les assureurs non-vie.

Nous obtenons les montants suivants :

	Classe « Attritional »	Classe « Large »	Classe « Exceptionnelle »	Portefeuille Anglais
Montant de l'ajustement pour risque	9,81	148,03	332,55	699,42

Conclusion

Les méthodes présentées précédemment permettent de calculer l'ajustement pour risque mais nécessitent de faire une hypothèse sur la distribution des provisions. Une des continuités possibles serait de développer des méthodes non paramétriques capables de prendre en compte les pondérations nécessaires au bon fonctionnement des méthodes.

Executive Summary

Context

Constrained by an inverted production cycle, insurers sell their products before knowing exactly their costs. This specificity has two consequences:

- Insurers must assess as well as possible the premium to ask (pricing);
- Insurers must assess the liabilities to book in their balance to be able to honor their engagements (reserving).

In this thesis, we will focus on reserving in the non-life insurance field. Claims reserve is one of the major items of the balance sheet of every general (non-life) insurance company. It can vary from one company to another, but in general, it constitutes a major part of technical provision. It is an estimate of the expenses required to settle all claims incurred but unpaid. It consists of the file/file reserve, the IBNER (Incurred But Not Enough Reported) and the IBNYR (Incurred But Not Yet Reported).

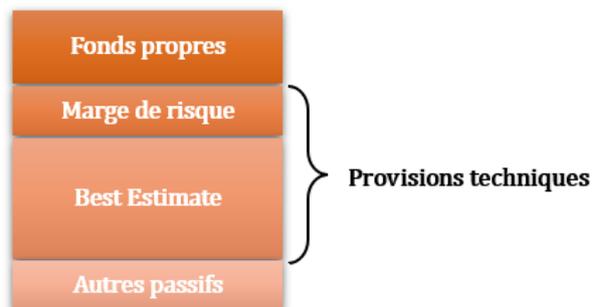
According to the standards, reserves are not calculated in the same way:

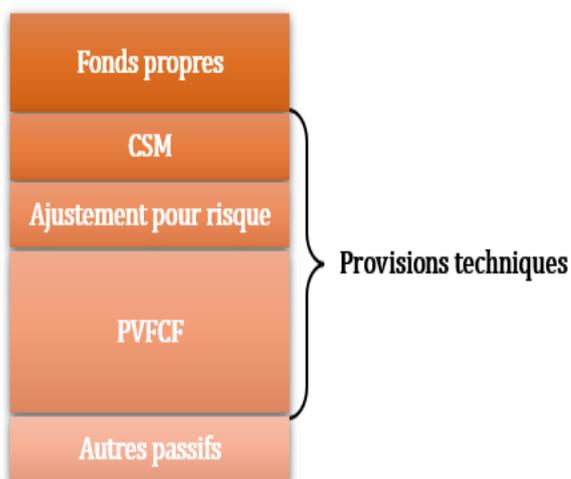


**Arbitrage possible
(Marge de prudence)**

According to the French standard, claims reserves must be « *suffisantes pour le règlement intégral de leurs engagements vis-à-vis des assurés, des souscripteurs et bénéficiaires de contrats et des entreprises réassurées* » (Section 141-1).

According to Directive 2009/138/CE (Solvency 2), claims reserves correspond to « *the sum of the Best Estimate and Risk Margin* » (Section 77).





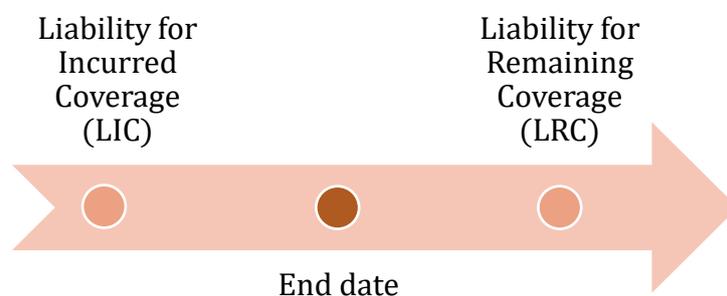
The new IFRS 17 accounting rules impose the booking of:

- A PVFCF (which is the “Best Estimate” of the insurer’s engagements),
- A risk adjustment (to reflect the compensation that the entity requires for bearing the uncertainty about the amount and timing of the cash flows that arise from non-financial risk),
- A CSM (for unearned profit).

In this thesis ,we will focus on the new IFRS 17 accounting standard which is applicable on 1st January 2023. To assess the liabilities, three measurement models exist:

- One applicable to all contracts (BBA) ;
- One applicable for liability of remaining coverage equal to one year or less, or applicable if this approach is not materially different from the BBA (PAA) ;
- One for participation features (VFA).

Under IFRS 17, provisions are divided into two parts:



LIC is for past events (for instance a claim has occurred but is not settled) and LRC for future ones (for example, future claims related to the coverage periods for which the insurance entity is already engaged).

In this thesis, we will focus on non-life insurance and most market participants will use the simplified model (PAA) for non-life portfolios.

In that case, LRC is calculated as unearned written premiums less acquisition costs and LIC is equal at the sum of a « Best Estimate » (BE LIC) and a risk adjustment (RA LIC).

Thus, dealing with PAA measurement models, liability equals to :

$$Liability_{PAA} = Unearned\ WP_{LRC} - aquisition\ costs_{LRC} + BE_{LIC} + RA_{LIC}$$

The evaluation will therefore focus on the estimation of the BE and the RA where the BE corresponds to the claim reserve in economic vision, and the RA corresponds to the level quantile α of the distribution of economic claim reserve. It is therefore necessary to know the distribution to calculate the RA.

For our study, we have two actual databases of automobile liability.

Our first data set consists solely of direct business claims on French policies that occurred between 1996 and 2021. In order to have homogeneous groups, the data were separated by cost band to differentiate:

- « Attritional » claims which are characterized by high frequency but low severity,
- « Large » claims which are characterized by high severity but low frequency,
- « Exceptional » claims.

The second dataset relates to reinsurance acceptance claims on English policies that occurred between 2002 and 2021.

Objectives

The objective of this thesis is to present stochastic extensions to the most commonly used methods (Chain Ladder, Bornhuetter-Ferguson, Average Cost, Verbeek-Taylor) in order to accurately estimate the associated prediction errors.

Indeed, to calculate the Best Estimate provisions, there is a wide range of more or less complex mathematical models. On the other hand, to estimate the volatility of the provisions (i.e., the prediction error) necessary to calculate the risk adjustment, the range of methods available to the actuary is more limited.

In most cases, insurers use the Mack method, which provides a formula for the prediction error associated with the Chain Ladder model. However, in practice, insurers do not only use the Chain Ladder method to calculate their "Best Estimate". To calculate the prediction error in the event that they use another method, insurers generally use a scaling approach. By calculating the "Best Estimate" with the chosen method and then with the Chain Ladder method whose prediction error can be calculated using the Mack formula, they approximate the prediction error of the chosen method by multiplying the Chain Ladder model prediction error by the "Best Estimate / Chain Ladder Estimate" ratio.

One of the problems is the consistency of that proxy.

The approaches described in this paper will be illustrated with their application to the real triangles arising from the non-life insurance portfolios presented above and compared in a simulation study.

Deterministic methods

Deterministic methods allow us to calculate the « Best Estimate ». Nonetheless, to use the methods, we need to verify their assumptions and sometimes, the data we use do not allow us to verify them.

One way to apply the method is then to use weights to select the data we use. Once the assumptions have been validated, we can use the methods to calculate the “Best Estimate”.

The amount of claim reserves varies according to the methods used. Here are the IBNR we obtain using the method on our data:

	« Attritional »	« Large »	« Exceptional »	English portfolio
Chain Ladder		402		
Weighted Chain Ladder	-28	509	241	388
Weighted Bornhuetter-Ferguson	-6	556	375	336
Average Cost			336	
Weighted Average Cost			280	
Verbeek-Taylor			246	

Table 3 : IBNR for each method

Stochastic methods

Stochastic methods will allow us to estimate the prediction error associated with the «Best Estimate» calculated previously. The measure used to quantify the error is the MSEP (Mean Square Error of Prediction) define as:

$$MSEP_{\theta}(\hat{\theta}) = E \left[(\theta - \hat{\theta})^2 \middle| D \right]$$

with θ a random variable, $\hat{\theta}$ a predictor and D the data.

By applying stochastic methods to our data, we obtain the table below.

	« Attritional »	« Large »	« Exceptional »	English portfolio
Chain Ladder		8,3%		
Weighted Chain Ladder	15,0%	9,8%	13,3%	9,1%
Weighted Bornhuetter-Ferguson	10,7%	7,8%	11,7%	7,6%
Average Cost			10,4%	
Weighted Average Cost			13,2%	
Verbeek-Taylor			11,1%	

Table 4 : Summary of Prediction Errors (% reserves)

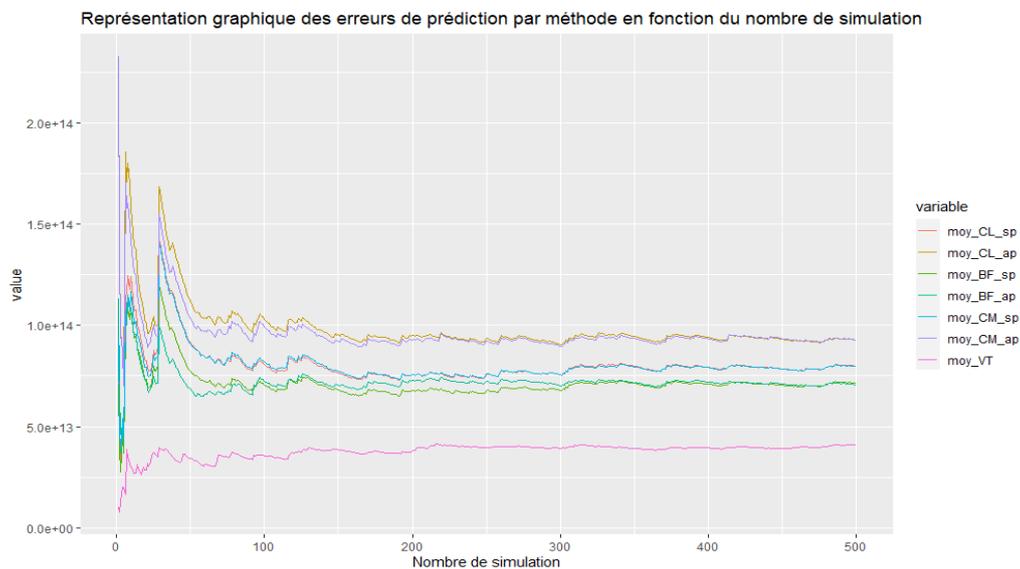
We may choose to keep the method that gives the lowest prediction error.

This does not mean that the estimation of the method with the lowest error is better, but simply that the volatility is lower and therefore the risk of error is reduced.

The Bornhuetter-Ferguson method appears to have the lowest volatility for the English portfolio data, as well as for the Attritional and Large classes of the French portfolio data. For the “Exceptional” class, it is the average cost method, followed very closely by the Verbeek-Taylor method, which gives the lowest error.

Comparing methods

The observations we made above apply to our data only. It is interesting to see if it is possible to generalize these observations. Using simulated data, we can observe the real prediction error calculated as the difference between the true final amount of claims and the expectations we calculated thanks to the deterministic methods:



The observations made previously seem to be true and by testing several sensitivities, we notice that the hierarchies are confirmed.

Method's aggregation

In practice, it is sometimes possible to use several methods on the same database by differentiating the years of occurrence or by segmenting our data into classes. That's what happened to the French portfolio.

If it is possible to add the « Best Estimate » reserve calculated from several different methods, it is not the case for prediction errors that are not additive. That's why the global prediction error cannot be calculated as a sum of the errors of each method.

To approximate the global prediction error, the following formula can be used:

$$\sum_{m=1}^M MSEP(R_m) + 2 \sum_{1 \leq m < l \leq M} \rho_{l,m} \left(\sqrt{Var(\hat{R}_m|D)} \sqrt{Var(\hat{R}_l|D)} + \sqrt{Var(R_m|D)} \sqrt{Var(R_l|D)} \right)$$

In this formula, M represents the number of different methods, m is the index of the method used, R is the amount of reserves and R is the estimated amount of those reserves.

By applying this formula to the French portfolio, we obtain an overall error percentage of 7.6% and benefit from the diversification effect.

Calculation of the risk adjustment (RA)

No standard method is prescribed by IFRS 17. The calculation should simply verify the following general principles (Subsection B91):

- **Consistency with the level of risk:** if the distribution of Branch 1 reserves has a thick tail (high risk level) and if the distribution of Branch 2 reserves to a thin tail (lower risk level) then $RA_1 > RA_2$;
- **Consistency with the duration:** the risk adjustment must increase with the projection period of cash flows ;
- **Consistency with data:** if there is a high uncertainty, the risk adjustment should be higher than if the uncertainty is low ;
- **Consistency with severity :** if the risks are infrequent but severe, the risk adjustment will be higher than if the risks are frequent but not severe.

The two most commonly used methods are the cost of capital approach derived from the Risk Margin calculation under Solvency 2 and the quantile approach. In non-life insurance, we note that a majority of actors have adopted the second approach, which is already frequently implemented to set an explicit level of prudence in statutory accounts. In this approach, the risk adjustment is calculated as the difference between a risk measure of a confidence level α and the mean (Best Estimate).

We have chosen to use the 80% level VaR to calculate the value of the risk adjustment because this risk measure has the advantage of being easily interpretable.

The stochastic models we used give an estimate of the mean (thanks to the “Best Estimate”) and variance (thanks to prediction errors) but do not provide the complete distribution of provisions (except for the case of the separation method that uses the Bootstrap). However, to use the risk measures defined above, we need to formulate an assumption about this distribution. We have chosen to look at the lognormal, normal and gamma laws, which are the distributions usually retained in practice by non-life insurers.

We get the following results (which are different from those calculated with *scaling*):

	« Attritional »	« Large »	« Exceptional »	English Portfolio
Risk Adjustment	9,81	148,03	332,55	699,42

Conclusion

The methods presented above make it possible to calculate the risk adjustment but require an assumption on the reserve distribution. One of the possible continuities would be to develop non-parametric methods capable of considering the weightings necessary for the proper functioning of the methods.

Remerciements

Je tiens tout d'abord à remercier l'ensemble du cabinet PwC pour son accueil et pour m'avoir donné l'opportunité de finaliser mes études au sein de leur société. Merci également à l'ensemble des équipes actuariat et audit pour m'avoir bien intégrée.

Plus particulièrement, je remercie mon tuteur Alexandre YOU pour son investissement, ses conseils et pour toutes les connaissances qu'il m'a apportées tout au long de mon stage.

Je tiens également à remercier tous mes anciens collègues de Chubb, auprès de qui j'ai pu réaliser deux années d'alternance, pour leur gentillesse, leur patience alors que je débutais dans le monde de l'actuariat ainsi que pour tous les bons moments passés ensemble. Vous avez contribué à la réussite de ma première expérience professionnelle en actuariat.

J'exprime également ma gratitude envers le corps enseignant, tant de l'ENSAE que de l'Université Gustave Eiffel, pour la qualité de la formation qui m'a été dispensée.

Enfin, je remercie mes amis et ma famille pour leurs précieux encouragements et leur soutien sans faille.

Table des matières

Introduction	21
1. Présentation du contexte.....	23
1.1 Généralités sur l'assurance non-vie.....	23
1.1.1 Historique de l'assurance et définition.....	23
1.1.2 Caractéristiques de l'assurance	25
1.2 Présentation des provisions	26
1.2.1 La réglementation des provisions	26
1.2.2 La PSAP et le cycle de vie des sinistres	34
1.3 Notations et triangles de provisionnement	36
2. Présentation des données	37
2.1.1 Présentation des jeux de données réelles.....	38
2.1.2 Base simulée.....	39
3. Méthodes déterministes classiques de provisionnement.....	47
3.1 Méthode Chain Ladder.....	47
3.1.1 Méthode classique.....	47
3.1.2 Extension pondérée.....	53
3.2 Méthode de Bornhuetter-Ferguson	58
3.2.1 Méthode classique.....	58
3.2.2 Méthode indépendante (Mack (2006)).....	60
3.2.3 Méthode indépendante pondérée.....	65
3.3 Méthode des coûts moyens.....	70
3.4 Méthodes factorielles	75
3.4.1 Méthode de De Vylder	75
3.4.2 Méthode par séparation	77
4. Evaluation de l'incertitude dans l'estimation des provisions techniques	84
4.1 Extension stochastique de la méthode Chain Ladder	85
4.1.1 Chain Ladder stochastique (Mack (1993)).....	85
4.1.2 Chain Ladder stochastique pondéré (Mack (1999)).....	88
4.2 Méthode de Bornhuetter-Ferguson stochastique	91
4.2.1 Modèle de Mack (2008).....	91
4.2.2 Extension pondérée.....	95

4.2.3	Application aux données	96
4.3	Méthode du coût moyen stochastique	100
4.3.1	Calcul des termes d'espérance	101
4.3.2	Calcul des termes de variance	102
4.3.3	Calcul et estimation de la MSEP	105
4.3.4	Application aux données	108
4.4	Modèle factoriel stochastique	110
4.4.1	Une version stochastique de la méthode de De Vylder : le modèle de Christofides	110
4.4.2	Méthode de Verbeek-Taylor stochastique : le Bootstrap paramétrique	114
4.5	Agrégation des méthodes par formule fermée pour le calcul de l'erreur globale	118
4.5.1	Formule fermée pour le calcul de l'erreur globale pour deux méthodes ...	118
4.5.2	Formule fermée pour le calcul de l'erreur globale avec M méthodes.....	119
4.5.3	Application aux données	120
5.	Comparaison des méthodes.....	123
5.1	Variation des paramètres de fréquence	125
5.2	Variation des paramètres de coûts	128
5.3	Variation de la durée de déclaration	130
5.4	Variation de la durée de traitement.....	131
5.5	Variation de l'inflation	132
6.	Application des méthodes stochastiques au calcul de l'ajustement pour risque au titre des sinistres survenus sous IFRS 17	134
6.1	Méthodes de calcul de l'ajustement pour risque	134
6.2	Calcul de l'ajustement pour risque à l'aide de l'approche par quantiles.....	135
6.2.1	Définition des paramètres	135
6.2.2	Application aux données	137
	Conclusion	143
	Bibliographie.....	145
	Annexes	147

Introduction générale

Le secteur de l'assurance est caractérisé par un cycle de production inversé : les compagnies d'assurances proposent des couvertures dont le coût réel n'est connu que postérieurement aux prestations versées. Afin de protéger les assurés et les tiers, la réglementation impose aux entités d'assurance de constituer des provisions afin de pouvoir honorer en toutes circonstances leurs prestations. Les assureurs doivent donc évaluer au mieux leurs engagements pour constituer les provisions nécessaires à la poursuite de leurs activités.

Avec l'entrée en vigueur de la norme IFRS 17 au 1^{er} janvier 2023, les entités d'assurance soumises à la norme devront évaluer leurs engagements comme la somme d'une provision « Best Estimate » et d'un ajustement pour risque. L'ajustement pour risque reflète la compensation requise par une entité d'assurance pour supporter l'incertitude des flux de trésorerie futurs résultants de risques non financiers. Dans l'approche dite par quantile, l'ajustement pour risque est calculé comme la différence entre le quantile au niveau choisi et le « Best Estimate ».

En assurance non-vie, pour calculer les provisions « Best Estimate », il existe une vaste palette de modèles mathématiques plus ou moins complexes. En revanche, pour estimer la volatilité des provisions (*i.e.* l'erreur de prédiction) nécessaire au calcul de l'ajustement pour risque, le panel des méthodes à disposition de l'actuaire est plus restreint.

Dans la majorité des cas, les assureurs ont recours à la méthode de Mack qui fournit une formule fermée de l'erreur de prédiction associée au modèle Chain Ladder. Cependant, en pratique, les assureurs n'utilisent pas uniquement la méthode Chain Ladder pour calculer leurs provisions « Best Estimate ». Pour calculer l'erreur de prédiction dans le cas où ils utilisent une autre méthode, les assureurs font généralement appel à une approche de « *scaling* » : en calculant le montant des provisions avec la méthode choisie puis avec la méthode Chain Ladder dont l'erreur de prédiction est calculable via la formule de Mack, on approxime l'erreur de prédiction de la méthode choisie en multipliant l'erreur de prédiction du modèle Chain Ladder par le ratio « Best Estimate / Provision Chain Ladder ». Une des problématiques est alors de savoir si cette approximation est cohérente.

L'objectif de ce mémoire est de présenter des extensions stochastiques aux méthodes les plus couramment utilisées (Chain Ladder, Bornhuetter-Ferguson, Coût Moyen, Verbeek-Taylor) afin de pouvoir estimer rigoureusement les erreurs de prédiction associées et ainsi être en mesure de calculer l'ajustement pour risque de la norme IFRS 17.

La première partie de ce mémoire abordera le contexte général de notre étude. Après une brève définition de l'assurance non-vie, nous présenterons les normes auxquelles les entités d'assurance doivent se soumettre ainsi que les provisions à constituer qui en découlent.

Nous présenterons ensuite les données utilisées dans l'ensemble du mémoire. Nous commencerons par les données issues de portefeuille d'assurance puis nous introduirons des données simulées qui nous permettront dans la suite de comparer les différentes méthodes.

Dans une troisième partie, nous introduirons les méthodes déterministes usuelles de provisionnement et les appliquerons aux données de nos portefeuilles. Nous commencerons par présenter la méthode Chain Ladder, véritable référence dans le domaine du provisionnement non-vie. Nous présenterons ensuite la méthode de Bornhuetter-Ferguson, alternative populaire pour pallier les points faibles de Chain Ladder. Enfin, nous introduirons la méthode du coût moyen qui utilise un procédé similaire à celui de la tarification ainsi que la méthode par séparation qui permet notamment d'appréhender l'inflation.

Nous chercherons ensuite des extensions stochastiques aux méthodes présentées dans le chapitre précédent. Nous nous intéressons ainsi aux travaux de T. Mack qui permettent d'étendre au cas stochastique les méthodes de Chain Ladder et de Bornhuetter-Ferguson et adapterons ces modèles pour y intégrer des pondérations. Nous proposerons ensuite une extension stochastique pour la méthode du coût moyen. Enfin, nous nous intéresserons aux travaux de S. Björkwall qui permet, grâce au Bootstrap, d'estimer l'erreur de prédiction de la méthode de Verbeek-Taylor.

Une fois que nous aurons appliqué les extensions stochastiques des méthodes sur nos données de portefeuille, nous allons chercher à étudier leurs performances de manière plus générale en utilisant des données simulées. En effet, en utilisant le montant ultime simulé, il est possible de calculer l'erreur de prédiction réelle comme la différence entre le montant ultime simulé et le montant ultime calculé par les différentes méthodes.

Enfin, dans le dernier chapitre, nous appliquerons les méthodes au calcul de l'ajustement pour risque de la norme IFRS 17.

1. Présentation du contexte

Nous allons commencer par définir l'assurance et présenter plus particulièrement l'assurance non-vie qui sera l'objet de ce mémoire. Nous allons ensuite détailler le principe du provisionnement en assurance non-vie.

1.1 Généralités sur l'assurance non-vie

1.1.1 Historique de l'assurance et définition

L'assurance est une pratique très ancienne dont les premières traces remontent à l'Antiquité, période où les commerçants se répartissaient les coûts engendrés par les vols et les pillages. Il faudra néanmoins attendre le XIV^e siècle et les grandes expéditions maritimes pour trouver les prémices de l'assurance moderne sous la forme du « prêt à la grande aventure ». Le principe était très simple : les expéditions étaient financées par des investisseurs qui inscrivaient leur nom sous le bateau qu'ils acceptaient de couvrir (pour « souscrire le contrat ») et si le bateau faisait naufrage, le marchand n'avait rien à rembourser. En revanche, si le bateau arrivait à destination, les marchands devaient rembourser leurs investisseurs et ajouter une compensation financière pour le risque pris.

De nos jours, la définition de l'assurance est plus précise et est explicitée dans des textes de loi comme le Code Civil par exemple. Le contrat d'assurance peut être simplement défini comme une convention par laquelle, en contrepartie d'une prime, l'assureur s'engage à garantir le souscripteur en cas de réalisation d'un risque aléatoire prévu au contrat. Le caractère aléatoire du risque est primordial et l'évènement déclencheur doit être extérieur et indépendant de l'assuré.

Le contrat d'assurance réunit donc plusieurs parties :

- L'assureur comme entité qui perçoit les primes et fournit une prestation en cas de réalisation du risque,
- Le souscripteur : la personne (physique ou morale) qui conclut le contrat avec l'assureur et lui verse les primes,
- Le bénéficiaire : la personne qui reçoit la prestation en cas de réalisation du risque.

Pour pouvoir exercer, les entreprises d'assurance doivent obtenir l'agrément nécessaire auprès de l'autorité de contrôle (ACPR). Les agréments sont octroyés par branche. Voici le détail des branches classées par type (assurances de personnes ou assurances de dommages) et par branche générale (vie ou non-vie/IARD).

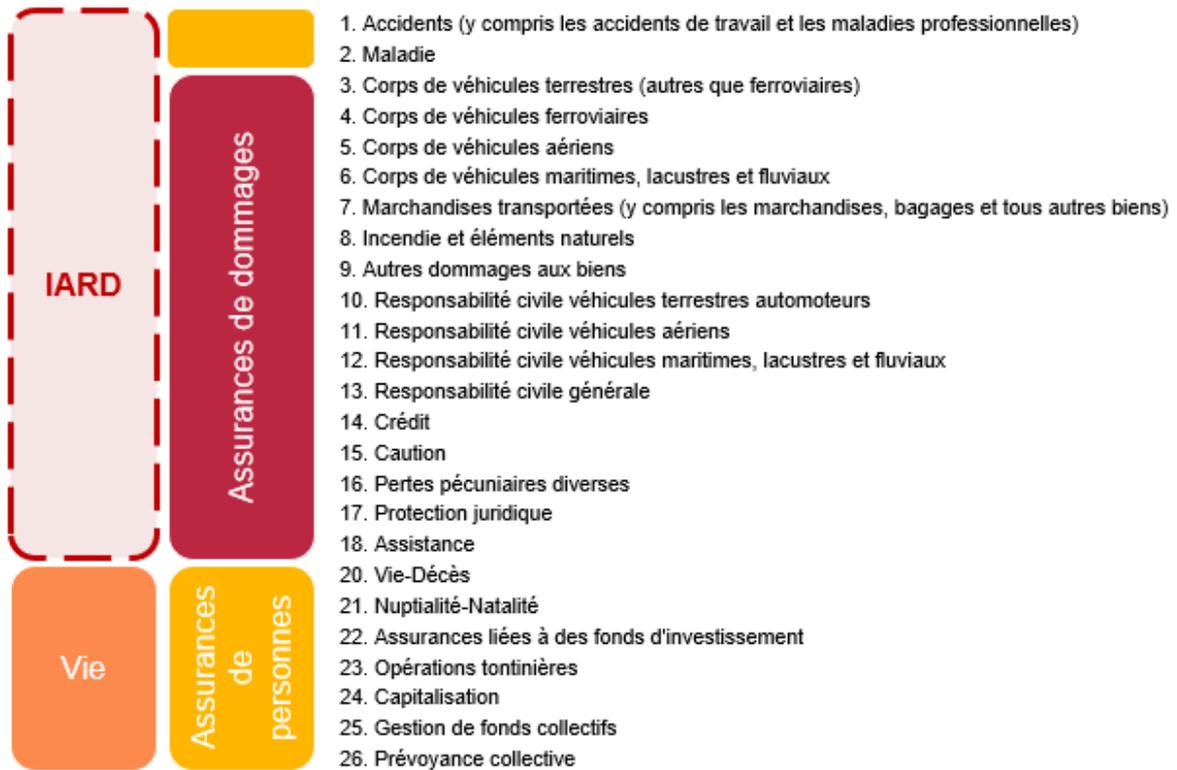


Figure 1.1: Présentation des branches

La différence entre les deux types d'assurance (personnes ou dommages) provient de la manière d'indemniser l'assuré : principe indemnitaire pour les assurances dommages et principe forfaitaire pour les assurances de personnes.

La différence entre les deux grandes branches d'assurance existantes (l'assurance vie et l'assurance non-vie) provient de la manière de gérer les primes. En effet, la réglementation française impose une gestion par capitalisation pour l'assurance vie et par répartition pour l'assurance non-vie.

- La gestion par capitalisation consiste à placer les primes collectées pour générer des revenus financiers. Ces revenus financiers seront ajoutés aux montants de primes initiaux pour constituer le capital permettant d'indemniser les assurés. Cette gestion n'est possible que pour des contrats souscrits pour une longue période.
- La gestion par répartition implique quant à elle que les cotisations payées par l'ensemble des assurés sur l'année soient utilisées pour indemniser les victimes de cette même année.

Le cadre général de l'assurance peut être résumé par le schéma ci-dessous.

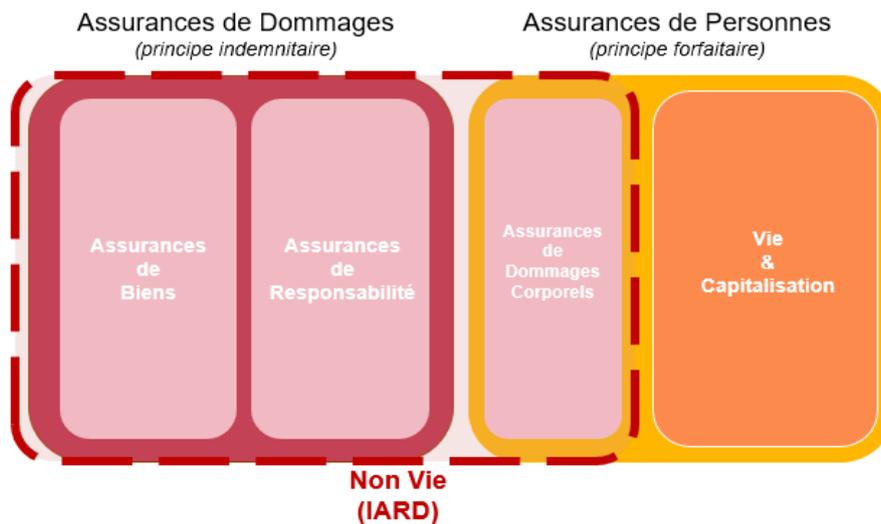


Figure 1.2 : Schéma de la classification des contrats d'assurance

1.1.2 Caractéristiques de l'assurance

La principale caractéristique de l'assurance est d'avoir un cycle de production inversé. Dans la majorité des cas, les coûts d'un produit sont connus à l'avance et le prix est fixé en conséquence. Dans le cas de l'assurance, les coûts sont incertains et l'assureur demande une prime sans connaître ni le montant réel des sinistres ni leur date de survenance.

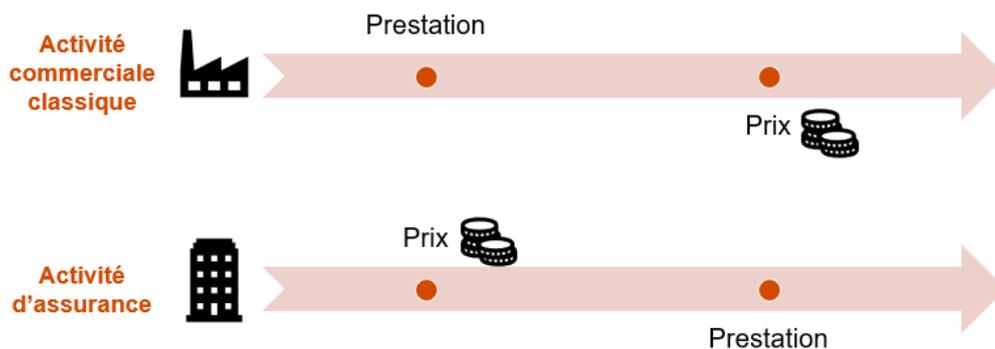


Figure 1.3 : Schéma de l'inversion du cycle de production

Cette spécificité a deux conséquences :

- La nécessité de mettre en place des outils mathématiques pour évaluer le montant de la prime à demander à l'assuré (la tarification),
- La nécessité de mettre en place des outils mathématiques pour garantir qu'il y aura toujours au passif du bilan les sommes nécessaires au respect des engagements de l'assureur (le provisionnement).

Une autre spécificité de l'assurance est l'asymétrie d'information car l'assureur ne connaît pas l'assuré. Pour se protéger contre son manque d'informations, l'assureur peut mettre en place des questionnaires, des franchises, des segmentations de tarifs...

1.2 Présentation des provisions

Du fait du cycle inversé de production, l'assureur ne connaît pas sa charge finale. La constitution de réserves est donc une étape essentielle car l'assureur doit toujours être en mesure d'honorer ses engagements. Nous nous intéressons dans ce mémoire plus particulièrement à l'évaluation des provisions nécessaires en assurance non-vie.

Les provisions techniques (provisions propres à l'activité d'assurance) sont inscrites au passif du bilan des compagnies d'assurances et représentent une large part de ce dernier (environ 80%). Il est important d'avoir une bonne estimation des provisions car une sous-estimation peut conduire à l'insolvabilité et une surestimation à une baisse du résultat (et donc une baisse de la charge fiscale). De ce fait, le montant des provisions est fortement contrôlé par le régulateur et l'administration fiscale.

1.2.1 La réglementation des provisions

Selon le Code des Assurances, il existe 7 catégories de provisions techniques en assurance IARD (Article R343-7) :

- La provision mathématique (PM) des rentes qui correspond à la valeur actuelle des engagements de l'entreprise en ce qui concerne les rentes en service ;
- La provision pour primes non acquises (PPNA) qui permet de constater, pour l'ensemble des contrats en cours, la part des primes émises et des primes restant à émettre se rapportant à la période comprise entre la date de l'inventaire et la date de la prochaine échéance de prime ou, à défaut, du terme du contrat ;
- La provision pour risque en cours (PREC) qui couvre la charge de sinistres et les frais pour la période comprise entre l'inventaire et la fin du contrat (pour les coûts qui ne sont pas couverts par la PPNA) ;
- La provision pour sinistre à payer (PSAP) qui correspond à une estimation des dépenses nécessaires pour régler tous les sinistres survenus et non payés. Elle est définie comme « la valeur estimative des dépenses en principal et en frais, tant internes qu'externes, nécessaires au règlement de tous les sinistres survenus et non payés, y compris les capitaux constitutifs des rentes non encore mises à la charge de l'entreprise » (4° Art. R343-7 du Code des assurances) ;
- La provision pour risques croissants pouvant être exigée pour les opérations d'assurance contre les risques de maladie et d'invalidité et qui est égale à la

différence entre les valeurs actuelles des engagements de l'assureur et de l'assuré ;

- La provision pour risque d'exigibilité qui permet de faire face aux engagements dans le cas de moins-value des actifs ;
- La provision pour égalisation qui permet notamment de faire face aux charges exceptionnelles sur les risques spécifiques (risque atomique, risque spatial, terrorisme...).

Pour calculer ces provisions, différentes normes sont utilisées en fonction du type de l'entreprise d'assurance (cotée ou non, opérant en France et/ou dans l'Union Européenne ...).

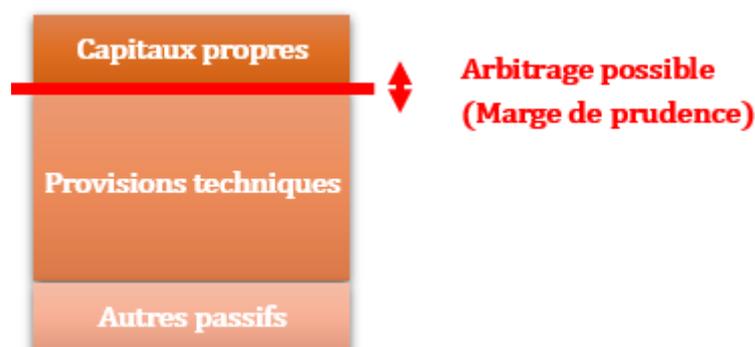
1.2.1.1 Les normes françaises

Toutes les sociétés d'assurance en France doivent répondre aux exigences des normes françaises.

Selon l'Autorité des Normes Comptables (ANC), les provisions techniques des entreprises d'assurance doivent être « *suffisantes pour le règlement intégral de leurs engagements vis-à-vis des assurés, des souscripteurs et bénéficiaires de contrats et des entreprises réassurées* » (Article 141-1).

Le caractère suffisant des provisions autorise une marge de prudence dans l'estimation des provisions mais n'impose aucun calcul précis. La politique de provisionnement influe donc très fortement sur le bilan et plus précisément sur le niveau des capitaux propres. En effet, une politique prudente va augmenter la proportion des provisions techniques et donc diminuer le montant des fonds propres et à l'inverse, une réduction du niveau des provisions augmente le résultat et donc augmente les capitaux propres.

Le passif du bilan peut être schématisé comme :



Il y a une nécessité absolue d'avoir des provisions suffisantes pour indemniser les assurés mais le résultat (sur la reprise de provision) est également un élément important notamment pour :

- Les actionnaires (via la rémunération des investissements),
- Les salariés (via les dispositifs d'intéressement et de participation),
- L'Etat (via le paiement des impôts).

Pour garantir la fiabilité du calcul des provisions et notamment de la marge de prudence considérée, les commissaires aux comptes ainsi que l'ACPR s'assurent que les engagements de l'assureur sont correctement provisionnés et en conformité avec la réglementation.

Depuis le 1^{er} janvier 2016, les dispositions relatives à l'évaluation et la comptabilisation des provisions techniques dans les comptes statutaires sont regroupées dans le Règlement ANC N°2015-11.

1.2.1.2 Solvabilité 2

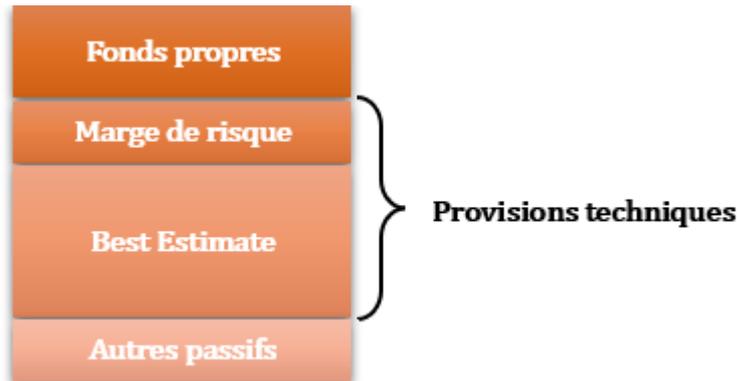
Entrée en application le 1^{er} janvier 2016, la norme Solvabilité 2 est un ensemble de règles fixant le régime de solvabilité applicable aux entreprises d'assurances dans l'Union Européenne. L'avantage est de pouvoir comparer la solidité financière de deux entreprises européennes de nationalité différente, ce qui n'était pas possible auparavant du fait des différences dans les normes nationales.

Dans ce référentiel, selon l'article 77 de la directive 2009/138/CE, les provisions techniques correspondent à « *la somme de la meilleure estimation (Best Estimate ou BE) et de la marge de risque (Risk Margin ou RM)* », ces deux quantités étant calculées séparément.

Le Best Estimate correspond à « *la moyenne pondérée par leur probabilité des flux de trésorerie futurs, compte tenu de la valeur temporelle de l'argent (valeur actuelle attendue des flux de trésorerie futurs), estimée sur la base de la courbe des taux sans risque pertinents* ». Inclure une marge de prudence est donc ici contraire à la directive.

La marge de risque est « *calculée de manière à garantir que la valeur des provisions techniques est équivalente au montant que les entreprises d'assurance et de réassurance demanderaient pour reprendre et honorer les engagements d'assurance et de réassurance* ». Les modalités de détermination de la marge de risque sont encadrées de façon très précise (Art. 37 à 39 du Règlement Délégué (UE) 2015/35, Notice « Solvabilité II » Provisions Techniques de l'ACPR). En particulier, elle est calculée selon une approche dite « Coût du Capital » (différente de l'approche par quantile à laquelle nous nous intéressons dans la suite). Elle est interprétée comme le coût actualisé de portage du risque (SCR) sur l'horizon de liquidation des engagements.

Le passif du bilan peut être représenté comme suit :



1.2.1.3 Les normes IFRS

Les normes IFRS (International Financial Reporting Standards) concernent les comptes consolidés des entreprises d'assurance cotées ou émettant de la dette sur un marché public.

La norme IFRS 17 publiée en mai 2017 et amendée en juin 2020 va entrer en application en janvier 2023. Conçue pour uniformiser les états financiers et permettre ainsi aux investisseurs de comparer plus facilement les différentes entités internationales, elle concerne environ 450 sociétés d'assurance. La principale différence avec la norme IFRS 4, qui était jusqu'à présent appliquée, est la comptabilisation du passif. Sous IFRS 4 les montants comptabilisés correspondaient aux montants enregistrés dans les comptes statutaires. Sous IFRS 17, la comptabilisation se fait suivant une vision économique. Cela permet de donner une cohérence avec l'actif qui était déjà comptabilisé en valeur de marché dans IFRS 4.

Il existe trois modèles d'évaluation des éléments du passif :

- Un modèle général (BBA) applicable à tous les contrats,
- Un modèle pour les contrats ayant une durée courte (inférieure à 1 an) ou ceux dont l'estimation par ce modèle n'est pas fondamentalement différente du modèle classique (PAA) et,
- Un modèle pour les contrats à participation directe (VFA).

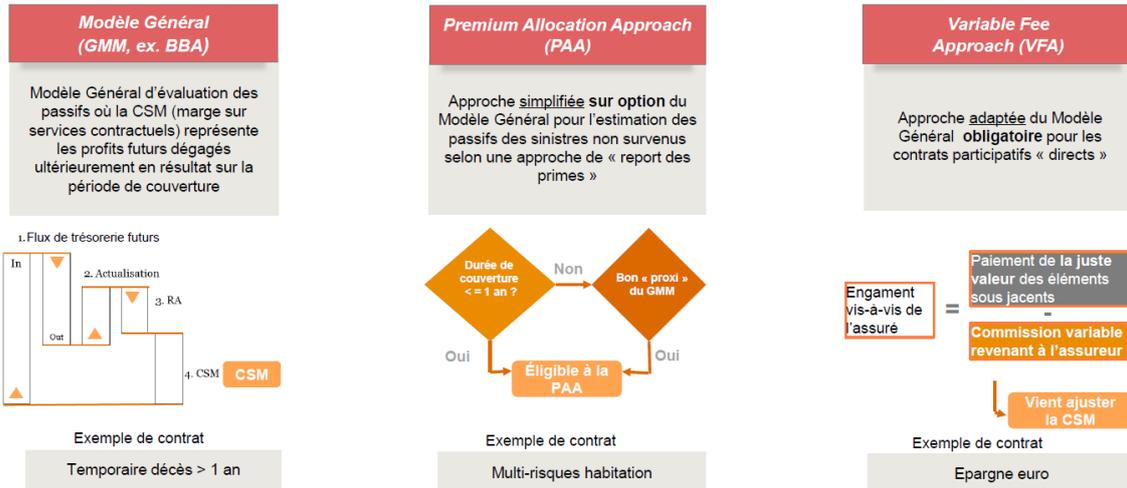


Figure 1.4 : Les modèles comptables IFRS 17 (Source : PwC)

Selon le modèle général, pour évaluer les passifs des contrats d'assurance, il y a 3 composantes :

- La valeur actuelle des flux de trésorerie futurs probabilisés (PVFCF) qui se compose d'une estimation des flux futurs et de la prise en compte de la valeur temps de l'argent via un taux d'actualisation,
- L'ajustement pour risque (RA) et
- La marge de service contractuelle (CSM).

Ces composantes sont évoquées dans le paragraphe 32 de la norme IFRS 17 :

“On initial recognition, an entity shall measure a group of insurance contracts at the total of:

(a) the fulfilment cash flows, which comprise:

(i) estimates of future cash flows (paragraphs 33–35);

(ii) an adjustment to reflect the time value of money and the financial risks related to the future cash flows, to the extent that the financial risks are not included in the estimates of the future cash flows (paragraph 36); and

(iii) a risk adjustment for non-financial risk (paragraph 37).

(b) the contractual service margin, measured applying paragraphs 38–39”

Chaque composante est ensuite définie. Pour la PVFCF, l'estimation des flux futurs est définie dans le paragraphe 33.

“ The estimates of future cash flows shall:

(a) incorporate, in an unbiased way, all reasonable and supportable information available without undue cost or effort about the amount, timing and uncertainty of those future cash flows. To do this, an entity shall estimate the expected value (i.e., the probability-weighted mean) of the full range of possible outcomes.

- (b) reflect the perspective of the entity, provided that the estimates of any relevant market variables are consistent with observable market prices for those variables.*
- (c) be current—the estimates shall reflect conditions existing at the measurement date, including assumptions at that date about the future.*
- (d) be explicit—the entity shall estimate the adjustment for nonfinancial risk separately from the other estimates. The entity also shall estimate the cash flows separately from the adjustment for the time value of money and financial risk, unless the most appropriate measurement technique combines these estimates.”*

L'actualisation est quant à elle définie dans le paragraphe 36.

- “An entity shall adjust the estimates of future cash flows to reflect the time value of money and the financial risks related to those cash flows, to the extent that the financial risks are not included in the estimates of cash flows. The discount rates applied to the estimates of the future cash flows described in paragraph 33 shall:*
- (a) reflect the time value of money, the characteristics of the cash flows and the liquidity characteristics of the insurance contracts;*
 - (b) be consistent with observable current market prices (if any) for financial instruments with cash flows whose characteristics are consistent with those of the insurance contracts, in terms of, for example, timing, currency and liquidity; and*
 - (c) exclude the effect of factors that influence such observable market prices but do not affect the future cash flows of the insurance contracts.”*

D'après ces définitions, la PVFCF correspond à la meilleure estimation des engagements de l'assureur envers l'assuré. Elle prend notamment en compte les montants actualisés des primes et des sinistres (survenus ou restant à venir) mais aussi tous les coûts attribuables au contrat (acquisition, frais de gestion...) à partir du moment où les flux sont dans les limites des contrats. Pour l'actualisation, aucune méthode n'est prescrite mais la méthode choisie doit refléter les caractéristiques des flux de trésorerie du passif. Cette définition se rapproche de celle du Best Estimate proposée par Solvabilité 2 mais les résultats peuvent être éloignés du fait de la différence au niveau de l'agrégation des contrats (par Line of Business pour Solvabilité 2 et par Unit of Account pour IFRS 17) ou de la courbe utilisée pour l'actualisation.

Concernant l'ajustement pour risque, sa définition se trouve dans le paragraphe 37 de la norme :

“An entity shall adjust the estimate of the present value of the future cash flows to reflect the compensation that the entity requires for bearing the uncertainty about the amount and timing of the cash flows that arises from non-financial risk.”

Il reflète la compensation que l'assureur requiert pour supporter l'incertitude des flux de trésorerie futurs résultants de risques non financiers, tant sur les montants que les

dates d'occurrence. Il est donc directement lié à l'appétit au risque et au degré de diversification des assureurs (paragraphe B88).

Le paragraphe B86 précise que seuls les risques assurantiels découlant des contrats d'assurance comptabilisés rentrent dans le périmètre de l'ajustement pour risque :

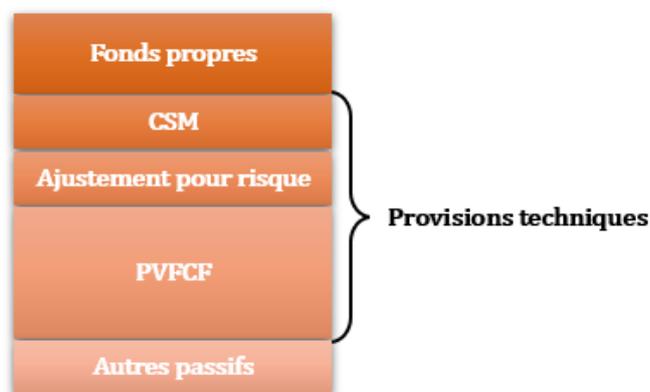
"The purpose of the risk adjustment for non-financial risk is to measure the effect of uncertainty in the cash flows that arise from insurance contracts, other than uncertainty arising from financial risk. Consequently, the risk adjustment for non-financial risk shall reflect all non-financial risks associated with the insurance contracts. It shall not reflect the risks that do not arise from the insurance contracts, such as general operational risk."

La marge de service contractuelle est définie dans le paragraphe 38 :

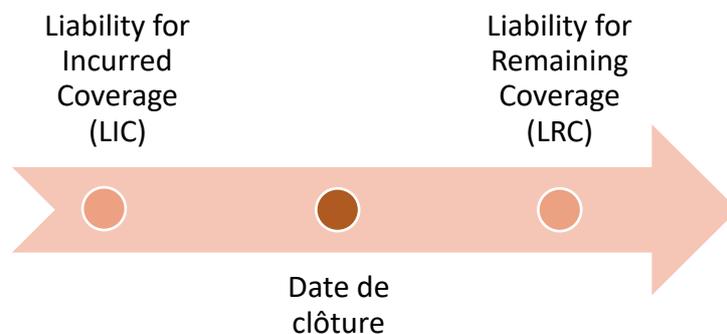
"The contractual service margin is a component of the asset or liability for the group of insurance contracts that represents the unearned profit the entity will recognize as it provides insurance contract services in the future."

Elle représente les profits futurs attendus du contrat et est équivalente aux primes diminuées de la PVFCF et du RA. Elle permet de lisser le résultat sur toute la durée du contrat. Elle est notamment impactée par les changements d'hypothèses non financières.

Le passif du bilan peut être représenté comme suit :



Dans le référentiel IFRS 17, les provisions se décomposent en deux parties :



Le LIC fait référence aux évènements passés (par exemple les sinistres survenus mais non clôturés) et le LRC aux évènements futurs (par exemple les sinistres futurs rattachés aux périodes de couverture pour lesquelles l'entité d'assurance est déjà engagée).

Dans ce mémoire, nous nous plaçons dans le cadre de l'assurance non-vie. La majorité des acteurs du marché va retenir le modèle simplifié (PAA) pour la comptabilisation des portefeuilles non-vie. Le fonctionnement de ce modèle et sa comparaison avec le modèle général ainsi qu'avec IFRS 4 sont résumés dans le schéma ci-dessous.

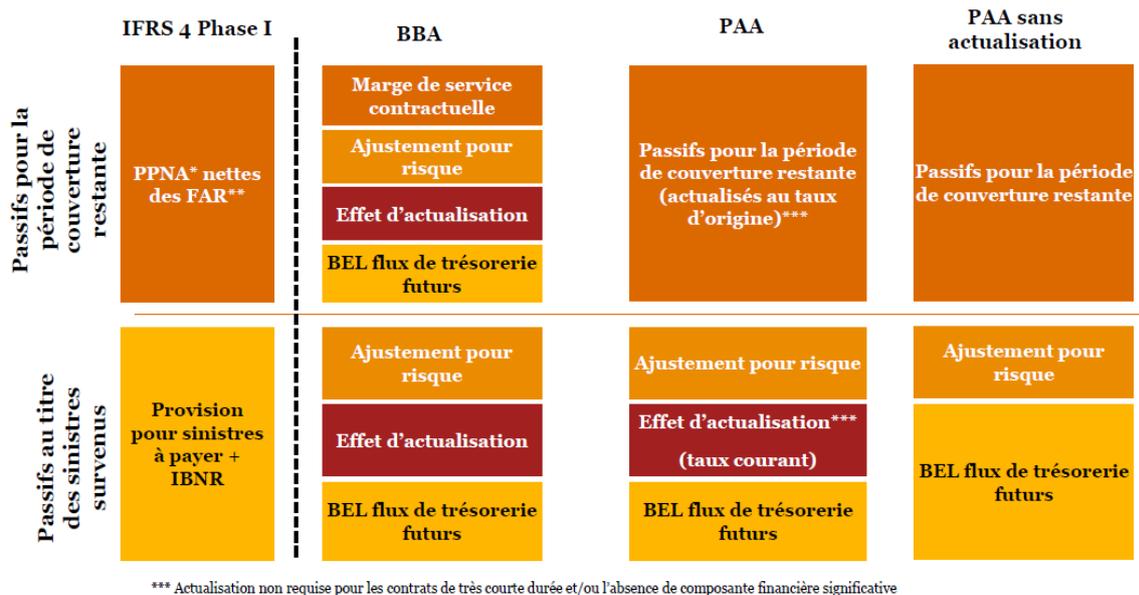


Figure 1.5 : Le modèle PAA - Comparatif avec l'approche actuelle et le modèle général (Source : PwC)

Dans ce modèle, le LRC se calcule comme les Primes Emises Non Acquises (PENA) diminuées des frais d'acquisition. Il suffira donc que les entités d'assurance estiment la part des primes et frais relatifs aux prochains exercices.

De plus, ce modèle simplifié permet également aux entités qui l'utilisent de ne pas calculer de CSM pour le LIC. Le LIC est donc égal à la somme d'un « Best Estimate » (BE LIC) et de l'ajustement pour risque (RA LIC).

Ainsi, dans le cadre du modèle PAA, le passif vaut :

$$Passif_{PAA} = PPNA_{LRC} - frais_{LRC} + BE_{LIC} + RA_{LIC}$$

L'évaluation va donc être centrée sur l'estimation du BE et du RA où le BE correspond à la PSAP en vision économique et le RA correspond au quantile de niveau α de la distribution des PSAP économiques.

1.2.2 La PSAP et le cycle de vie des sinistres

Le calcul de la PSAP est directement lié à la position d'un sinistre au sein de son cycle de vie puisque cette provision ne concerne que les sinistres survenus et non payés.

Il existe 3 catégories de sinistres en fonction de leur position dans le cycle :

- Les sinistres clos survenus et déclarés pendant l'exercice comptable qui ont entièrement été payés à la date de clôture. Ce type de sinistres ne nécessite pas de provision puisque l'assureur n'a plus d'engagement envers l'assuré.

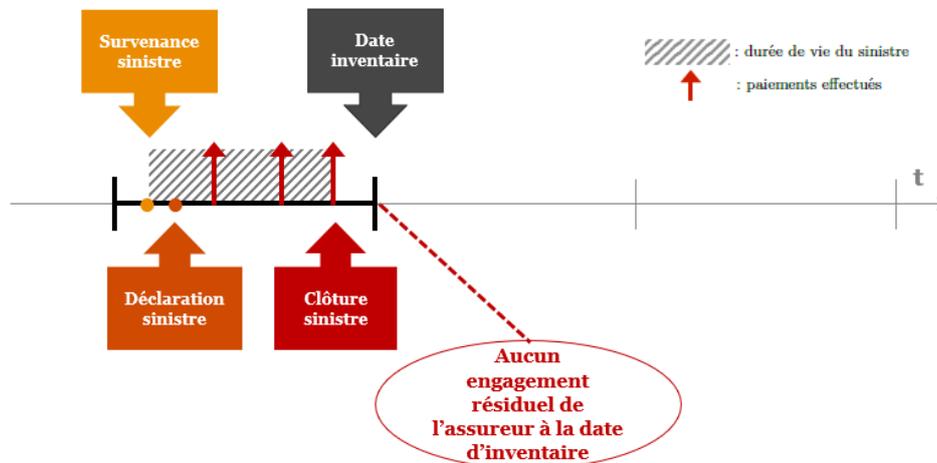


Figure 1.6 : Représentation de la vie d'un sinistre - Cas d'un sinistre survenu, déclaré et clôturé

- Les sinistres en cours qui ont été déclarés à l'assureur durant l'exercice mais dont le montant final n'est pas connu. Ce type de sinistre nécessite des provisions à la date d'inventaire car l'assureur a toujours un engagement envers l'assuré et il doit indemniser le sinistre dans sa totalité mais ne connaît pas le coût final du sinistre.

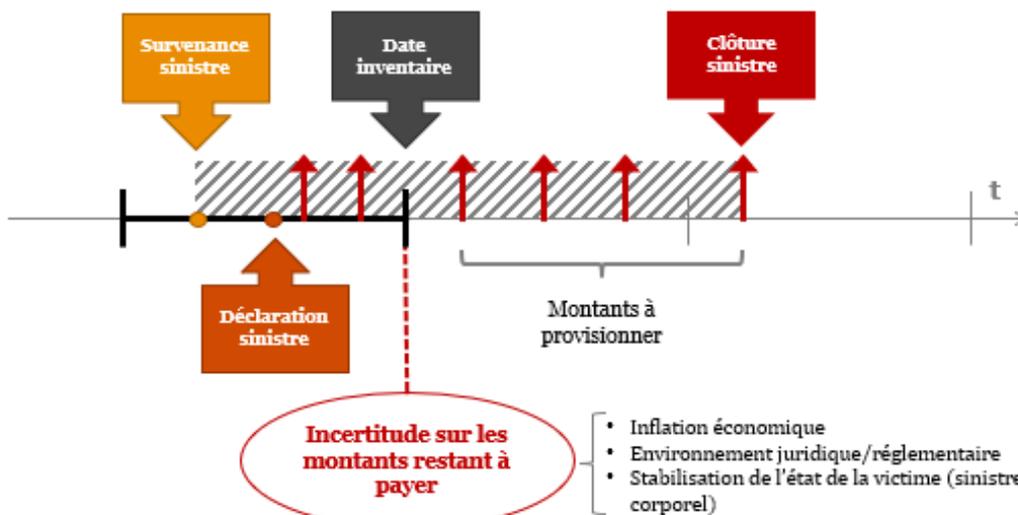


Figure 1.7 : Représentation de la vie d'un sinistre - Cas d'un sinistre survenu, déclaré et mais pas clôturé

- Les sinistres tardifs qui n'ont pas encore été déclarés à l'assureur bien qu'ils soient survenus au cours de la période de couverture. Ce type de sinistre nécessite également d'être provisionné puisque l'assureur doit encore indemniser le sinistre dans sa totalité même si son coût final est inconnu.

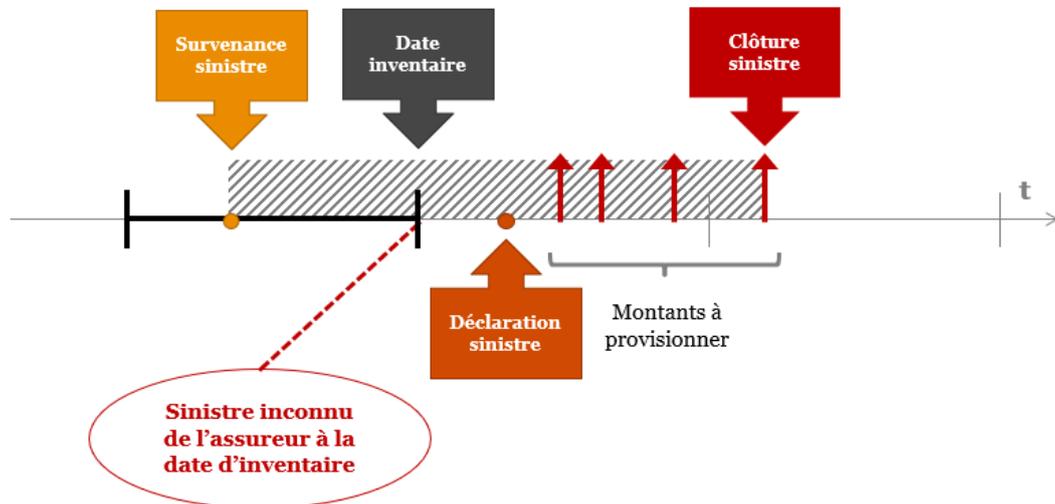


Figure 1.8 : Représentation de la vie d'un sinistre - Cas d'un sinistre survenu, non déclaré et non clôturé

La PSAP se compose de 3 montants :

- Le montant des provisions initialement calculé par les gestionnaires sinistres à la survenance du sinistre à partir des informations du dossier (provision dossier/dossier) ;
- Le montant des IBNER (Incurred But Not Enough Reported) sur les sinistres déclarés qui vient diminuer ou augmenter la provision dossier/dossier à la suite de nouvelles informations (passage d'un expert, changement d'environnement économique, juridique ou réglementaire...) ;
- Le montant des IBNYR (Incurred But Not Yet Reported) qui correspond aux sinistres tardifs est évalué à partir de méthodes statistiques.

Le coût ultime peut donc se décomposer comme suit :

Cout Ultime	PSAP	IBNR	IBNYR
		Provision d/d	IBNER
	Montant payé	Montant payé	Montant total encouru

Figure 1.9 : Décomposition d'un sinistre

1.3 Notations et triangles de provisionnement

Les modèles utilisés pour le provisionnement se basent sur des triangles de liquidation semblables à celui-ci-dessous :

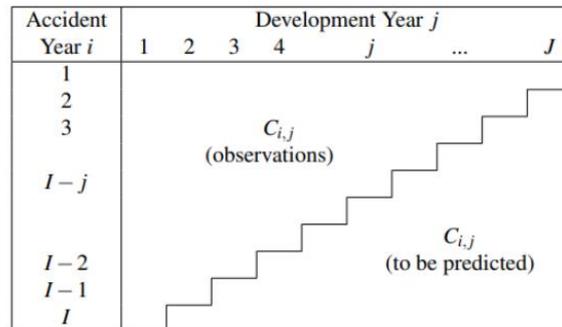


Figure 1.10 : Représentation d'un triangle de liquidation

Les lignes représentant les années de survenance des sinistres sont notées $i : 1 \leq i \leq I$ et les colonnes représentant les délais de règlement des sinistres sont notées $j : 1 \leq j \leq J$. Pour la suite et par soucis de simplification, nous allons prendre $I = J = n$. Les diagonales représentent les années calendaires.

Dans ce mémoire, nous utiliserons des développements en années mais il est aussi possible d'étudier les développements au trimestre ou au mois (à condition de disposer des données selon la périodicité requise). Les montants observés peuvent être incrémentaux ou cumulés et les triangles peuvent représenter les charges, les paiements ou le nombre de sinistre. Dans la suite, nous allons noter :

- $N_{i,j}$ le nombre cumulé de sinistres survenus l'année i regardés l'année j ,
- $S_{i,j}$ les montants incrémentaux de l'année i réglés en j ,
- $C_{i,j}$ les montants cumulés, définis comme la somme des montant incrémentaux

$$C_{i,j} = \sum_{k=1}^j S_{i,k}.$$

L'objectif est de prédire la partie basse du triangle à partir des données observées dans la partie haute. Compléter le triangle va permettre d'estimer la charge sinistre ultime notée $U_i = \hat{C}_{i,n}$ ainsi que le montant des provisions pour chaque année noté \hat{R}_i . La provision globale est définie comme la somme des provisions de chaque année :

$$\hat{R} = \sum_{i=1}^n \hat{R}_i.$$

Dans la majorité des cas, les méthodes statistiques de provisionnement utilisent les données historiques de sinistralité et supposent que le passé est indicatif du futur. Ainsi, lors de l'étude d'un triangle de charges, les montants antérieurs vont permettre d'estimer les charges futures et ainsi les provisions. Ces méthodes sont d'autant plus performantes que le passé et le futur sont stables.

2. Présentation des données

Tout au long de ce mémoire, nous étudierons des triangles contenant des informations sur des sinistres automobiles.

L'assurance automobile représente environ 40% de l'ensemble des primes collectées au sein des assurances de biens et responsabilités. La branche que nous allons étudier est celle de la « responsabilité civile des véhicules terrestres automoteurs » qui couvre les dommages causés par l'assuré à un tiers (dommages matériels ou corporels). Il existe 2 types de sinistres :

- Les sinistres matériels (dommages causés à un véhicule tiers ou à des équipements urbains par exemple) ;
- Les sinistres corporels (indemnisation des piétons, cyclistes ou passagers consécutive à un accident responsable par exemple).

Ces deux types de sinistres représentent 50% de la charge sinistre totale de l'assurance automobile bien qu'ils ne soient pas majoritaires en fréquence (2% seulement).

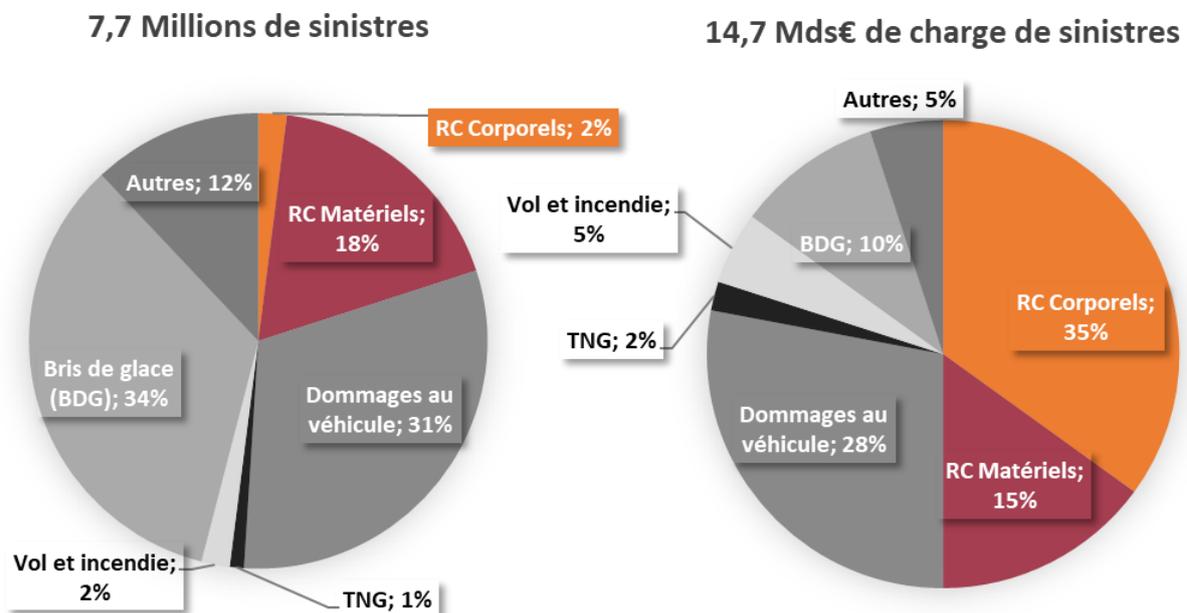


Figure 2.1 : Chiffres clés de l'assurance automobile

Ces deux types de sinistres n'ont pas du tout le même développement puisque les sinistres matériels ont un développement court à intermédiaire (moins de 5 ans en général) alors que les sinistres corporels ont un développement long (parfois plus de 20 ans).

Cette différence s'explique par le fait que le coût final d'un sinistre corporel (qui dépend de l'état de santé, de l'âge, de la profession de la victime...) est long à évaluer. En effet, dans le cas d'un sinistre impliquant un dommage corporel avec séquelles à vie, l'assureur dédommage dans un premier temps uniquement les soins médicaux (médicaments, hospitalisations...). Ce n'est qu'après la consolidation (certificat médical d'un retour à un état de santé stable avec d'éventuelles séquelles qui ne sont plus supposées évoluer dans le temps), que l'assureur fera l'offre d'indemnisation correspondant à l'évaluation des préjudices (nécessité d'une tierce personne à vie, perte de revenus...). Les montants de sinistres peuvent ainsi varier pendant plus de 20 ans, avec parfois beaucoup de volatilité surtout en cas d'aggravation des séquelles.

Nous disposons pour notre étude de deux bases de données réelles de responsabilité civile automobile avec les informations sur le montant des charges et des paiements, ainsi que des informations partielles sur le nombre de sinistres et les primes. Les montants sont nets de recours (ce qui peut entraîner la présence d'incrément négatifs). Pour des raisons de confidentialité, les deux jeux de données réelles seront présentés en base 100.

Nous avons également intégralement construit une base de sinistres pour laquelle nous connaissons toutes les informations afin de pouvoir tester nos méthodes.

Les données sont présentées en [Annexe 1](#).

2.1.1 Présentation des jeux de données réelles

Notre premier jeu de données est composé uniquement de sinistres en affaires directes sur des polices françaises survenus entre 1996 et 2021. Afin d'avoir des groupes homogènes, les données ont été séparées par tranche de coût pour différencier les sinistres attritionnels (notés dans la suite « Attritional »), les sinistres graves (notés « Large ») et les sinistres exceptionnels. Ces différents types de sinistres n'ont pas les mêmes caractéristiques et il est nécessaire de les séparer pour prendre en compte les spécificités de chacune de ces classes. En effet, les sinistres attritionnels (qui auront tendance à être beaucoup plus nombreux mais avec une sévérité relativement faible) n'auront pas forcément le même comportement que les sinistres graves (qui ont une grande sévérité et une plus faible fréquence). Les sinistres étudiés sont ici déjà regroupés en classe homogène (Attritionnels, Graves et Exceptionnels) mais si cela n'avait pas été le cas, un retraitement aurait été nécessaire.

Il faut néanmoins faire attention à ne pas faire trop de classes pour toujours être en mesure d'appliquer la loi des grands nombres. Dans notre cas, 70% des sinistres sont dans la classe « Attritional », 20% dans la classe « Large » et 10% dans la classe « Exceptionnelle ».

Le deuxième jeu de données concerne des sinistres en acceptation de réassurance sur des polices anglaises survenus entre 2002 et 2021. Au niveau de la réglementation nationale, il existe des différences notamment sur le provisionnement des rentes. Au Royaume-Uni, il existe un taux légal pour l'actualisation des rentes (le taux OGDEN) alors qu'il n'existe aucune réglementation en France. Récemment, ce taux a changé deux fois, fin 2017 et mi 2019. Il est passé de 2.5% en 2001 (période où les taux d'intérêts étaient hauts) à -0.75% en 2017. Cela a conduit à une forte augmentation des engagements sur les rentes. En effet, plus le taux est faible, plus le niveau des provisions à constituer est élevé car les indemnités exigibles augmentent. En 2019, le taux OGDEN est remonté à -0.25%.

Si les données ne sont pas retraitées, il y aura donc un effet diagonal important dans le triangle des charges. Or, le but est de projeter l'évolution des sinistres et non pas l'évolution du taux OGDEN. Nous verrons comment prendre en compte ces informations dans la suite du mémoire.

2.1.2 Base simulée

En plus des données réelles précédentes, des données ont été simulées à partir du package « SynthETIC »¹ disponible sur le logiciel R. Créé en 2021 par Avanzi, Taylor, Wang et Wong, ce package nous permet de simuler un historique de données d'assurance (à partir d'hypothèses de distribution) cohérent avec l'expérience d'un portefeuille de responsabilité civile automobile anonyme. La simulation nécessite de modéliser 8 modules :

1. La fréquence des sinistres et la détermination de leur période de survenance,
2. Le montant total des sinistres sans inflation,
3. Le délai de déclaration du sinistre (temps entre la survenance et la notification du sinistre à l'assureur),
4. Le délai de traitement (temps entre la déclaration du sinistre à l'assureur et la clôture),
5. Le nombre de paiement,
6. Le montant des paiements partiels,
7. Le délai entre le paiement des montants partiels,
8. L'inflation.

Pour chaque module, une hypothèse de loi est faite et les paramètres sont choisis. Pour les hypothèses de lois, les lois les plus utilisées en pratique ont été choisies. 20 années de sinistres ont été simulées et nous allons présenter dans la suite les sélections réalisées pour simuler la base.

¹ <https://cran.r-project.org/package=SynthETIC>

Module 1

Nous supposons que la fréquence suit une loi de Poisson. Le paramètre est estimé à partir des données du portefeuille de responsabilité civile automobile anonyme. La fréquence moyenne attendue est d'environ 90 sinistres par an.

La date de survenance de ces sinistres est un tirage d'une loi uniforme sur l'année où ils surviennent. Voici le nombre de sinistres simulé par année :

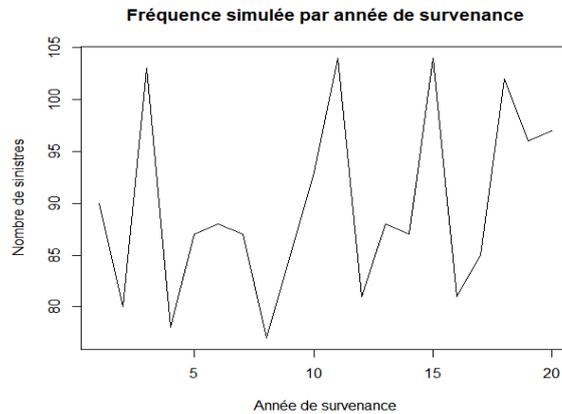


Figure 2.2 : Représentation graphique de la fréquence des sinistres

Module 2

Pour le montant des sinistres, nous allons utiliser une loi de Weibull. En effet, les distributions à valeurs positives ont un schéma similaire. Elles commencent avec une fréquence d'apparition nulle, croissent jusqu'à un maximum puis décroissent. Il est alors possible de trouver une loi de Weibull qui ne s'éloigne pas trop des données du package en calculant les paramètres à partir de la moyenne et de la variance observée.

Sa densité est :

$$f(x; \alpha, \beta) = \frac{\alpha}{\beta} \left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^\alpha}$$

Le paramètre α est le paramètre de forme et β celui d'échelle. Dans notre cas, le paramètre de forme vaut environ 0.7 et celui d'échelle environ 120 000.

Voici les montants simulés :

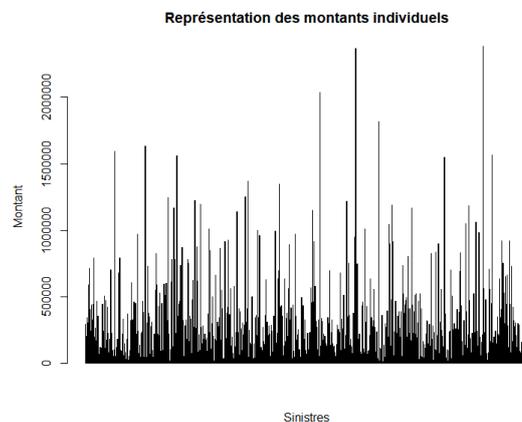


Figure 2.3 : Représentation graphique des montants individuels

Module 3

La durée avant la déclaration du sinistre à l'assureur suit aussi une loi de Weibull. Les paramètres estimés sont 1.3 pour la forme et 2.3 pour l'échelle.

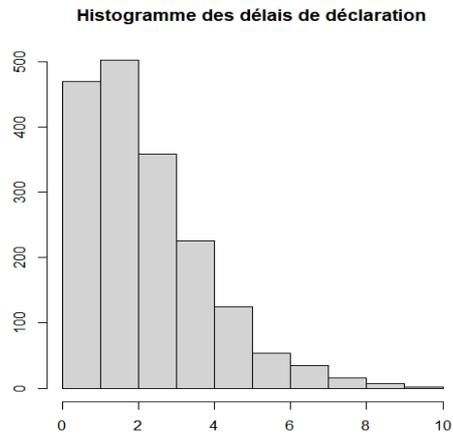


Figure 2.4 : Histogramme des délais de déclaration des sinistres

La majorité des sinistres sont déclarés en moins de 2 ans et seul une minorité de sinistres sont déclarés très tardivement.

Module 4

La durée avant la clôture du sinistre suit également une loi de Weibull avec comme paramètres 2.47 pour la forme et 10.7 pour l'échelle.

Sur l'histogramme suivant, nous pouvons remarquer que dans la majorité des cas, il faut environ 6 ans en moyenne pour qu'un sinistre soit clôturé.

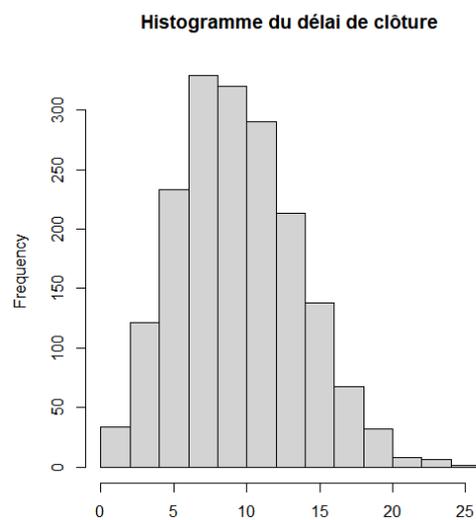


Figure 2.5 : Histogramme des délais de clôture

Module 5

Pour le nombre de paiements, nous allons utiliser une loi qui dépend du montant estimé du sinistre. Ainsi, les sinistres qui ont un montant faible auront moins de paiements que les sinistres avec un montant élevé.

Voici un tableau récapitulatif des montants simulés :

Min	Q1	Q2	Q3	Max
0	20 000	67 000	192 000	2 400 000

Nous considérons qu'en dessous du premier quartile, le nombre de paiement vaut 1 ou 2 avec équiprobabilité. Dans le cas où le montant est compris entre le premier et le troisième quartile, le nombre de paiement vaut 2 dans un cas sur trois et vaut 3 sinon. Enfin, si le montant est supérieur au troisième quartile, le nombre de paiement suit une loi géométrique de paramètre 1/2 translattée de 4.

Voici une représentation graphique du nombre de paiements en fonction du montant :

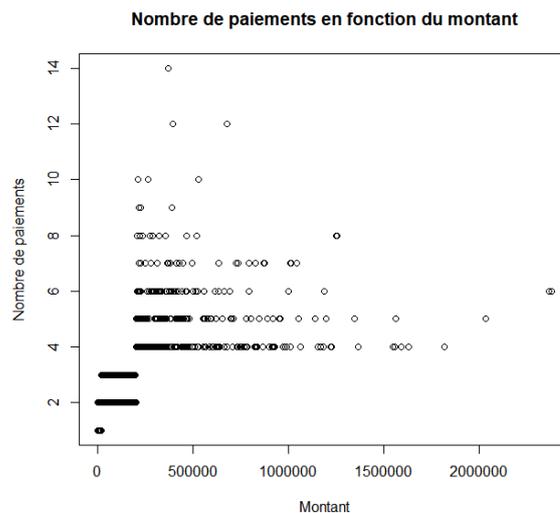


Figure 2.6 : Représentation graphique du nombre de paiements

Module 6

Une fois le nombre de paiements simulé, il faut répartir la charge totale entre ces différents paiements. Il serait possible de ne choisir qu'une seule loi pour modéliser le montant des paiements mais nous avons choisi de définir la distribution différemment en fonction du nombre de paiements pour être au plus proche de la réalité. Ainsi :

- Pour un unique paiement, le montant partiel est équivalent au montant total.

- Si le nombre de paiements vaut 2 ou 3, alors nous allons utiliser une loi bêta de densité :

$$f(x; \alpha, \beta) = \begin{cases} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1} du} & \text{pour } x \in [0; 1] \\ 0 & \text{sinon} \end{cases}$$

Les paramètres utilisés pour tirer des proportions correspondant aux paiements partiels garantissent que l'espérance soit de $1/\text{nombre_paiements}$ et que la variance soit de 0.10.

Par exemple, dans le cas de deux paiements, nous avons obtenu comme tirage 0.61 et 0.54. Nous avons ensuite renormalisé ces tirages : $\frac{0.61}{0.61+0.54} = 53\%$ et $\frac{0.54}{0.61+0.54} = 47\%$. Dans notre exemple, 53% du montant final sera payé au premier versement et le reste sera versé au second.

- Dans le cas où le nombre de paiements est supérieur à 4, il y a une distinction à faire entre les premiers paiements et les deux derniers supposés beaucoup plus élevé. En effet, si le nombre de paiements est important, cela signifie généralement que le montant total est élevé et que des procédures sont en cours. Les paiements les plus importants seront donc à la fin de la procédure. En assurance automobile, il est par exemple courant d'attendre la consolidation avant de faire une offre d'indemnisation. Les premiers versements correspondent donc uniquement aux frais médicaux tandis que les deux derniers correspondent à l'indemnité principale.

Nous commençons par simuler la part que les deux derniers paiements vont représenter en utilisant la quantité $1 - \text{Beta}(\alpha, \beta)$ avec α et β tels que l'espérance de la loi vaut 0.05 et la variance 0.2. Nous avons choisi de considérer que le montant le plus important serait l'avant-dernier, le dernier représentant les derniers frais résiduels. Nous allons donc attribuer des poids différents en tirant à nouveau une loi $\text{Beta}(\alpha, \beta)$ garantissant une espérance de 0.9 et une variance de 0.03.

Pour les autres paiements, nous allons utiliser la loi Beta définie dans le point précédent.

Par exemple, dans le cas de 6 paiements, nous allons commencer par déterminer la part du montant total attribuée aux deux derniers paiements. Nous tirons 95%. Ensuite, entre ces deux paiements, nous souhaitons accorder un poids différent et le tirage nous donne 91% pour l'avant-dernier et 9% pour le dernier. Pour les 4 premiers paiements, nous répétons la procédure du 2^{ème} point et obtenons les proportions 0.23, 0.26, 0.25 et 0.23.

Pour obtenir des pourcentages du montant total, nous renormalisons :

$$\begin{aligned} \text{Premier paiement} &: (1 - 0.95) \times \frac{0.23}{0.23 + 0.26 + 0.25 + 0.23} = 1.18\% \\ \text{Deuxième paiement} &: (1 - 0.95) \times \frac{0.26}{0.23 + 0.26 + 0.25 + 0.23} = 1.33\% \\ \text{Troisième paiement} &: (1 - 0.95) \times \frac{0.25}{0.23 + 0.26 + 0.25 + 0.23} = 1.28\% \\ \text{Quatrième paiement} &: (1 - 0.95) \times \frac{0.23}{0.23 + 0.26 + 0.25 + 0.23} = 1.18\% \\ \text{Cinquième paiement} &: (0.95) \times 0.91 = 86.90\% \\ \text{Sixième paiement} &: (0.95) \times 0.09 = 8.13\% \end{aligned}$$

Ainsi, si le montant total est de 100 000€, il sera réparti comme suit :

N° Paiement	1	2	3	4	5	6
Montant	1180	1330	1280	1180	86900	8130

Le choix de la loi beta se justifie par le fait qu'en fonction des paramètres choisis, il existe différentes asymétries sur l'intervalle $[0 ; 1]$ ce qu'une loi uniforme, par exemple, ne permet pas. Voici une illustration des lois utilisées pour répartir le montant total entre les paiements :

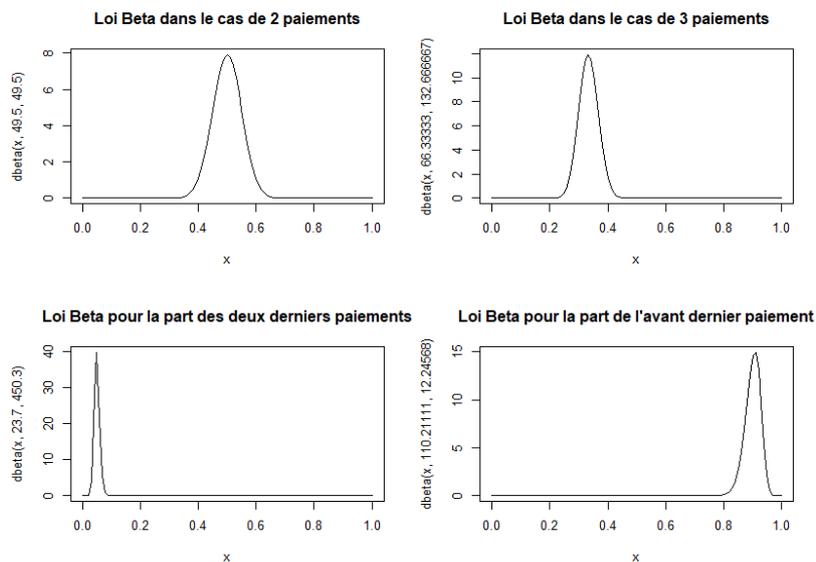


Figure 2.7 : Illustration des lois Bêta utilisées

Module 7

Le délai entre chaque paiement suit une loi de Weibull. Le paramètre de forme dépend du délai de clôture calculé au module 4 (environ 9.46) divisé par le nombre de paiements et le paramètre d'échelle dépend d'une variance fixée à 35%.

Le résultat est renormalisé de la même manière qu'au module 6 pour tenir compte de la durée maximale imposée par la date de clôture.

Module 8

Pour modéliser l'inflation, nous avons choisi une inflation constante de 2% par an (inflation passée et prédiction pour l'inflation future). Cette valeur correspond à l'inflation moyenne passée calculée à partir de l'indice des prix pour l'entretien et la réparation des véhicules particuliers de l'INSEE.

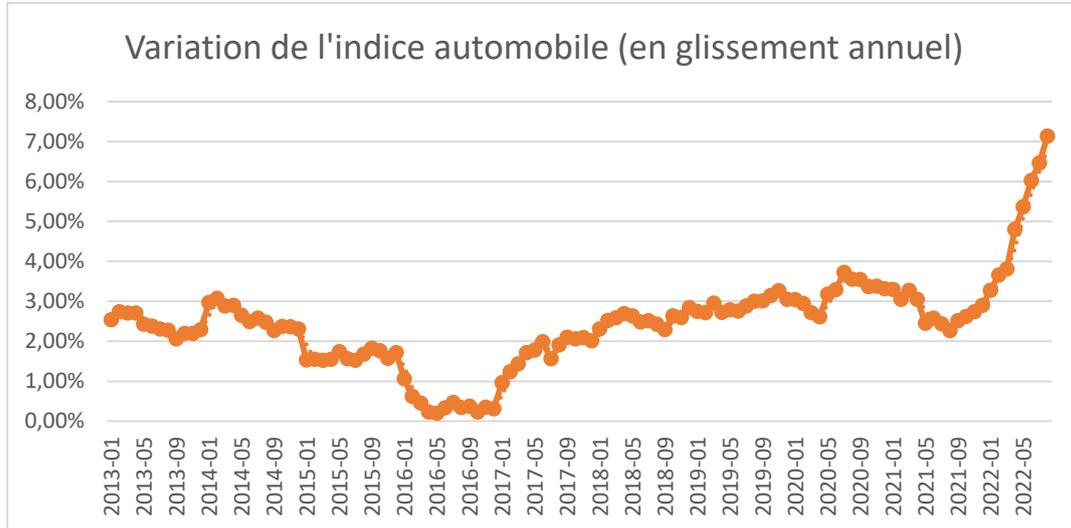


Figure 2.8 : Graphique des taux de l'IPC automobile

Nous pouvons récapituler notre démarche dans le tableau suivant :

Module	Loi	Paramètres
Fréquence des sinistres	Poisson	90
Montant des sinistres	Weibull	Shape = 0.667 ; Scale = 119350
Délai de déclaration	Weibull	Shape = 1.31 ; Scale = 2.32
Délai de traitement	Weibull	Shape = 2.47 ; Scale = 10.669
Nombre de paiements		<ul style="list-style-type: none"> • Si montant < Q1, $\mathbb{P}(\text{Nombre} = 1) = \mathbb{P}(\text{Nombre} = 2) = \frac{1}{2}$ • Si $Q1 \leq \text{montant} \leq Q3$, $\mathbb{P}(\text{Nombre} = 2) = \frac{1}{3}$ et $\mathbb{P}(\text{Nombre} = 3) = \frac{2}{3}$ • Sinon, la distribution est géométrique $G(1/2)$ translatée de 4
Montant des paiements partiels	Beta conditionnelle au nombre de paiements	<ul style="list-style-type: none"> • Nombre de paiement = 2 ou 3 $Beta(\alpha, \beta)$ d'espérance $\frac{1}{\text{nombre de paiements}}$ et de variance 0.10 <ul style="list-style-type: none"> • Paiement ≥ 4 Les 2 derniers paiements en % du montant total : $1 - Beta(\alpha, \beta)$ d'espérance 0.05 et de variance 0.2 L'avant-dernier paiement : $Beta(\alpha, \beta)$ d'espérance 0.9 et de variance 0.03 Les autres suivent le cas précédent (paiement = 2 ou 3)
Délai entre les paiements partiels	Weibull	Shape : paramètre qui permet d'obtenir une espérance égale au $\frac{\text{Délai de clôture moyen (9.46)}}{\text{Nombre de paiements}}$ Scale = 3.128794
Inflation	$f(t) = (1 + \alpha)^t$	$\alpha = 2\%$

A partir des deux bases de données réelles, nous allons chercher à modéliser la charge ultime de l'assureur et déterminer le montant de provisions adéquat ainsi que l'erreur de prédiction commise.

Avec la base simulée, nous allons évaluer la qualité des méthodes en comparant les estimations des cash flows futurs (cash flows projetés) et les cash flows réels (cash flows futurs simulés).

3. Méthodes déterministes classiques de provisionnement

Les méthodes déterministes permettent de projeter à l'ultime les sinistres afin d'évaluer les montants de provisions nécessaires. Un panorama détaillé de ces méthodes est notamment présenté dans l'ouvrage de Partrat et al. (2007).

3.1 Méthode Chain Ladder

Apparue en 1938, la méthode de Chain Ladder est aujourd'hui la méthode de référence la plus largement utilisée sur le marché. Elle permet de donner une estimation de la charge ultime uniquement à partir des données historiques.

3.1.1 Méthode classique

3.1.1.1 Présentation de la méthode

La méthode Chain Ladder est la plus utilisée en pratique du fait de sa simplicité. Elle peut être employée sur de nombreux types de triangles cumulés : règlements, nombre de sinistres, charges... Le principe est de calculer un coefficient permettant de passer d'une année de développement à une autre. Les facteurs de développement individuels sont définis par :

$$f_{i,j} = \frac{C_{i,j+1}}{C_{i,j}}$$

Pour utiliser cette méthode, il est nécessaire de vérifier une hypothèse centrale : les facteurs de développement sont indépendants de l'année de survenance. Ainsi, la première hypothèse est :

(H1)

$$\forall j \in \{1, \dots, n-1\} : \frac{C_{1,j+1}}{C_{1,j}} = \dots = \frac{C_{i,j+1}}{C_{i,j}} = \dots = \frac{C_{n,j+1}}{C_{n,j}}$$

Si le rapport de proportionnalité est identique, cela signifie que les sinistres se développent de la même manière indifféremment de leur année de survenance. Les années de développement sont les seules variables explicatives du comportement des sinistres futurs et il est donc légitime dans ces conditions d'utiliser une unique cadence de développement par colonne, que l'on notera f_j . Il est donc possible d'écrire :

$$\forall j \in \{1, \dots, n-1\} \text{ et } \forall i \in \{1, \dots, n\}, \exists f_j \text{ tel que : } E(C_{i,j+1} | C_{i,1}, \dots, C_{i,j}) = f_j C_{i,j}$$

Cette méthode suppose également l'indépendance des années de survenance, c'est-à-dire :

(H2)

$$\{C_{i,1}, \dots, C_{i,n}\} \text{ est indépendant de } \{C_{b,1}, \dots, C_{b,n}\} \text{ pour } i \neq b$$

Suite aux hypothèses, il est naturel d'estimer le facteur de développement relatif à une année de développement de la manière suivante :

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

Une fois ces facteurs calculés, il suffit de multiplier les montants cumulés par les facteurs appropriés pour compléter la partie basse du triangle. Les montants de la partie basse sont donc estimés par :

$$\forall j \geq n-i+1 : \hat{C}_{i,j}^{CL} = C_{i,n-i+1} \cdot \prod_{k=n-i+1}^{j-1} \hat{f}_k$$

Les estimations de la partie basse du triangle sont obtenues à partir de l'extrapolation de la dernière diagonale connue.

Les provisions par année de survenance (sauf dans le cas de la première année de survenance disponible où la charge de sinistres est supposée être déjà à l'ultime) sont calculées en faisant la différence entre la charge ultime estimée et la dernière information connue :

$$\hat{R}_i = \hat{C}_{i,n} - C_{i,n-i} \quad \text{et} \quad \hat{R} = \sum_{i=2}^n \hat{R}_i$$

Avant d'utiliser la méthode, nous devons en vérifier les hypothèses.

3.1.1.2 Vérification des hypothèses

Nous allons étudier la base de données française segmentée (pour chacune des classes) ainsi que les données issues du portefeuille d'acceptations anglais.

Données françaises – Classe « Attritional »

La première hypothèse peut se vérifier graphiquement en traçant les couples $(C_{i,j}; C_{i,j+1})$. Si l'hypothèse est vérifiée, pour chaque période, les points doivent être relativement bien alignés et il doit être possible de pouvoir tracer une droite qui passe par l'origine et l'ensemble de points.

Dans nos graphiques, par soucis de clarté, nous n'allons présenter que les trois premières périodes de développement car ce sont celles qui sont le plus volatiles mais le test a été réalisé sur l'ensemble des périodes de développement.

Le graphique obtenu à partir des données de notre base est présenté ci-dessous.

- La deuxième étape est la comptabilisation des -1 et des 1 par diagonale. En l'absence d'effet calendaire, il est censé y avoir autant de valeurs négatives (nb_{neg}) que de valeurs positives (nb_{pos}) sur la diagonale. En posant $Z_k = \min(nb_{neg}; nb_{pos})$, il est possible de remarquer que $Z_k \sim \text{Binomiale}(n, \frac{1}{2})$ où n est le nombre total d'éléments non nuls sur la diagonale. Les premiers moments sont donnés par :

$$E(Z_k) = \frac{n}{2} - \binom{n-1}{\lfloor \frac{n-1}{2} \rfloor} \times \frac{n}{2^n}$$

$$V(Z_k) = \frac{n(n-1)}{4} - \binom{n-1}{\lfloor \frac{n-1}{2} \rfloor} \times \frac{n(n-1)}{2^n} + E(Z_k) - E(Z_k)^2$$

- La dernière étape est la construction de l'intervalle de confiance à 95% permettant de tester notre hypothèse. Comme la somme de variables aléatoires binomiales converge vers une loi normale, en définissant $Z = \sum_k Z_k$ l'intervalle de confiance est donné par :

$$\left[E(Z) - 1.96\sqrt{V(Z)} ; E(Z) + 1.96\sqrt{V(Z)} \right]$$

Pour nos données, l'intervalle de confiance à 95% est [112 ;133] alors que la variable Z vaut 102. L'hypothèse d'absence d'effet calendaire est donc rejetée.

Toutes les hypothèses ne sont pas vérifiées, il n'est donc pas possible en l'état d'appliquer la méthode Chain Ladder sur ces sinistres.

Dans le cas où toutes les hypothèses ne sont pas vérifiées, il faut différencier les cas où :

- La première hypothèse n'est pas vérifiée car les couples $(C_{i,j}; C_{i,j+1})$ sont alignés selon une droite ne passant pas par l'origine et il faut alors changer de méthode : il est possible de rajouter un facteur pour avoir une relation affine (méthode London Chain par exemple).
- La seconde hypothèse n'est pas vérifiée et il est possible de ne sélectionner que certaines années afin de ne pas prendre en compte les effets calendaires (version pondérée).

Données françaises – Classe « Large »

Pour les données des sinistres classés « Large », les deux hypothèses sont vérifiées puisque la statistique de test Z vaut 118 et l'intervalle de confiance est [112 ; 133]. De plus, les couples $(C_{i,j} ; C_{i,j+1})$ sont alignés comme le montre le graphique ci-dessous.

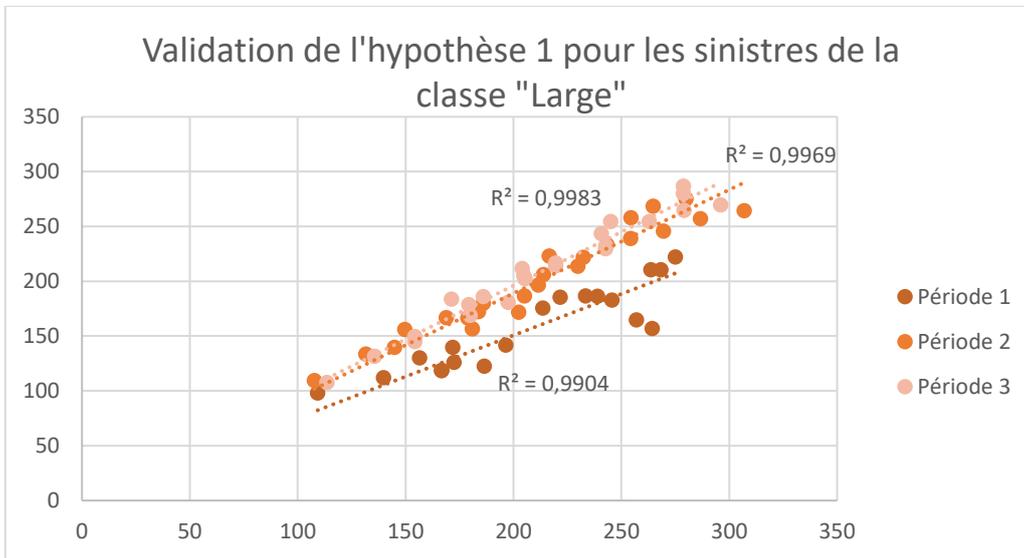


Figure 3.3 : Vérification graphique de l'hypothèse 1 pour la classe « Large »

Données françaises – Classe « Exceptionnelle »

Pour les données de cette classe, la deuxième hypothèse n'est pas vérifiée puisque la statistique qui vaut 107 est en dehors de l'intervalle de confiance calculé [112 ; 133]. La première hypothèse est néanmoins vérifiée :

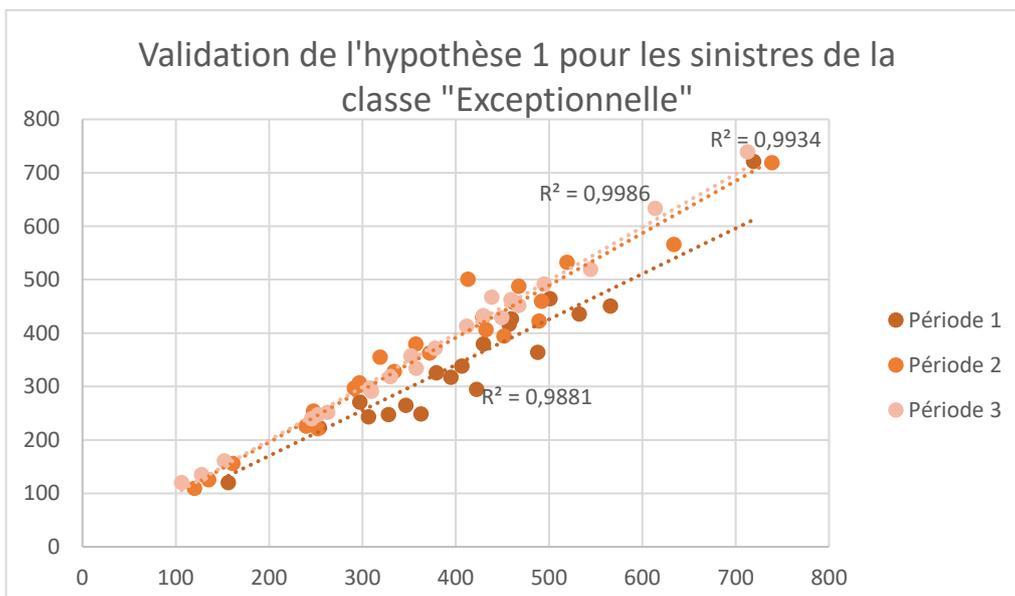


Figure 3.4 : Vérification graphique de l'hypothèse 1 pour la classe "Exceptionnelle"

Données d'acceptations anglaises

Nous allons tester de la même manière les hypothèses de Chain Ladder sur nos sinistres anglais. La première hypothèse, vérifiée grâce au graphique suivant, est validée.

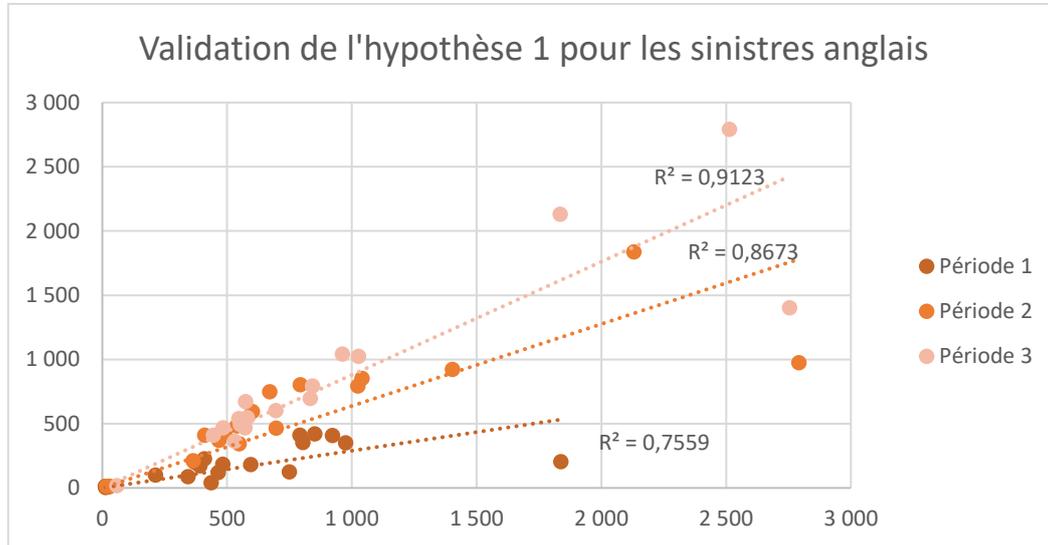


Figure 3.5 : Vérification graphique de l'hypothèse 1 pour les données anglaises

Pour la seconde hypothèse, le test d'indépendance calendaire a été réalisé. L'intervalle de confiance est [60 ;76] et notre statistique vaut 55. Cette hypothèse n'est donc pas validée et cela est dû à l'impact des variations du taux OGDEN. Pour avoir cette hypothèse vérifiée, il faudra alors supprimer cet effet.

L'ensemble des hypothèses n'est vérifié que pour les sinistres français de la classe « Large », nous n'appliquerons donc la méthode que sur ces données.

3.1.1.3 Application de la méthode

Après avoir calculé les facteurs de développement pour chaque période, nous complétons le triangle en multipliant la dernière valeur connue par le produit des facteurs de développement restants. Voici le triangle complété :

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26		
1996	100	134	132	136	135	135	133	131	129	128	126	117	116	117	117	117	117	117	117	117	117	117	117	117	119	119	120	120
1997	131	156	150	154	143	138	132	128	129	127	126	131	131	130	131	129	130	135	138	139	141	141	142	142	142	142	148	148
1998	222	258	254	245	221	209	202	203	205	204	195	197	199	197	199	200	203	202	203	204	201	210	211	215	215	220	221	221
1999	181	223	217	219	203	191	210	209	207	202	200	202	201	201	202	204	205	207	210	209	212	210	213	215	215	220	221	221
2000	150	206	214	220	223	213	207	198	202	199	198	194	193	192	192	192	196	198	203	201	202	203	205	210	211	211	211	211
2001	153	180	186	186	189	178	178	179	171	170	173	178	173	174	179	176	178	184	184	182	179	181	183	184	188	188	189	189
2002	112	140	145	154	157	160	158	157	151	160	160	165	168	169	169	177	182	182	180	178	177	179	181	182	186	186	187	187
2003	118	167	179	179	187	182	182	184	162	174	175	172	171	177	178	180	182	187	187	188	187	189	191	192	197	197	197	197
2004	98	109	108	114	115	114	108	106	105	105	106	102	102	101	100	98	104	108	108	108	108	109	110	111	114	114	114	114
2005	142	196	211	204	191	196	199	201	196	205	197	198	198	200	199	201	196	199	201	201	200	202	204	206	211	211	211	211
2006	119	167	169	180	180	178	172	169	168	170	166	163	172	167	163	166	168	170	171	172	171	173	174	176	180	180	180	180
2007	126	172	184	171	180	177	175	160	147	143	139	138	132	130	129	130	131	133	134	134	133	135	136	137	140	141	141	141
2008	185	222	232	243	259	268	267	249	240	228	226	226	225	230	231	232	234	238	239	240	239	241	243	245	251	252	252	252
2009	123	186	205	203	207	201	185	174	174	178	177	176	170	170	170	171	173	176	177	177	176	178	180	181	185	186	186	186
2010	130	156	181	197	209	214	226	225	228	213	209	218	217	218	218	220	222	225	226	227	226	228	230	232	238	238	238	238
2011	175	214	230	243	240	215	213	217	216	212	212	212	212	212	213	214	216	219	220	221	220	222	224	226	231	232	232	232
2012	165	257	287	279	279	291	274	266	274	277	273	274	273	274	274	276	278	283	284	285	284	287	290	292	299	299	300	300
2013	140	172	202	205	221	214	220	213	222	221	219	219	218	219	219	221	223	226	227	228	227	229	231	233	239	240	240	240
2014	186	233	243	241	246	251	242	236	234	234	231	231	230	231	231	233	235	238	240	240	239	242	244	246	252	252	253	253
2015	186	239	254	263	276	273	263	255	254	253	250	250	249	250	250	252	254	258	260	260	259	262	264	266	273	274	274	274
2016	222	275	280	279	273	255	252	245	244	243	240	240	239	240	240	242	244	248	249	250	249	252	254	256	262	263	263	263
2017	210	268	265	279	289	283	280	272	270	270	266	267	266	267	267	269	271	275	277	277	276	279	282	284	291	292	292	292

Figure 3.6 : Triangle complété par la méthode Chain Ladder classique

Les résultats sont récapitulés dans le tableau suivant :

	Charge Ultime	Provisions D/D	Montant des IBNR
1996	120	2	-
1997	148	10	1
1998	221	16	6
1999	221	4	7
2000	211	1	9
2001	189	6	10
2002	187	3	9
2003	197	10	10
2004	114	12	6
2005	211	12	15
2006	180	13	14
2007	141	1	12
2008	252	18	22
2009	186	12	16
2010	238	39	20
2011	232	47	20
2012	300	57	23
2013	240	49	17
2014	253	102	17
2015	274	98	11
2016	263	128	7
2017	292	166	3
2018	301	218	5
2019	318	243	11
2020	288	240	24
2021	352	247	104
Total	5 928	1 753	402

Figure 3.7 : Résultats de la méthode Chain Ladder sans pondération pour les sinistres français de la classe "Large"

Le développement de la classe « Large » est assez long puisque les sinistres survenus en 1997 nécessitent encore de constituer des provisions pour les IBNR.

3.1.2 Extension pondérée

3.1.2.1 Présentation de la méthode

Cette méthode introduit des pondérations dans la version classique de Chain Ladder afin de se ramener dans un cadre où les hypothèses de la méthode sont vérifiées. Les facteurs de développement vont prendre en compte ces pondérations :

$$\forall j \in \{1, \dots, n-1\}, \quad \hat{f}_j = \frac{\sum_{i=1}^{n-j} w_{i,j} \cdot C_{i,j}^\alpha \cdot f_{i,j}}{\sum_{i=1}^{n-j} w_{i,j} \cdot C_{i,j}^\alpha}$$

Les $w_{i,j} \in [0,1]$ sont les poids qui peuvent être choisis par l'actuaire pour accorder moins de poids à un ou plusieurs facteurs individuels $f_{i,j}$. Le cas précédent correspond au cas $w_{i,j} = 1$ et $\alpha = 1$. Si $\alpha \neq 1$, on n'utilise plus les facteurs individuels historique (moyenne pondérée par les volumes des montants cumulés) mais une moyenne arithmétique dans le cas où $\alpha = 0$ et le résultat de la minimisation de la régression de $C_{i,j+1}$ sur $C_{i,j}$ dans le cas où $\alpha = 2$.

Pour choisir les poids, il existe différentes méthodes. Les plus utilisées sont :

- La méthode dite « d'expert » où l'actuaire peut se servir d'information exogène au triangle pour enlever un ou plusieurs coefficients qu'il sait non représentatifs de la réalité ;
- La méthode min-max où l'actuaire décide de supprimer les facteurs minimum et maximum pour chaque période de développement afin d'enlever les événements extrêmes ;
- La vision récente où l'actuaire choisit de ne garder que les années les plus récentes. Cela permet de prendre en compte un changement survenu sur le portefeuille (politique de souscription, mode de gestion de sinistres ...).

Sur la figure 3.2, les diagonales semblent scindées en deux. Ainsi, les facteurs -1 se trouvent majoritairement dans la partie haute de la diagonale du triangle alors que les facteurs 1 sont majoritairement dans la partie basse pour les diagonales récentes alors que l'inverse se produit sur les anciennes années.

La méthode de la vision récente pourrait être appropriée pour les pondérations des sinistres issus du portefeuille français car il semble y avoir un changement entre les années passées et récentes. Pour les sinistres de ce portefeuille, nous choisissons de ne conserver que les 6 dernières années.

Pour les sinistres issus du portefeuille anglais, il est intéressant de faire des sélections d'expert afin d'enlever les effets liés au changement du taux OGDEN sur les années de calendaires concernées par les changements du taux. En effet, nous ne souhaitons pas projeter les évolutions de ce taux mais uniquement les montants des sinistres.

3.1.2.2 Application aux données

Données françaises – Classe « Attritional »

Nous répétons le test fait précédemment sur les sinistres français de la classe « Attritional » en gardant uniquement les 6 dernières années pour vérifier que la deuxième hypothèse est maintenant vérifiée. L'intervalle de confiance est maintenant [32 ; 43] et notre valeur de Z est de 34. La deuxième hypothèse est maintenant validée.

Comme les hypothèses sont validées, la méthode Chain Ladder pondérée peut donc être utilisée. Voici les résultats obtenus :

	Charge Ultime	Provisions D/D	Montant des IBNR
1996	84	0	-
1997	69	1	0
1998	80	2	0
1999	84	0	0
2000	81	3	0
2001	86	0	0
2002	86	0	0
2003	78	0	0
2004	66	0	0
2005	64	0	0
2006	62	0	0
2007	63	0	0
2008	63	1	1
2009	65	1	1
2010	69	1	1
2011	69	1	1
2012	70	2	1
2013	66	1	1
2014	63	4	1
2015	64	4	2
2016	66	5	2
2017	69	8	3
2018	68	11	4
2019	71	19	7
2020	58	24	8
2021	71	43	4
Total	1 835	118	28

Figure 3.8 : Résultats de la méthode Chain Ladder avec pondération pour les sinistres de la classe "Attritional"

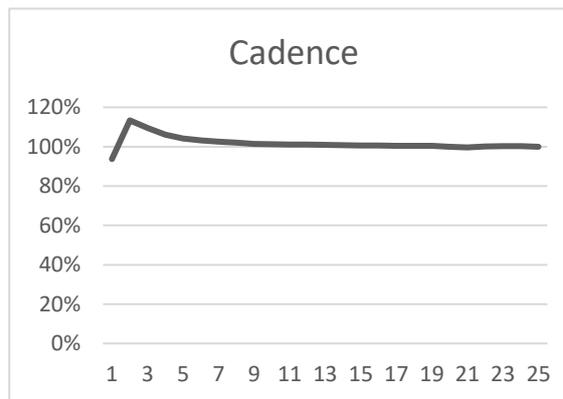


Figure 3.9 : Cadence de développement des sinistres de la classe "Attritional"

Le montant total des IBNR est négatif et cela est dû à la forme de la cadence de développement. La première année, le coefficient de développement calculé est supérieur à 1 mais ensuite, les coefficients deviennent tous inférieurs à 1. C'est pour cela que la cadence de développement croît au-dessus de 100% puis diminue ensuite. Cela s'explique, d'une part, par des procédures de recours et d'autre part, par la libération des provisions dossier/dossier lorsqu'elles étaient trop prudentes (boni sur les réserves). Le montant final du sinistre est obtenu après environ 10 ans.

Données françaises – Classe « Exceptionnelle »

En pondérant de la même manière la classe « Exceptionnelle », la deuxième hypothèse est aussi vérifiée puisque l'intervalle de confiance est [61 ; 76] et la statistique vaut 70. Il est donc aussi possible d'appliquer une version pondérée de la méthode Chain Ladder sur la base Exceptionnelle. Voici les montants totaux obtenus :

	Charge Ultime	Provisions D/D	Montant des IBNR
Total	10 179	3 525	241

Figure 3.10 : Résultats de la méthode Chain Ladder avec pondération pour les données de la classe "Exceptionnelle"

Données françaises – Classe « Large »

Bien que les hypothèses de Chain Ladder soient vérifiées pour cette classe, nous allons à titre indicatif sélectionner les mêmes diagonales que pour les classes « Attritional » et « Exceptionnelle » afin de comparer les résultats de la méthode non pondérée avec la version pondérée.

	Charge Ultime	Provisions D/D	Montant des IBNR
1996	120	2	-
1997	148	10	1
1998	221	16	6
1999	221	4	7
2000	211	1	9
2001	189	6	10
2002	187	3	9
2003	197	10	10
2004	114	12	6
2005	212	12	16
2006	182	13	15
2007	143	1	14
2008	254	18	23
2009	187	12	17
2010	239	39	21
2011	233	47	21
2012	301	57	24
2013	236	49	14
2014	255	102	20
2015	279	98	17
2016	265	128	10
2017	297	166	8
2018	314	218	18
2019	336	243	30
2020	308	240	44
2021	387	247	140
Total	6 035	1 753	509

Figure 3.11 : Résultats de la méthode Chain Ladder avec pondération pour les données de la classe "Large"

Le montant à provisionner varie à la hausse en ne sélectionnant que les années récentes. Cela est dû à la cadence de développement qui est impactée par les poids :

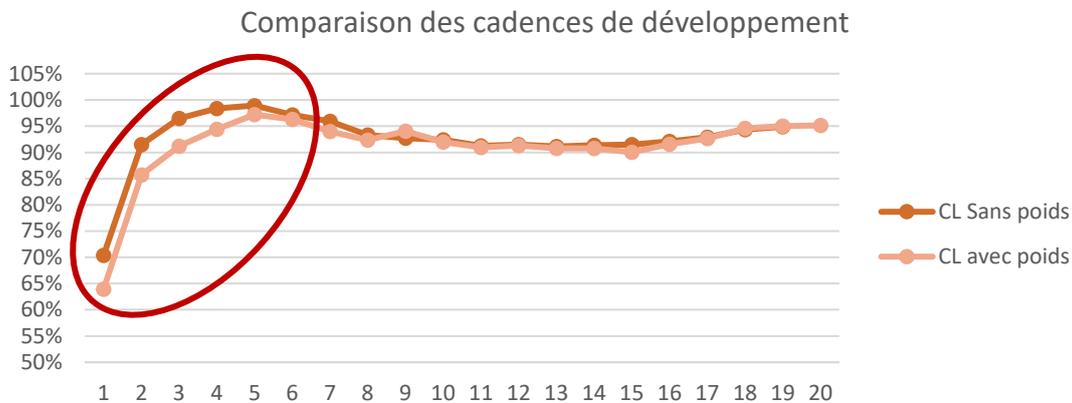


Figure 3.12 : Comparaison des cadences de développement entre une version pondérée et non pondérée

La cadence avec poids suppose que les sinistres se développent plus lentement que la version sans poids. Pour une même période de développement, plus de provisions seront donc nécessaires pour la version pondérée. L'écart entre les deux cadences tend à s'estomper après la 10^e année de développement ce qui explique que les montants de provisions soient identiques dans les tableaux ci-dessus pour les années anciennes.

Nous notons l'importance de la répartition en groupe homogène pour provisionner. En effet, les sinistres n'ont pas du tout le même comportement (cadences très différentes) en fonction de leur classe.

Données d'acceptations anglaises

Avec le retraitement des variations du taux OGDEN (neutralisation des diagonales 2017 et 2019), l'intervalle de confiance pour la deuxième hypothèse est de [46 ;61] et notre statistique vaut 50. La deuxième hypothèse est donc validée.

En appliquant la méthode, voici les résultats obtenus :

	Charge Ultime	Provisions D/D	Montant des IBNR
2002	467	14	-
2003	1 121	527 -	3
2004	1 076	449 -	13
2005	642	320 -	9
2006	453	194 -	5
2007	925	622 -	21
2008	834	416 -	23
2009	659	291 -	27
2010	894	612 -	37
2011	1 039	593	2
2012	1 460	887	7
2013	1 698	625 -	5
2014	2 195	1 184	10
2015	2 181	1 527	12
2016	1 733	1 580 -	14
2017	566	511	9
2018	60	55	1
2019	29	28	1
2020	32	25	7
2021	710	214	496
Total	18 774	10 675	388

Figure 3.14 : Résultats de la méthode Chain Ladder avec pondération pour les polices anglaises

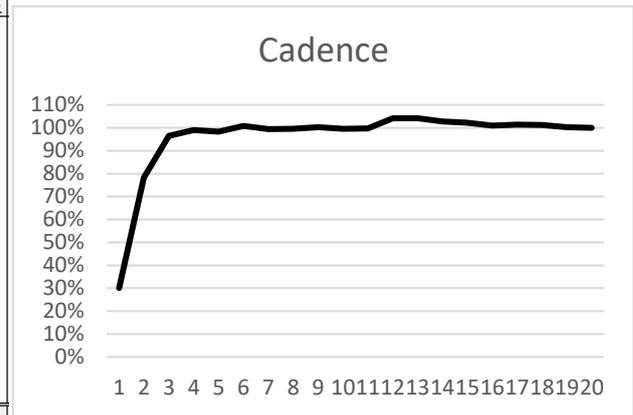


Figure 3.13 : Cadence de développement des polices anglaises

La cadence de développement pour ce portefeuille est plus lente que pour le portefeuille des polices françaises. Cela signifie que beaucoup d'IBNR sont nécessaires sur la dernière année observée car seulement 30% du montant final est observé dans les données.

Conclusion sur la méthode Chain Ladder

L'avantage de cette méthode est sa simplicité d'utilisation et c'est ce qui fait d'elle la méthode la plus utilisée par les actuaires non-vie aujourd'hui. Néanmoins, elle présente un gros inconvénient pour les années récentes. En effet, nous ne disposons que de peu d'informations et l'estimation peut donc être très incertaine voire erronée sur les exercices de survenance récents dans le cas où la charge connue est proche de ou égale à 0 (sous-estimation de la charge ultime compte tenu du caractère multiplicatif de la méthode). Ce cas de figure peut être très courant en réassurance par exemple où les informations des cédantes sont connues plus tardivement par le réassureur.

3.2 Méthode de Bornhuetter-Ferguson

3.2.1 Méthode classique

Proposée en 1972, cette méthode permet de combler certaines failles de la méthode Chain Ladder puisque les provisions ne seront pas déterminées uniquement à partir des dernières observations connues mais également à partir de données exogènes au triangle grâce à des ratios sinistres/primes ultimes a priori (noté ULR dans la suite). La propriété fondamentale de cette méthode est l'indépendance entre les provisions à constituer et le dernier montant connu. En effet, cette méthode suppose que les montants déclarés à ce jour ne sont pas prédictifs des montants des sinistres non déclarés.

Pour utiliser la méthode de Bornhuetter-Ferguson, il faut estimer la charge de sinistres. Elle est estimée par le produit suivant :

$$\hat{U}_i^{BF} = ULR_i \times v_i$$

avec v_i les primes acquises l'année i et ULR_i le ratio à l'ultime estimé. En pratique, ce sont les données de la tarification, du business plan ou des données de marché qui sont utilisées pour déterminer ces ratios.

L'estimateur de la charge ultime pour l'exercice de survenance i est :

$$\hat{C}_{i,n}^{BF} = C_{i,n+1-i} + (1 - z_{n+1-i}) \hat{U}_i^{BF}$$

où z_{n+1-i} est la cadence de développement après $(n + 1 - i)$ années de développement.

Il est courant de reprendre les estimations données par la méthode Chain Ladder pour la cadence de développement :

$$z_n = 1, \hat{z}_{n-1} = \frac{1}{\hat{f}_n}, \dots, \hat{z}_1 = \frac{1}{\prod_{k=2}^n \hat{f}_k} \text{ où les } \hat{f}_k \text{ sont les coefficients calculés précédemment.}$$

L'estimation de la charge ultime peut être considérée comme une moyenne pondérée suivant la crédibilité des estimations données par la méthode Chain Ladder et par les montants ultimes estimées à partir des ULR. Plus les facteurs de développement sont élevés (années récentes) et plus il y a de poids sur les estimations basées sur les ULR. Pour les anciennes années, la crédibilité est donnée à la méthode Chain Ladder.

La provision par exercice de survenance est donnée par :

$$\hat{R}_i^{BF} = (1 - z_{n+1-i}) \hat{U}_i^{BF}$$

Une autre méthode déterministe fréquemment utilisée par les assureurs, notamment pour la projection des exercices récents, est la méthode dite du « Loss Ratio » (ou méthode du « S/P cible »). Cette méthode, extrêmement simple, consiste à appliquer un ratio sinistres à primes *a priori* (i.e. un ULR) pour déterminer la charge ultime et en soustraire les sinistres connus pour obtenir une estimation de la provision :

L'estimateur de la charge ultime pour l'exercice de survenance i est donné par :

$$\hat{C}_{i,n}^{LR} = \hat{U}_i^{LR}$$

avec

$$\hat{U}_i^{LR} = ULR_i \times v_i$$

La provision correspondante est donnée par : $\hat{R}_i^{LR} = \hat{U}_i^{LR} - C_{i,n+1-i}$

La méthode du loss ratio peut ainsi être vue comme un cas particulier de la méthode Bornhuetter-Ferguson en considérant la cadence de développement « triviale » suivante :

$$z_{n+1-i} = \frac{C_{i,n+1-i}}{\hat{U}_i^{LR}}$$

Nous n'aborderons plus dans la suite la méthode du loss ratio mais nos développements concernant la méthode Bornhuetter-Ferguson sont applicables sans perte de généralité à la méthode du loss ratio.

Nous allons appliquer la méthode Bornhuetter-Ferguson uniquement sur le triangle des charges issu du portefeuille d'acceptations anglais. En effet, il s'agit de la seule base pour laquelle les ULR sont fournis par les souscripteurs. La cadence utilisée est celle de Chain Ladder pondérée (car les hypothèses étaient toutes vérifiées).

	ULR	Montant Ultime	IBNR
2002	88%	467	-
2003	106%	1 121	- 3
2004	74%	1 075	- 15
2005	83%	640	- 10
2006	80%	452	- 5
2007	115%	920	- 26
2008	122%	827	- 31
2009	82%	651	- 35
2010	83%	885	- 46
2011	80%	1 040	3
2012	109%	1 462	9
2013	94%	1 696	- 7
2014	108%	2 198	13
2015	109%	2 184	15
2016	118%	1 729	- 18
2017	123%	570	14
2018	88%	61	2
2019	88%	39	10
2020	70%	144	118
2021	64%	979	766
Total		19 142	756

Figure 3.15 : Résultat de la méthode classique

Le montant total d'IBNR est assez proche de celui donné par Chain Ladder pour les années anciennes. Jusqu'en 2017, les ordres de grandeurs sont les mêmes car le poids donné à Chain Ladder est très important. Après 2017, c'est l'information exogène via les ULR qui a le plus de poids. Ainsi, même si les années récentes sont pour le moment assez peu sinistrées (et ont donc peu d'IBNR par la méthode Chain Ladder), les ratios ultimes prévus sont assez élevés et la méthode de Bornhuetter-Ferguson va conduire à estimer un montant supérieur d'IBNR.

Conclusion sur la méthode Bornhuetter-Ferguson classique

Les données éventuellement aberrantes des années récentes ont un impact mineur sur les projections et cette méthode peut être utilisée même si les données sont volatiles. Cependant, elle dépend fortement des ULR choisis initialement et n'est donc pas totalement robuste (dépend des choix de l'actuaire/souscripteur). De plus, ce n'est pas une méthode à part entière puisqu'elle dépend des facteurs de développement de la méthode Chain Ladder.

3.2.2 Méthode indépendante (Mack (2006))

3.2.2.1 Présentation de la méthode

En 2006, T. Mack propose une nouvelle version de la méthode de Bornhuetter-Ferguson, indépendante des facteurs de développement de Chain Ladder.

Pour pouvoir utiliser cette méthode, il est nécessaire de vérifier les hypothèses suivantes :

(BF1)

Les incréments $S_{i,1}, \dots, S_{i,n}$ sont indépendants

(BF2)

$\exists x_i, z_j$ tels que $E(C_{i,j}) = x_i z_j$

où x_i représente le montant ultime pour chaque année i et z_j la cadence de développement.

La formule utilisée pour déterminer les provisions est toujours : $\widehat{R}_t^{BF} = (1 - \hat{z}_{n+1-i}) \widehat{U}_i$

avec $\widehat{U}_i = v_i \hat{q}_i$ où \hat{q}_i est l'ULR pour l'année de survenance i , v_i est le montant de prime et $z_j \in [0,1]$ est la proportion de sinistres qui n'ont pas encore été reportés (cadence de développement).

Cependant ici, les ULR et la cadence de développement vont être estimés à partir des données. Ce choix est justifié par le fait qu'utiliser le développement de la méthode de Chain Ladder dans l'estimation Bornhuetter-Ferguson ne respecte pas la condition d'indépendance requise entre les montants cumulés observés et les montants de provision estimés.

- Estimation de l'ULR

Deux étapes sont nécessaires pour pouvoir calculer les ULR. La première est de calculer les ILR (Incremental Loss Ratio) définis comme :

$$\hat{m}_j = \frac{\sum_{i=1}^{n+1-j} S_{i,j}}{\sum_{i=1}^{n+1-j} v_i}$$

La somme des ILR permet d'avoir une estimation a priori de l'ULR pour une année de survenance moyenne.

La deuxième étape est de calculer le niveau relatif d'adéquation \hat{r}_i pour chaque année de survenance i . Le niveau d'adéquation mesure le changement entre le ratio technique des polices à leur date d'expiration par rapport à la date de leur renouvellement. Si une année de survenance a un niveau relatif d'adéquation en dessous du niveau moyen, cela signifie que le montant des primes est inférieur à celui qui serait observé sur une année moyenne. Du point de vue des primes, \hat{r}_i désigne donc le facteur par lequel les montants de primes v_i doivent être multipliés pour s'ajuster au niveau moyen. Pour calculer \hat{r}_i , nous utilisons la moyenne pondérée des ratios $\frac{S_{i,j}}{v_i}$ et \hat{m}_j définie comme :

$$\hat{r}_i = \sum_{j=1}^{n+1-i} \frac{\hat{m}_j}{\sum \hat{m}_k} \cdot \frac{S_{i,j}/v_i}{\hat{m}_j}$$

Il convient de l'incorporer dans le calcul des ILR :

$$\tilde{m}_j = \frac{\sum_{i=1}^{n+1-j} S_{i,j}}{\sum_{i=1}^{n+1-j} v_i \hat{r}_i}$$

Il peut être nécessaire de lisser les derniers coefficients pour qu'ils convergent vers 0. Pour cela, il faut ajuster une courbe aux points calculés afin d'obtenir une estimation des paramètres de la loi choisie pour le lissage. Nous avons testé deux lois pour notre lissage : la loi exponentielle et la loi puissance inverse.

Cas exponentiel

La densité de la loi exponentielle à deux paramètres (a et b) est donnée par :

$$f(x) = b \cdot \exp(-ax)$$

Voici le résultat de la superposition de nos coefficients calculés et du lissage exponentiel :

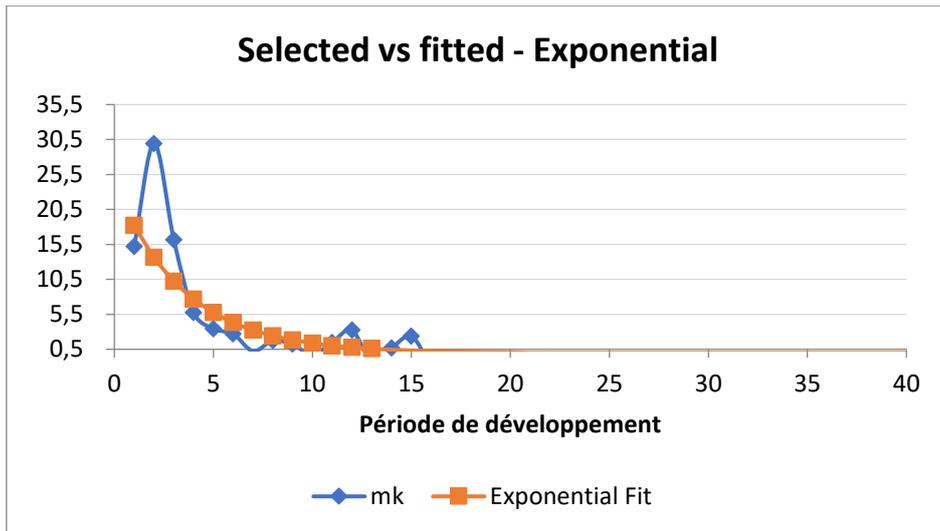


Figure 3.16 : Lissage exponentiel

Les coefficients \tilde{m}_j sont très volatiles au départ et il est assez compliqué d'ajuster la loi sur les premières périodes de développement. Néanmoins, sur les dernières périodes, l'ajustement est mieux réalisé.

La qualité de l'adéquation entre nos données et la loi est mesurée par le coefficient R^2 . Dans notre cas, le coefficient R^2 vaut 83%.

Cas puissance inverse

La densité de la loi puissance inverse à deux paramètres (a et b) est donnée par :

$$f(x) = bx^{-a}$$

L'ajustement avec la loi puissance inverse est meilleur puisque le R^2 est de 96%.

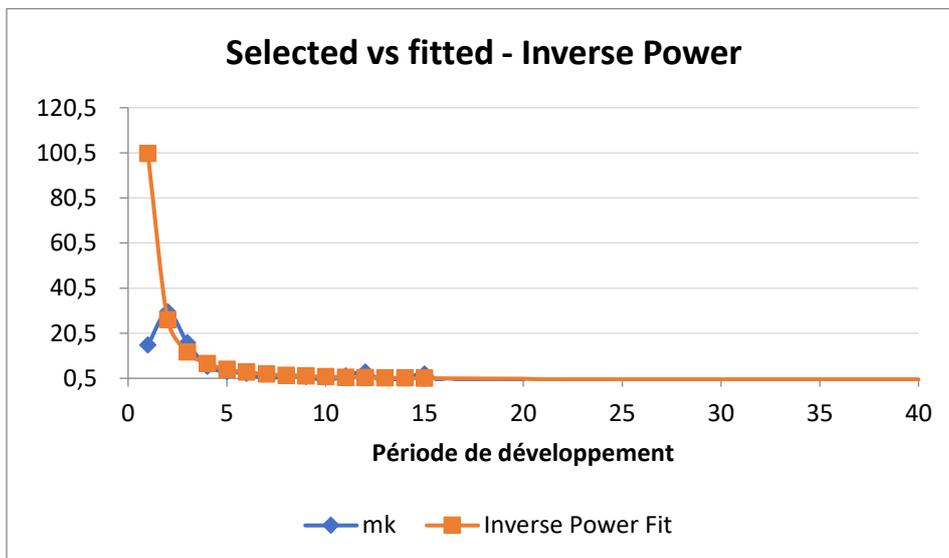


Figure 3.17 : Lissage puissance inverse

Nous allons donc utiliser le lissage à partir de la loi puissance inverse pour les derniers coefficients \tilde{m}_j .

Finalement, l'ULR est défini comme $\hat{q}_i = \hat{r}_i \cdot \tilde{m}_j$

- Estimation de la cadence de développement

La cadence est définie comme $\hat{z}_j = \frac{\hat{m}_1 + \dots + \hat{m}_j}{\hat{m}_1 + \dots + \hat{m}_n + \hat{m}_{n+1}}$ où \hat{m}_{n+1} est un ratio de queue moyen (pour prendre en compte la possibilité d'un développement après l'année n). Typiquement, si le développement est terminé, il prend la valeur 0.

3.2.2.2 Application aux données

Dans un premier temps, cette méthode sera uniquement appliquée sur le triangle incrémental des données d'acceptations anglaises car nous pouvons comparer les résultats avec la version classique. Pour vérifier la première hypothèse, les montants ne doivent plus seulement être indépendants selon les années de survenance comme dans la méthode Chain Ladder mais ils doivent être indépendants selon tous les axes temporels du triangle (par années de survenance, par années calendaires et par années de développement).

L'indépendance des années de survenance peut être vérifiée à l'aide du test d'absence d'effet calendaire sur les montants incrémentaux. Notre statistique vaut 65 et l'intervalle de validation est [60 ; 75]. L'indépendance selon les années de survenance est acceptée.

Pour l'indépendance des années de développement, il est possible de réaliser un test statistique pour juger si une séquence est aléatoire.

Soit $A_i = \{X_{i,j} \geq x_i\}$ l'évènement correspondant à un incrément supérieur à la médiane notée x_i et $B_i = \{X_{i,j} < x_i\}$ l'évènement contraire. Pour chaque année de survenance, les incréments vont donc former une suite de A et de B en fonction de leur position par rapport à la médiane. Le nombre de A sera noté p et le nombre de B sera noté m . La suite comprend au total n termes avec $n = m + p$.

S'il existe une tendance, pendant une certaine durée, les observations successives seront situées soit uniquement au-dessus soit uniquement en dessous de la médiane.

Soit S la variable aléatoire égale au nombre de séquences homogènes (une séquence correspond au changement de lettre).

Le résultat obtenu avec nos données se trouve ci-dessous. Pour la première ligne du triangle, un total de 8 séquences est observé.

Cependant, comme dans notre exemple, les incréments du triangle peuvent ne pas tous vérifier les hypothèses et cette méthode ne permet pas d'intégrer des pondérations pour se ramener à un cadre où les hypothèses seraient validées.

3.2.3 Méthode indépendante pondérée

3.2.3.1 Présentation de la méthode

Lorsque les hypothèses de Chain Ladder ne sont pas vérifiées, une possibilité est d'utiliser des poids pour ne sélectionner que des années qui vérifient les conditions nécessaires. L'utilité des poids est aussi d'écarter les sinistres (ou les années) identifiés comme « anormaux ». La cadence estimée avec la méthode de Mack (2006) ne permet pas de faire ces sélections. Nous proposons dans la suite une extension de son modèle au cas pondéré.

La cadence est toujours définie par :

$$\hat{z}_j = \frac{\tilde{m}_1 + \dots + \tilde{m}_j}{\tilde{m}_1 + \dots + \tilde{m}_n + \tilde{m}_{n+1}}$$

En revanche, pour prendre en compte une pondération, la définition des ILR doit être modifiée. Les ILR vont maintenant être définis par :

$$\tilde{m}_j = \frac{\sum_{i=1}^{n+1-k} w_{i,k} S_{i,k}}{\sum_{i=1}^{n+1-k} w_{i,k} v_i} \text{ où } w_{i,k} \in \{0,1\} \text{ est le poids souhaité.}$$

3.2.3.2 Application aux données

La version non pondérée de la méthode n'était pas applicable. Pour vérifier les hypothèses, nous avons mis à 0 les années calendaires 2017, 2018 et 2019 (années où le taux OGDEN a eu un impact) et les hypothèses sont maintenant vérifiées.

Voici les résultats obtenus avec la méthode indépendante pondérée :

	Primes	Provisions D/D	Estimations			
			ULR	Montant ultime	% restant à développer	IBNR
2002	531	14	73,7%	391	0,0%	0
2003	1 083	527	86,4%	936	0,0%	0
2004	1 594	449	55,2%	880	0,0%	0
2005	870	320	59,9%	521	0,0%	0
2006	663	194	54,0%	358	0,0%	0
2007	987	622	71,9%	710	0,0%	0
2008	921	416	73,5%	677	0,0%	0
2009	1 029	291	47,4%	488	0,0%	0
2010	1 313	612	47,6%	625	0,1%	0
2011	1 657	593	42,1%	697	0,1%	1
2012	1 783	887	53,7%	958	0,2%	2
2013	2 395	625	45,5%	1 089	0,4%	4
2014	2 623	1 184	51,6%	1 353	0,7%	10
2015	2 492	1 527	54,4%	1 355	1,3%	18
2016	1 921	1 580	53,9%	1 037	2,4%	25
2017	670	511	48,5%	325	4,4%	14
2018	291	55	11,2%	32	8,1%	3
2019	333	28	4,9%	16	12,3%	2
2020	776	25	2,2%	17	26,4%	5
2021	1 705	214	21,4%	364	69,5%	253
Total	25 637	10 675	50%	12 830		336

Figure 3.19 : Résultats de la méthode indépendante pondérée

Les montants d'IBNR donnés par la méthode indépendante pondérée sont assez proches de la version Chain Ladder pondérée même si, à cause du lissage, il n'y a pas d'IBNR négatifs.

En comparant les cadences de développement estimées par chacune des méthodes, nous remarquons un petit écart entre la 4^e et 6^e période de développement. Néanmoins, la cadence estimée par la méthode Bornhuetter-Ferguson indépendante pondérée semble tout aussi cohérente.

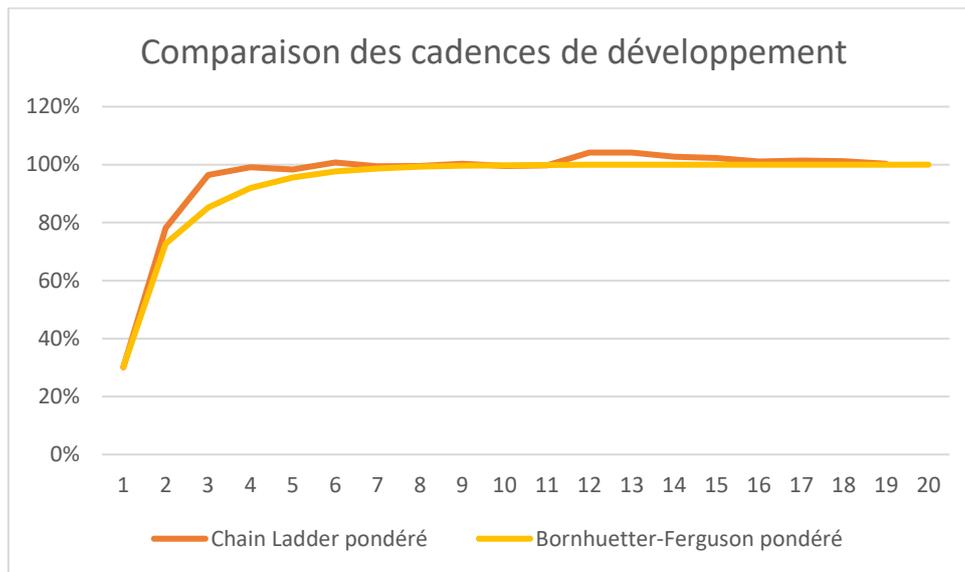


Figure 3.20 : Comparaison des cadences de développement

En revanche, en étudiant l'estimation des ULR, nous remarquons que les ULR estimés sont assez différents des ULR fournis par les souscripteurs. En traçant ces ULR, nous obtenons le graphique suivant :

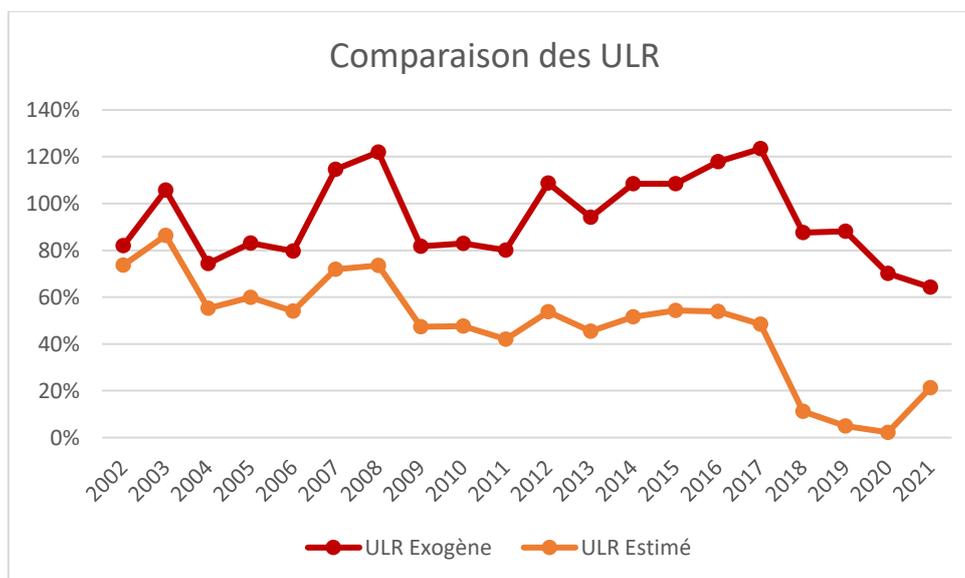


Figure 3.21 : Comparaison des ULR

Les ULR estimés et exogènes suivent les mêmes tendances mais les ULR dérivés des données sont sensiblement inférieurs aux ULR des souscripteurs. Sur les années récentes, cela peut être expliqué par le manque d'information (seulement 1, 2 ou 3 observations dans le triangle pour les 2019 à 2021). Sur les années anciennes, il y a deux possibilités : soit les souscripteurs surestiment le montant des sinistres attendus, soit il y a encore d'importants sinistres qui ne sont pas encore arrivés.

Les ULR calculés en utilisant Chain Ladder sont assez proches des ULR estimés par la méthode Bornhuetter-Ferguson, confirmant qu'avec les données actuelles, nous nous attendons à avoir des ULR faibles. Néanmoins, il pourrait être judicieux pour la suite de conserver une version avec les ULR donnés par les souscripteurs, sur les années récentes au moins, car des ratios ultimes inférieurs à 20% en 2018, 2019 et 2020 semblent peu probables. Voici les résultats obtenus en conservant les ULR des souscripteurs et en utilisant la cadence estimée :

	Primes	Provisions D/D	Estimations			
			ULR	Montant ultime	% restant à développer	IBNR
2002	531	14	82,0%	435	0,0%	0
2003	1 083	527	105,8%	1 146	0,0%	0
2004	1 594	449	74,3%	1 184	0,0%	0
2005	870	320	83,1%	723	0,0%	0
2006	663	194	79,7%	529	0,0%	0
2007	987	622	114,6%	1 132	0,0%	0
2008	921	416	121,9%	1 122	0,0%	0
2009	1 029	291	81,7%	841	0,0%	0
2010	1 313	612	83,0%	1 089	0,1%	1
2011	1 657	593	80,1%	1 327	0,1%	2
2012	1 783	887	108,7%	1 938	0,2%	4
2013	2 395	625	94,2%	2 255	0,4%	9
2014	2 623	1 184	108,5%	2 846	0,7%	20
2015	2 492	1 527	108,5%	2 704	1,3%	35
2016	1 921	1 580	117,8%	2 263	2,4%	54
2017	670	511	123,4%	827	4,4%	36
2018	291	55	87,5%	254	8,1%	21
2019	333	28	88,1%	293	12,3%	36
2020	776	25	70,1%	544	26,4%	144
2021	1 705	214	64,2%	1 095	69,5%	761
Total	25 637	10 675	96%	24 549		1 123

Figure 3.22 : Résultats de la méthode indépendante pondérée avec des ULR donnés

Les IBNR augmentent drastiquement puisque les ULR souscripteurs sont largement plus élevés que les ULR estimés par la méthode Bornhuetter-Ferguson pondérée. Cette estimation est extrême et ne représente sûrement pas la réalité puisque nous utilisons une cadence qui ne permet pas d'IBNR négatifs à cause du lissage avec des ULR très élevés.

Nous en concluons que tout comme pour la version classique de Bornhuetter-Ferguson, l'estimation que nous faisons des ULR influe très fortement sur le montant des provisions.

Comme l'estimation de la cadence est satisfaisante et que celle des ULR semble cohérente au moins au niveau des tendances, nous allons appliquer la méthode aux données françaises. Pour chacune des 3 catégories, il a été nécessaire de sélectionner uniquement les années qui permettent que les hypothèses soient vérifiées. Les résultats obtenus sont présentés ci-dessous.

Classe « Attritional »

				Estimations			
	Primes	Provisions D/D		ULR	Montant ultime	% restant à développer	IBNR
1996	129	0		66,0%	85	0,0%	0
1997	142 -	1		49,0%	70	0,0%	0
1998	137 -	2		58,3%	80	0,1%	0
1999	137	0		61,5%	84	0,1%	0
2000	141 -	3		57,9%	82	0,2%	0
2001	146 -	0		59,3%	87	0,2%	0
2002	160 -	0		53,6%	86	0,3%	0
2003	161	0		48,5%	78	0,3%	0
2004	149 -	0		44,6%	66	0,4%	0
2005	142	0		45,3%	64	0,5%	0
2006	137	0		45,4%	62	0,6%	0
2007	132	0		47,6%	63	0,7%	0
2008	130	1		48,3%	63	0,8%	1
2009	132	1		49,7%	65	1,0%	1
2010	136	1		51,6%	70	1,0%	1
2011	139	1		50,2%	70	0,9%	1
2012	144	2		48,9%	70	0,8%	1
2013	144	1		46,4%	67	0,6%	0
2014	127	4		50,0%	64	0,0%	0
2015	121	4		53,1%	64	-0,4%	0
2016	122	5		54,9%	67	-1,0%	-1
2017	121	8		57,5%	70	-2,0%	-1
2018	124	11		55,4%	69	-4,2%	-3
2019	125	19		57,2%	72	-7,1%	-5
2020	129	24		45,3%	59	-11,2%	-7
2021	132	43		53,7%	71	7,4%	5
Total	3 541	118		52%	1 847		-6

Figure 3.23 : Résultats du Bornhuetter-Ferguson indépendant - Classe "Attritional"

Classe « Large »

				Estimations			
	Primes	Provisions D/D		ULR	Montant ultime	% restant à développer	IBNR
1996	3 161	2		4,0%	126	0,0%	0
1997	3 504	10		4,4%	155	1,0%	2
1998	3 374	16		6,8%	229	2,1%	5
1999	3 378	4		6,8%	229	3,2%	7
2000	3 470	1		6,3%	218	4,4%	10
2001	3 596	6		5,4%	195	5,6%	11
2002	3 934	3		4,9%	193	6,9%	13
2003	3 958	10		5,2%	204	8,2%	17
2004	3 669	12		3,2%	118	8,7%	10
2005	3 490	12		6,2%	217	10,1%	22
2006	3 370	13		5,5%	185	10,9%	20
2007	3 248	1		4,4%	144	11,5%	17
2008	3 209	18		8,0%	258	11,6%	30
2009	3 234	12		5,9%	190	11,9%	23
2010	3 336	39		7,3%	244	11,5%	28
2011	3 421	47		7,0%	238	11,8%	28
2012	3 533	57		8,7%	307	10,6%	33
2013	3 550	49		6,9%	246	10,3%	25
2014	3 125	102		8,3%	260	9,8%	25
2015	2 975	98		9,5%	282	7,2%	20
2016	3 007	128		9,0%	271	6,0%	16
2017	2 986	166		10,1%	301	4,0%	12
2018	3 058	218		10,1%	310	4,5%	14
2019	3 085	243		10,6%	327	6,3%	21
2020	3 175	240		9,3%	296	11,2%	33
2021	3 245	247		11,1%	360	31,6%	114
Total	87 088	1 753		7%	6 105		556

Figure 3.24 : Résultats du Bornhuetter-Ferguson indépendant - Classe "Large"

Pour ces deux classes, le montant des IBNR est très proche de celui obtenu en appliquant la méthode Chain Ladder pondérée (respectivement -28 € pour la classe « Attritional » et 509 € pour la classe « Large »).

Classe « Exceptionnelle »

	Primes	Provisions D/D	Estimations			IBNR
			ULR	Montant ultime	% restant à développer	
1996	3 319	0	4,1%	136	0,0%	0
1997	3 679	8	13,9%	510	0,0%	0
1998	3 541	47	8,9%	313	0,1%	0
1999	3 546	6	4,8%	172	0,2%	0
2000	3 643	83	10,6%	385	0,2%	1
2001	3 775	3	7,7%	292	0,3%	1
2002	4 129	14	7,1%	295	0,4%	1
2003	4 154	3	3,2%	133	0,4%	1
2004	3 851	31	12,5%	481	0,5%	3
2005	3 663	9	10,6%	388	0,6%	2
2006	3 537	90	10,8%	382	0,7%	3
2007	3 409	56	12,1%	414	0,9%	4
2008	3 368	18	8,1%	273	1,0%	3
2009	3 395	149	14,5%	493	1,2%	6
2010	3 501	232	17,4%	611	1,4%	8
2011	3 591	109	13,4%	480	1,6%	8
2012	3 709	280	18,1%	672	1,9%	13
2013	3 726	120	12,6%	468	2,2%	10
2014	3 280	158	13,9%	457	2,7%	12
2015	3 123	103	7,7%	241	3,2%	8
2016	3 156	262	14,6%	461	3,9%	18
2017	3 134	264	11,9%	373	4,8%	18
2018	3 210	415	15,8%	506	6,2%	31
2019	3 238	414	15,8%	513	8,6%	44
2020	3 333	334	11,5%	384	13,5%	52
2021	3 406	315	12,7%	433	29,8%	129
Total	91 416	3 525	11%	10 265		375

Figure 3.25 : Résultats du Bornhuetter-Ferguson indépendant - Classe "Exceptionnelle"

Le montant d'IBNR calculé avec la méthode Bornhuetter-Ferguson est plus élevé que celui calculé avec Chain Ladder (241 €) car l'estimation de la cadence est assez différente entre les deux méthodes :

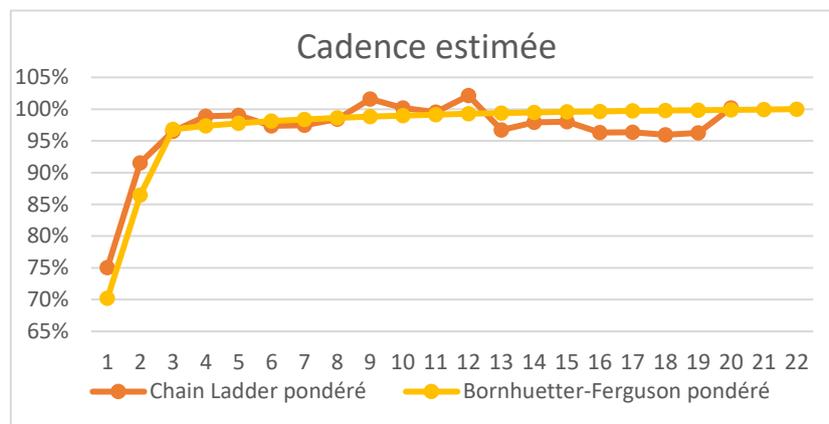


Figure 3.26 : Comparaison des cadences estimées

Sans la possibilité d'évaluer l'erreur de prédiction, il est difficile de choisir la méthode la plus optimale.

Conclusion sur la méthode Bornhuetter-Ferguson indépendante pondérée

Cette méthode estime des cadences stables et proches de celles trouvées avec la méthode Chain Ladder pondérée tout en respectant l'idée d'indépendance entre les derniers montants connus et le niveau des réserves à constituer.

Concernant l'estimation des ULR, l'estimation semble correcte en tendance mais l'absence de données sur les années récentes met en doute la cohérence de l'estimation des ULR pour les dernières années observées.

3.3 Méthode des coûts moyens

Le coût moyen d'un sinistre est défini comme le rapport de la charge et du nombre de sinistres. La méthode du coût moyen consiste à déterminer le coût moyen ultime par année de survenance d'un côté et le nombre de sinistres ultime de l'autre.

Pour cela, on calcule à l'aide de la méthode Chain Ladder, une première cadence de développement qui sera appliquée au triangle des coûts moyens, et une seconde cadence qui sera appliquée à celui des nombres de sinistres. Cela nous permet de compléter les deux triangles et d'obtenir les montants ultimes du nombre de sinistres et des coûts moyens. La charge future finale sera calculée comme le produit des ultimes de chaque triangle.

Comme nous utilisons la méthode Chain Ladder sur le triangle des nombres et celui des coûts moyens, les hypothèses de la méthode doivent être vérifiées sur ces deux triangles. Nous allons utiliser les données des sinistres français de la classe « Exceptionnelle » sur lesquelles nous avons le détail des nombres de sinistres par année et période de développement.

- Triangle des nombres de sinistres

Le nombre de sinistres est répertorié dans le triangle suivant. Par soucis de confidentialité, les données du triangle ont été modifiées tout en préservant les ordres de grandeur.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1996	0,00031	0,00035	0,00036	0,00036	0,00036	0,00036	0,00036	0,00036	0,00036	0,00036	0,00036	0,00038	0,00038	0,00038	0,00038	0,00038	0,00038	0,00038	0,00038	0,00038	0,00038	0,00038	0,00038	0,00038	0,00038	0,00038
1997	0,00061	0,00071	0,00071	0,00071	0,00071	0,00071	0,00071	0,00071	0,00071	0,00071	0,00071	0,00071	0,00071	0,00071	0,00071	0,00071	0,00071	0,00071	0,00071	0,00071	0,00071	0,00071	0,00071	0,00071	0,00071	0,00071
1998	0,00050	0,00056	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059
1999	0,00035	0,00042	0,00042	0,00042	0,00042	0,00043	0,00043	0,00043	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045
2000	0,00050	0,00063	0,00064	0,00064	0,00064	0,00064	0,00064	0,00064	0,00066	0,00066	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068
2001	0,00042	0,00056	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00061	0,00061	0,00061	0,00061	0,00061	0,00061	0,00063	0,00063	0,00063	0,00063	0,00063	0,00063	0,00063	0,00063	0,00063
2002	0,00052	0,00064	0,00066	0,00066	0,00068	0,00069	0,00069	0,00069	0,00069	0,00069	0,00073	0,00073	0,00073	0,00073	0,00073	0,00073	0,00073	0,00073	0,00073	0,00073	0,00073	0,00073	0,00073	0,00073	0,00073	0,00073
2003	0,00031	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00047	0,00047	0,00047	0,00047	0,00047	0,00047	0,00047	0,00047	0,00047	0,00047
2004	0,00036	0,00047	0,00052	0,00054	0,00054	0,00054	0,00054	0,00056	0,00056	0,00056	0,00056	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057
2005	0,00045	0,00061	0,00063	0,00064	0,00066	0,00066	0,00066	0,00066	0,00066	0,00066	0,00066	0,00066	0,00066	0,00066	0,00066	0,00066	0,00066	0,00066	0,00066	0,00066	0,00066	0,00066	0,00066	0,00066	0,00066	0,00066
2006	0,00045	0,00056	0,00056	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057
2007	0,00036	0,00047	0,00049	0,00049	0,00052	0,00054	0,00054	0,00054	0,00054	0,00054	0,00054	0,00054	0,00054	0,00054	0,00054	0,00054	0,00054	0,00054	0,00054	0,00054	0,00054	0,00054	0,00054	0,00054	0,00054	0,00054
2008	0,00038	0,00052	0,00054	0,00056	0,00057	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059
2009	0,00042	0,00054	0,00061	0,00061	0,00064	0,00064	0,00066	0,00066	0,00066	0,00066	0,00066	0,00066	0,00066	0,00066	0,00066	0,00066	0,00066	0,00066	0,00066	0,00066	0,00066	0,00066	0,00066	0,00066	0,00066	0,00066
2010	0,00061	0,00075	0,00080	0,00080	0,00080	0,00078	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080
2011	0,00054	0,00071	0,00078	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080	0,00080
2012	0,00063	0,00087	0,00089	0,00089	0,00089	0,00089	0,00092	0,00092	0,00092	0,00092	0,00092	0,00092	0,00092	0,00092	0,00092	0,00092	0,00092	0,00092	0,00092	0,00092	0,00092	0,00092	0,00092	0,00092	0,00092	0,00092
2013	0,00045	0,00050	0,00057	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059	0,00059
2014	0,00083	0,00096	0,00096	0,00099	0,00101	0,00101	0,00101	0,00101	0,00101	0,00101	0,00101	0,00101	0,00101	0,00101	0,00101	0,00101	0,00101	0,00101	0,00101	0,00101	0,00101	0,00101	0,00101	0,00101	0,00101	0,00101
2015	0,00050	0,00054	0,00056	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057	0,00057
2016	0,00069	0,00080	0,00082	0,00082	0,00082	0,00082	0,00082	0,00082	0,00082	0,00082	0,00082	0,00082	0,00082	0,00082	0,00082	0,00082	0,00082	0,00082	0,00082	0,00082	0,00082	0,00082	0,00082	0,00082	0,00082	0,00082
2017	0,00043	0,00063	0,00064	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068
2018	0,00056	0,00063	0,00068	0,00069	0,00069	0,00069	0,00069	0,00069	0,00069	0,00069	0,00069	0,00069	0,00069	0,00069	0,00069	0,00069	0,00069	0,00069	0,00069	0,00069	0,00069	0,00069	0,00069	0,00069	0,00069	0,00069
2019	0,00042	0,00061	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068	0,00068
2020	0,00035	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045
2021	0,00035	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045	0,00045

Figure 3.27 : Triangle des nombres

Avant d'appliquer la méthode Chain Ladder, il faut vérifier les hypothèses :

- La deuxième hypothèse est vérifiée lorsque nous appliquons la sélection faite précédemment des 6 dernières années puisque la statistique vaut 3 et l'intervalle de confiance est de [1.8 ; 6.2].

- La première hypothèse de la méthode Chain Ladder est aussi vérifiée comme l'illustre le graphique suivant :

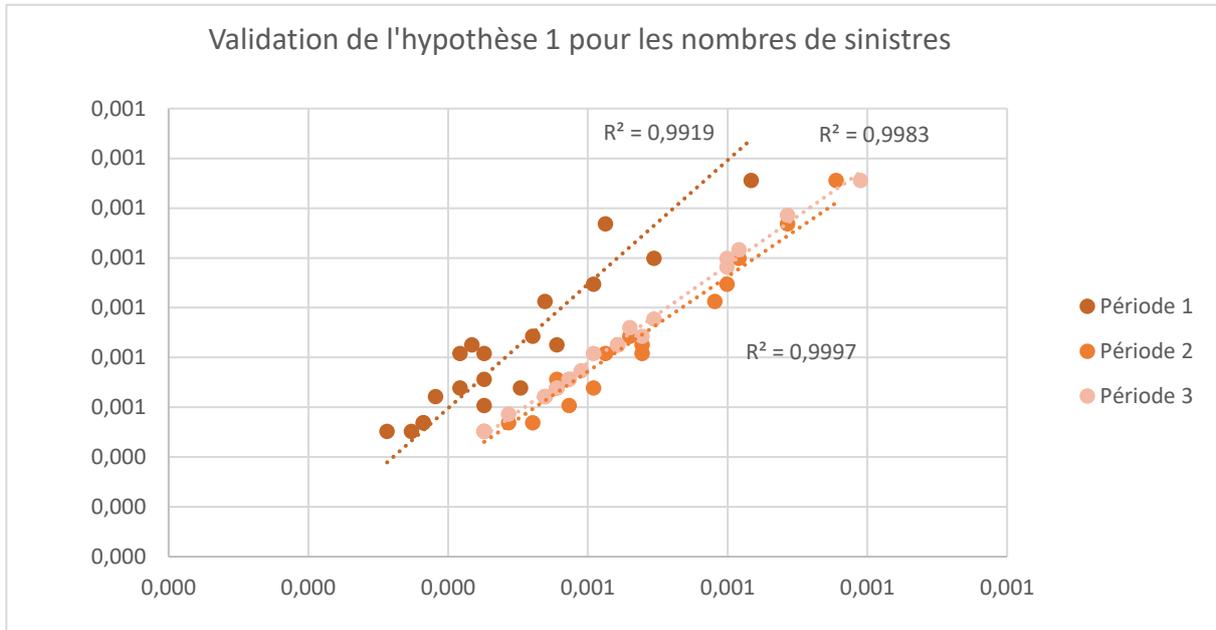


Figure 3.28 : Validation de l'hypothèse 1 de Chain Ladder pour le triangle contenant les nombres de sinistres

Après application de la méthode, le résultat est le suivant :

	Nombre final de sinistres	Nombre de tardifs
1996	0,0004	-
1997	0,0007	-
1998	0,0006	-
1999	0,0005	-
2000	0,0007	-
2001	0,0006	-
2002	0,0007	-
2003	0,0005	-
2004	0,0006	-
2005	0,0007	-
2006	0,0006	-
2007	0,0005	-
2008	0,0006	0,000004
2009	0,0007	0,000004
2010	0,0008	0,000005
2011	0,0008	0,000005
2012	0,0009	0,000006
2013	0,0006	0,000004
2014	0,0010	0,000006
2015	0,0006	0,000004
2016	0,0008	0,000005
2017	0,0007	0,000010
2018	0,0007	0,000014
2019	0,0007	0,000033
2020	0,0005	0,000048
2021	0,0005	0,000140
Total	0,0169	0,000288

Figure 3.29 : Résultats de la méthode Chain Ladder avec pondération appliquée au triangle de nombres

Il y a encore des sinistres tardifs attendus sur les années 2008 à 2021 pour un total de 0,000288.

- Triangle des coûts moyens

Le triangle des coûts moyens est obtenu en divisant pour chaque année et période de développement les charges par les nombres de sinistres. Voici le triangle des coûts moyens obtenu :

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1996	319 830	318 138	325 010	291 483	305 722	374 832	400 367	397 444	397 684	386 829	372 138	374 526	381 043	378 711	384 827	355 256	347 485	351 982	349 995	349 954	349 964	349 973	350 108	350 036	350 901	351 285
1997	517 457	498 494	447 828	463 145	475 119	498 587	523 485	567 043	578 721	583 623	587 080	610 057	623 371	663 206	691 312	703 678	704 267	699 486	734 108	730 925	729 468	727 883	708 706	708 607	708 528	
1998	413 391	407 149	418 452	428 435	445 977	457 586	427 891	461 566	438 840	466 909	473 824	479 450	507 597	506 939	497 392	487 227	488 431	493 101	501 845	500 629	538 674	537 658	523 827	524 509		
1999	348 885	301 283	324 497	306 093	277 511	350 568	299 756	368 352	331 996	326 520	327 523	336 584	353 708	362 489	371 408	372 658	377 967	372 466	372 590	373 533	375 368	375 675	375 734			
2000	505 940	489 881	462 040	480 562	487 039	496 027	481 063	478 015	484 642	510 437	519 431	508 079	484 396	476 078	470 422	465 708	477 204	479 494	478 502	561 024	575 368	568 625				
2001	355 979	398 095	426 740	444 930	446 186	452 063	471 290	498 774	485 052	481 872	487 935	503 681	517 096	513 534	506 721	475 015	472 441	467 043	466 160	466 756	468 847					
2002	520 138	462 451	441 347	469 253	454 632	432 325	426 953	397 979	384 836	391 569	439 212	444 271	444 671	458 246	390 909	391 224	392 845	391 898	400 647	399 734						
2003	384 393	345 904	356 644	337 091	286 580	277 381	273 061	276 824	278 836	280 241	275 760	275 748	275 779	275 116	276 386	276 301	276 157	276 494	273 701							
2004	927 153	856 800	830 120	798 238	832 861	849 182	815 891	838 937	852 001	868 086	872 653	848 817	822 522	838 244	851 777	818 513	814 919	808 902								
2005	805 683	802 675	747 943	682 129	583 115	565 143	597 023	598 290	591 944	598 037	584 540	572 977	586 818	584 672	573 803	564 166	568 474									
2006	721 635	682 422	642 777	614 129	633 034	655 066	634 799	625 976	599 053	614 772	625 839	628 418	630 750	657 615	649 685	643 968										
2007	681 311	773 470	764 908	776 854	818 865	847 377	894 214	845 439	829 752	808 506	770 921	798 176	751 040	751 004	751 360											
2008	636 972	588 336	550 600	554 011	570 674	571 629	586 565	547 077	524 650	483 825	459 785	464 972	451 090	452 108												
2009	761 087	733 047	743 258	769 206	718 054	709 205	715 052	711 453	757 107	829 150	805 302	846 907	721 390													
2010	740 901	757 756	793 166	768 359	752 608	706 328	678 841	715 351	707 850	722 962	761 375	774 709														
2011	809 242	747 711	664 239	681 420	674 034	649 669	626 472	599 385	599 897	587 407	593 164															
2012	1 153 086	828 083	834 171	804 427	781 361	743 493	776 709	776 876	792 643	725 868																
2013	1 028 875	994 418	720 812	697 169	671 749	674 531	687 650	721 320	800 496																	
2014	455 398	449 940	449 197	453 935	458 407	458 171	448 126	441 455																		
2015	442 271	471 793	445 768	440 220	456 639	410 572	405 247																			
2016	599 615	572 228	566 590	562 567	549 430	542 600																				
2017	570 251	524 805	513 965	527 828	538 734																					
2018	768 187	735 084	726 324	732 071																						
2019	707 499	694 840	722 117																							
2020	763 296	767 691																								
2021	910 070																									

Figure 3.30 : Triangle des coûts moyens

Nous vérifions les hypothèses Chain Ladder :

- La première hypothèse de la méthode Chain Ladder est vérifiée grâce au graphique suivant :

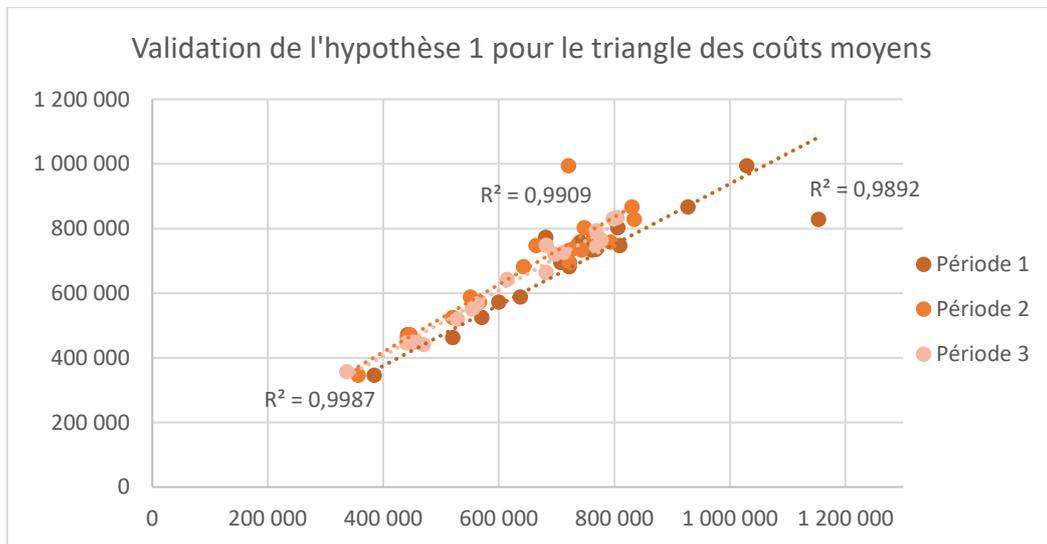


Figure 3.31 : Validation de l'hypothèse 1 de Chain Ladder pour le triangle des coûts moyens

- La deuxième hypothèse est vérifiée, sans avoir besoin de pondération puisque la statistique vaut 114 pour un intervalle de [111.9 ; 132.7].

Il est intéressant de regarder les résultats de la méthode en faisant les mêmes sélections que pour la méthode Chain Ladder (les hypothèses restent vérifiées).

L'application de la méthode nous permet d'estimer les coûts moyens ultimes. Voici le résultat global ainsi que le détail par année de survenance avec et sans pondération des coûts moyens :

	Charge Ultime
1996	351 265
1997	709 264
1998	525 443
1999	376 525
2000	560 431
2001	460 378
2002	399 799
2003	280 266
2004	838 997
2005	588 127
2006	667 676
2007	763 765
2008	456 030
2009	737 031
2010	776 958
2011	602 244
2012	738 327
2013	817 035
2014	450 923
2015	418 101
2016	559 050
2017	554 247
2018	728 213
2019	730 516
2020	748 971
2021	847 325
Total	15 686 908

Figure 3.33 : Résultats sans pondération

	Charge Ultime
1996	351 265
1997	709 264
1998	525 443
1999	376 525
2000	560 431
2001	460 378
2002	400 766
2003	284 536
2004	842 703
2005	589 811
2006	668 091
2007	765 752
2008	460 228
2009	743 728
2010	757 608
2011	595 314
2012	719 732
2013	786 000
2014	449 906
2015	418 468
2016	560 640
2017	539 814
2018	711 700
2019	718 750
2020	759 528
2021	870 230
Total	15 626 613

Figure 3.32 : Résultats avec pondération

- Calcul des provisions

Le montant ultime est obtenu en faisant le produit des coûts moyens ultimes avec le nombre final de sinistres par année de survenance. Le montant des IBNR est calculé comme la différence entre la charge ultime estimée et la charge connue (dernière diagonale du triangle de charges). Voici les résultats obtenus :

	Charge Ultime	Charge connue	Montant des IBNR
1996	134	134	-
1997	505	505	1
1998	310	310	1
1999	170	170	0
2000	380	385 -	6
2001	288	293 -	5
2002	292	292	0
2003	131	128	3
2004	481	464	17
2005	388	375	13
2006	383	369	14
2007	411	405	7
2008	271	267	4
2009	490	476	13
2010	625	619	6
2011	484	474	10
2012	684	668	16
2013	486	473	13
2014	457	445	12
2015	241	232	9
2016	459	443	16
2017	381	365	16
2018	516	495	21
2019	519	489	30
2020	374	347	27
2021	413	316	97
Total	10 274	9 938	336

Figure 3.34 : Résultats sans pondération du triangle des coûts moyens

	Charge Ultime	Charge connue	Montant des IBNR
1996	134	134	-
1997	505	505	1
1998	310	310	1
1999	170	170	0
2000	380	385 -	6
2001	288	293 -	5
2002	292	292	1
2003	133	128	5
2004	483	464	19
2005	389	375	14
2006	383	369	14
2007	412	405	8
2008	273	267	6
2009	494	476	18
2010	609	619 -	10
2011	479	474	5
2012	667	668 -	2
2013	467	473 -	6
2014	456	445	11
2015	241	232	9
2016	461	443	18
2017	371	365	6
2018	504	495	10
2019	511	489	22
2020	379	347	33
2021	425	316	108
Total	10 218	9 938	280

Figure 3.35 : Résultats avec pondération du triangle des coûts moyens

Les résultats du tableau de gauche sont supérieurs aux résultats obtenus par la méthode Chain Ladder pondérée (241 €) contrairement aux résultats du tableau de droite qui sont quant à eux très proches.

Conclusion sur la méthode du coût moyen

A partir d'un triangle de nombres et d'un triangle de coûts moyens, cette méthode permet d'estimer séparément les effets liés aux nombres et aux coûts. Sur chacun des deux triangles, il est nécessaire de satisfaire les hypothèses de la méthode utilisée pour le calcul de la cadence. Nous avons utilisé la méthode Chain Ladder car c'est la méthode la plus utilisée en pratique mais il n'existe aucune restriction quant au choix de la méthode.

Notons qu'il n'est pas nécessaire d'avoir des pondérations identiques entre les triangles de nombres et de coûts moyens du moment que les hypothèses sont vérifiées.

3.4 Méthodes factorielles

Les méthodes factorielles définissent les montants incrémentaux $S_{i,j}$ comme le produit :

- D'un facteur lié aux années de survenance x_i interprété comme le montant ultime attendu pour l'année i ,
- D'un facteur lié aux années de développement y_j interprété comme la proportion du montant final observée,
- D'un facteur lié aux années calendaires λ_{i+j} qui peut être interprété comme l'effet de l'inflation ou l'effet d'un changement dans la taille du portefeuille.

Ainsi, toutes les informations du triangle sont utilisées puisque les trois axes de lecture sont mobilisés. Dans sa forme générale, le modèle s'écrit :

$$S_{i,j} = x_i y_j \lambda_{i+j}$$

Cependant, sans contrainte supplémentaire, plusieurs solutions donnent le même résultat et il n'est donc pas possible d'identifier l'effet de chaque facteur.

3.4.1 Méthode de De Vylder

En 1982, De Vylder propose une résolution de ce problème en réécrivant les incréments comme le produit d'un effet ligne et d'un effet colonne. Les effets calendaires sont donc supposés constants ou retraités. Il ajoute également une contrainte pour rendre le modèle identifiable.

Le modèle se réécrit comme :

$$S_{i,j} = x_i y_j$$

sous la contrainte

$$\sum_{j=1}^n y_j = 1$$

La composante y_j est liée aux années de développement et représente la proportion de la charge ultime enregistrée l'année j . Il est supposé qu'à la fin de l'observation, l'ensemble des charges sont observées.

Les composantes sont estimées grâce à la méthode des moindres carrés, c'est-à-dire qu'ils minimisent la quantité :

$$Q = \sum_{i=1}^n \sum_{j=1}^{n-i} (S_{i,j} - x_i y_j)^2$$

Cependant, il n'y a pas de solution explicite pour x_i et y_j puisque :

Pour un i fixé :

$$\frac{\partial Q}{\partial x_i} = 0$$

$$\Leftrightarrow 2 * \sum_j (S_{i,j} - x_i y_j) * y_j = 0$$

$$\Leftrightarrow x_i \sum_j y_j^2 = \sum_j S_{i,j} * y_j$$

$$\Leftrightarrow x_i = \frac{\sum_j S_{i,j} * y_j}{\sum_j y_j^2}$$

Pour un j fixé :

$$\frac{\partial Q}{\partial y_j} = 0$$

$$\Leftrightarrow 2 * \sum_i (S_{i,j} - x_i y_j) * x_i = 0$$

$$\Leftrightarrow y_j \sum_i x_i^2 = \sum_i S_{i,j} * x_i$$

$$\Leftrightarrow y_j = \frac{\sum_i S_{i,j} * x_i}{\sum_i x_i^2}$$

Les solutions étant dépendantes l'une de l'autre, il est nécessaire de recourir à une méthode itérative ou d'utiliser un logiciel de calcul pour estimer les x_i et y_j .

En réalisant les itérations sur les données françaises de la classe « Large », nous obtenons les résultats suivants :

Année	Paiements	Provisions D/D	Charge D/D	PSAP	IBNR	Charge ultime
1996	118	2	120	2	-	120
1997	138	10	148	10	1	148
1998	199	16	215	23	7	222
1999	209	4	213	12	8	221
2000	201	1	202	9	8	210
2001	173	6	179	16	10	189
2002	175	3	178	10	7	185
2003	177	10	187	18	8	195
2004	96	12	108	19	7	115
2005	184	12	196	25	13	210
2006	154	13	166	25	12	178
2007	128	1	129	14	14	143
2008	212	18	230	38	20	250
2009	158	12	170	27	15	185
2010	179	39	218	53	14	232
2011	165	47	212	67	20	232
2012	220	57	277	74	17	294
2013	173	49	222	62	13	235
2014	134	102	236	117	15	251
2015	165	98	263	107	9	272
2016	127	128	255	135	7	262
2017	123	166	289	167	1	289
2018	78	218	296	220	2	298
2019	64	243	307	249	6	313
2020	23	240	264	262	21	285
2021	0	247	248	346	99	346
Total	3 773	1 753	5 526	2 107	354	5 880

Figure 3.36 : Résultats de la méthode de De Vylder sur les charges incrémentales du triangle de la classe « Large »

Le montant ultime calculé avec la méthode de De Vylder à partir du triangle de charges incrémental n'est pas significativement différent des montants ultimes proposés par la méthode de Chain Ladder (5 928 € pour la version sans poids et 6 035 € pour la version pondérée).

Le montant final des provisions est donc légèrement inférieur mais tout de même très proche des estimations fournies par Chain Ladder (2 107 € pour De Vylder contre 2 155 € avec Chain Ladder non pondéré et 2 262 € pour Chain Ladder pondéré).

Conclusion sur la méthode de De Vylder

La méthode proposée par De Vylder ne prend pas en compte les effets calendaires et ne permet pas de facilement calculer les estimateurs puisqu'il est nécessaire de procéder par itérations. Les résultats de cette méthode sur la classe « Large » des données françaises sont très proches de ceux obtenus avec Chain Ladder.

3.4.2 Méthode par séparation

Dans les années 70, comme l'illustre le graphique ci-dessous, l'inflation était élevée et instable notamment à la suite des chocs pétroliers de 1973 et 1979.

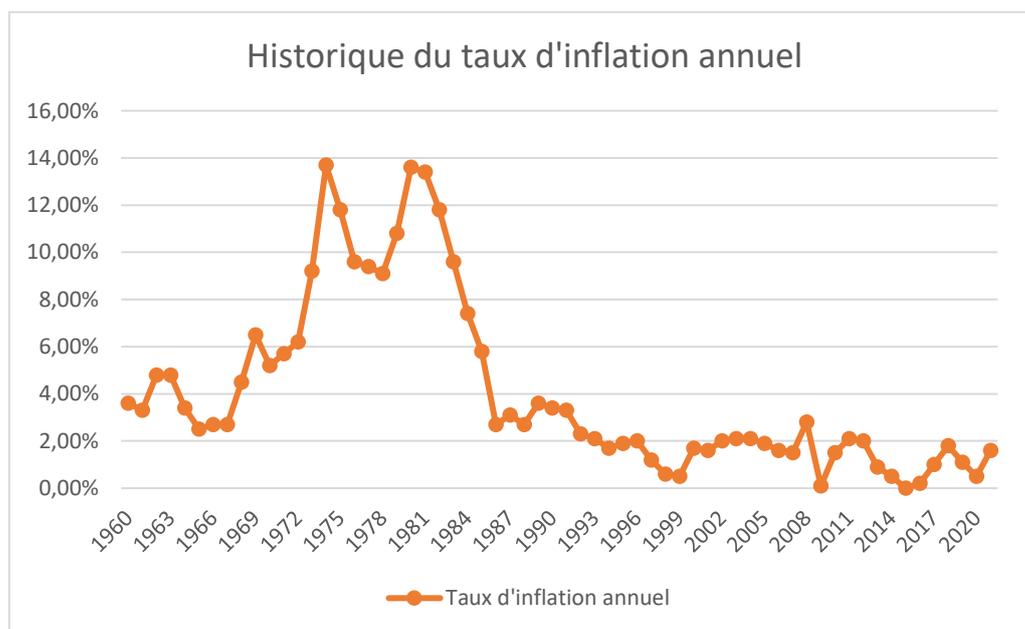


Figure 3.37 : Evolution de l'inflation au cours du temps

Sur des branches longues comme les branches de responsabilité, les sinistres sont réglés sur plusieurs années. La période d'incertitude est donc très élevée et la modélisation de l'inflation est un enjeu majeur. Afin de provisionner au mieux, il est donc nécessaire de pouvoir estimer l'inflation présente dans le triangle ainsi que l'inflation future.

Pour mieux appréhender l'inflation dans le provisionnement, Taylor développe en 1977 la méthode de séparation en se basant sur l'évaluation de l'évolution des sinistres proposée par Verbeek en 1972.

3.4.2.1 Présentation de la méthode

La méthode de séparation décompose les montants incrémentaux en un produit de facteurs exogènes liés à l'année calendaire (diagonales du triangle) et de facteurs liés à l'année de développement. Pour utiliser le modèle de Verbeek-Taylor, nous avons donc besoin d'un historique de sinistres où l'effet lié à la survenance est supprimé. Pour cela, il est possible de travailler sur le triangle des coûts moyens où chaque ligne du triangle est renormalisée par le nombre de sinistres. Cela permet de supprimer les effets liés à la taille du portefeuille et donc les effets liés à la survenance.

Nous nous intéressons dans cette partie au triangle des règlements. Les coûts moyens basés sur les paiements (ratios des paiements incrémentaux sur le nombre de sinistres) sont définis comme $\bar{X}_{i,j}^p$.

Ainsi, les montants des paiements renormalisés sont de la forme $\bar{X}_{i,j}^p = y_j \lambda_{i+j}$ avec y_j les facteurs constants liés à l'année de développement j tels que $\sum_{j=1}^n y_j = 1$ et λ_{i+j} les facteurs mesurant les influences exogènes liées à l'année de paiement $k = i + j$.

Le triangle utilisé dans la méthode est de la forme suivante :

Années de survenance	Années de développement						
	1	2	3	.	.	.	n
1	$y_1 \lambda_2$	$y_2 \lambda_3$	$y_3 \lambda_4$.	.	.	$y_n \lambda_n$
2	$y_1 \lambda_3$	$y_2 \lambda_4$	
3	$y_1 \lambda_4$	
.	
.	
.	
n	$y_1 \lambda_n$						

Le but est de séparer l'effet lié au développement représenté par les facteurs y_j et l'effet diagonal lié aux facteurs λ_k . Comme $\sum_{j=1}^n y_j = 1$, la somme de la dernière diagonale donne :

$$d_n = \lambda_n (y_1 + y_2 + \dots + y_{n-1} + y_n) = \lambda_n$$

Ainsi :

$$\hat{\lambda}_n = d_n$$

En sommant la diagonale précédente, on obtient :

$$d_{n-1} = \lambda_{n-1} (y_1 + \dots + y_{n-1}) = \lambda_{n-1} (1 - y_n)$$

Il est ici nécessaire d'avoir une estimation de y_k pour estimer λ_{n-1} . Un estimateur naturel de y_n est :

$$\hat{y}_n = v_n / \hat{\lambda}_n$$

où v_n est la somme de la colonne n .

En répétant cette procédure, nous arrivons au résultat général suivant :

$$\hat{\lambda}_k = \frac{d_k}{(1 - \hat{y}_n - \hat{y}_{n-1} - \dots - \hat{y}_{k+1})}$$

$$\hat{y}_j = \frac{v_j}{(\hat{\lambda}_j + \hat{\lambda}_{j+1} + \dots + \hat{\lambda}_n)}$$

En posant $\hat{m}_{i,j} = \hat{N}_i \hat{y}_j \hat{\lambda}_k$ où \hat{N}_i est le nombre final de sinistres pour l'année i (généralement estimé par la méthode Chain Ladder), le montant des réserves est donné par :

$$\hat{R}_i = \sum_{j > n+1-i} \hat{m}_{i,j} \text{ pour le montant par année et } \hat{R} = \sum_i \hat{R}_i \text{ pour le montant global}$$

Dans la formule du $\hat{m}_{i,j}$ nous remarquons la présence d'un $\hat{\lambda}_k$ où $k = i + j$. Nous connaissons $\{\hat{\lambda}_2, \dots, \hat{\lambda}_n\}$ mais comme $i \in \{1, \dots, n\}$ et $j \in \{1, \dots, n\}$, nous avons besoin d'extrapoler les valeurs des $\hat{\lambda}_k$ au-delà des valeurs calculées pour obtenir les valeurs $\{\hat{\lambda}_{n+1}, \dots, \hat{\lambda}_{2n}\}$ nécessaires pour calculer la partie inférieure du triangle.

Différentes méthodes peuvent être envisagées pour l'extrapolation comme mettre en œuvre un modèle stochastique d'inflation² (modèle de Vasicek par exemple), calibrer un modèle de régression ou faire des hypothèses à dire d'expert sur l'inflation future (prévisions FMI, BCE etc ... par exemple). Dans notre cas, nous allons regarder six hypothèses représentant des scénarios que nous jugeons possible. Voici les scénarios étudiés :

- VT1. Une régression linéaire
- VT2. Une régression avec un polynôme de degré 2
- VT3. Une inflation future supposée constante équivalente à la moyenne observée dans les données :

$$taux = \frac{1}{n-1} \sum_{m=1}^{n-1} \left(\frac{\hat{\lambda}_{m+1}}{\hat{\lambda}_m} - 1 \right) \approx 2,57\%$$

- VT4. Une inflation future supposée constante équivalente à la moyenne de l'indice des prix à la consommation (inflation $\approx 2\%$)
- VT5. Une inflation future basée sur les prévisions de la Banque de France et de la Banque Centrale Européenne
- VT6. Une inflation supposée constante équivalente à l'inflation observée sur l'année 2022 (inflation $\approx 11\%$ pour le secteur de l'automobile en France)

² Cf. par exemple le mémoire de C. Dossou (2022)

3.4.2.2 Application aux données

Nous allons appliquer la méthode par séparation sur les données des polices françaises pour la classe « Exceptionnelle ». Ce type de sinistres, dont le coût est très important, est particulièrement sensible à l'inflation car les délais de règlements sont étendus (attente d'une décision judiciaire par exemple). Le triangle des coûts moyens que nous allons étudier est celui des paiements renormalisés.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1996	10 992	50 322	8 164	14 183	106 669	51 517	54 292	11 531	6 922	6 190	9 520	17 071	9 070	3 329	5 073	10 323	484	4 094	3 258	2 573	273	1 554	3 331	19	39	85
1997	7 208	30 793	36 972	60 102	19 036	65 950	18 550	27 941	60 276	46 257	17 607	15 848	47 228	5 709	37 993	4 707	20 400	9 867	18 869	4 828	17 632	40 540	20 130	62 524	502	
1998	8 351	33 625	23 434	36 247	19 867	73 494	24 780	68 279	23 231	32 905	3 787	9 357	7 939	1 052	16 611	7 380	7 763	-3 535	4 531	-108	482	4 672	49 232	1		
1999	8 573	20 642	53 392	8 71	29 575	52 824	13 482	77 340	10 962	35 007	8 478	37 983	-16 860	968	20 138	12 552	234	2 553	22	5 611	17 211	3	0			
2000	20 139	42 176	21 478	26 545	61 883	45 542	15 786	32 135	20 502	17 321	26 726	16 342	5 185	5 210	1 964	4 295	38 080	3 197	4 250	1 600	11 850	23 596				
2001	6 881	38 943	46 489	29 960	43 410	27 926	52 514	26 435	38 728	17 206	2 216	1 919	19 188	9 238	16 852	77 673	16 976	-9 473	456	2 534	7 884					
2002	11 714	33 821	36 452	24 426	38 367	60 635	28 309	28 470	34 303	15 295	10 876	6 572	-2 799	18 080	9 150	200	2 726	8 730	11 599	495						
2003	13 026	35 457	28 110	27 873	54 427	38 433	28 258	533	9 778	28 449	5 841	444	138	9	-7 899	358	19	105	3 372							
2004	22 738	27 093	32 094	22 622	46 291	63 046	97 221	69 908	127 703	61 952	20 960	121 584	1 765	9 498	28 511	991	1 473	108								
2005	7 670	34 073	27 624	22 569	70 811	54 596	51 218	98 088	39 368	26 692	21 974	17 790	25 013	22 270	7 191	18 234	9 373									
2006	8 043	39 327	78 781	22 662	21 893	19 541	86 866	18 656	15 210	24 065	14 457	30 255	4 810	44 602	49 983	4 615										
2007	3 491	42 920	38 506	43 139	79 449	126 789	22 650	70 045	104 414	56 441	7 202	22 801	18 563	4 478	5 759											
2008	11 915	27 530	31 310	14 159	27 401	69 668	59 379	58 678	35 765	46 668	15 810	9 691	12 119	1 189												
2009	11 395	34 883	25 614	28 100	60 568	47 897	43 506	72 957	16 292	37 084	69 884	22 577	25 096													
2010	10 055	46 349	27 873	54 676	37 088	34 157	59 165	39 019	34 864	85 208	28 339	17 375														
2011	9 303	51 040	50 832	57 028	32 480	43 119	107 505	47 106	26 787	7 268	24 828															
2012	7 847	29 884	42 413	30 998	63 270	71 266	65 845	27 529	28 342	54 021																
2013	5 474	38 318	62 036	201 014	37 815	38 132	58 467	24 080	131 706																	
2014	10 685	57 301	43 564	48 878	38 785	34 119	18 914	38 047																		
2015	9 771	41 941	44 110	40 518	16 810	36 637	35 081																			
2016	8 060	45 617	33 336	48 796	56 606	29 038																				
2017	6 789	31 748	32 235	34 615	43 711																					
2018	5 647	35 591	22 635	50 899																						
2019	4 639	53 590	53 045																							
2020	4 564	22 892																								
2021	2 226																									

Figure 3.38 : Triangle incrémental des paiements renormalisés issu des données de la classe « Exceptionnelle »

Graphiquement, voici ce que donnerait l'extrapolation des lambdas sur nos données en fonction des méthodes utilisées :

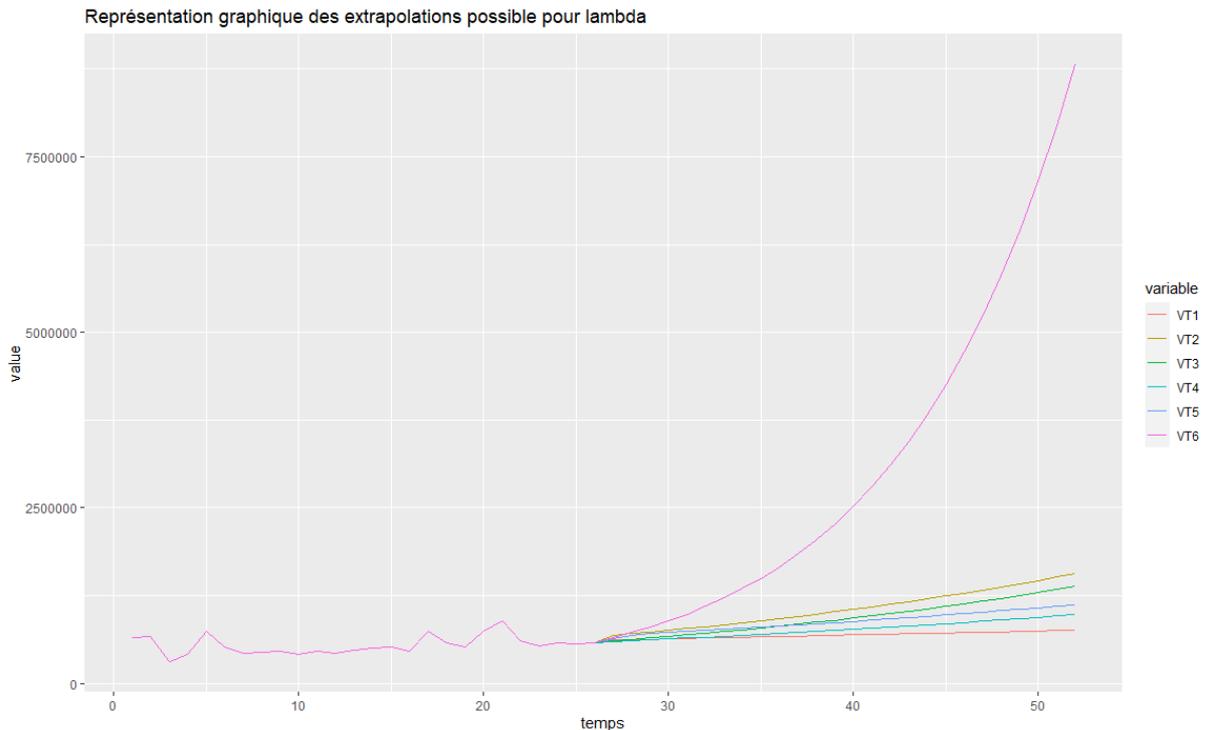


Figure 3.39 : Extrapolation des lambdas en fonction de la méthode choisie

Sur la période [0 ; 26] qui correspond au passé, nous retrouvons les $\hat{\lambda}_k$ calculés grâce à la méthode de Verbeek-Taylor. Sur la période [26 ; 52] qui correspond au futur, nous avons tracé les $\hat{\lambda}_k$ extrapolés en fonction des scénarios. Nous remarquons que l'hypothèse d'inflation future constante à 11% par an est très éloignée des autres. Nous n'allons pas garder cette hypothèse car même si elle est observée sur le passé récent, elle nous semble peu probable sur le long terme.

Pour rester proche des données, nous allons garder les extrapolations obtenues à partir des régressions (VT1 et VT2) ainsi que celle utilisant un taux constant égal au taux d'inflation moyen contenu dans les données (VT3).

Pour choisir entre les deux régressions nous comparons les R^2 ajustés. Nous ne remarquons pas de différence significative entre la régression linéaire et polynomiale (Régression linéaire $R^2_{adj} = 3\%$ vs Régression polynomiale $R^2_{adj} = 7\%$). La qualité d'ajustement pour les deux régressions est très faible à cause de fortes variations sur les années avant 2000 ainsi qu'entre 2011 et 2017 comme l'illustre le graphique :

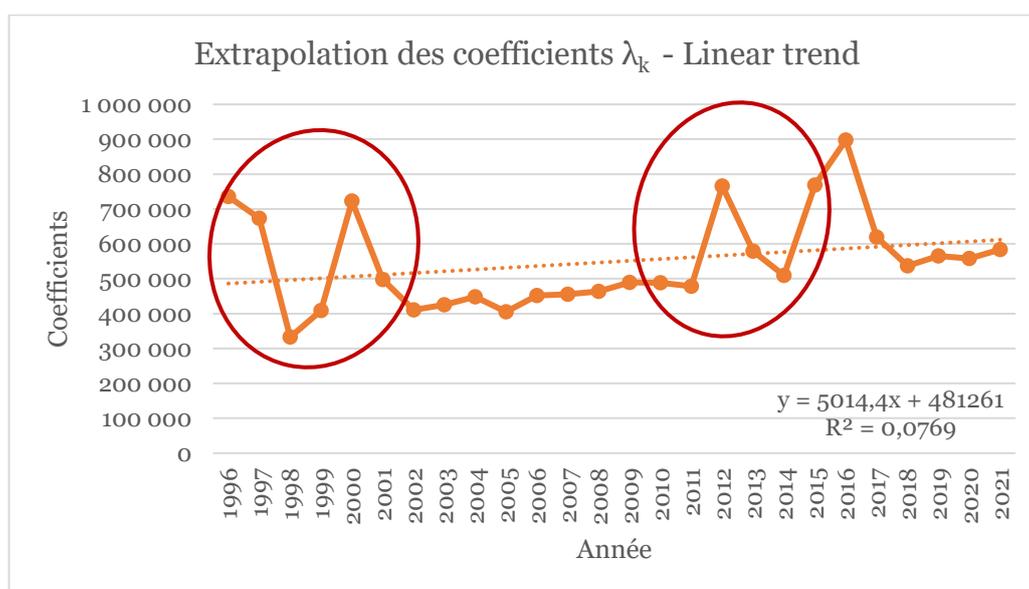


Figure 3.40 : Extrapolation des lambdas par une régression linéaire

L'historique étant assez profond, nous pouvons faire des sélections pour ne garder que certaines années et ainsi améliorer la qualité de l'extrapolation. Nous choisissons de ne pas retenir les années avant 2002 pour ajuster notre droite car elles sont trop volatiles. Nous allons également enlever les deux « pics » des années 2012 et 2016. Le R^2_{adj} obtenu est maintenant de 82% pour la régression linéaire de même que pour la régression polynomiale. Pour ajuster nos régressions, nous allons donc uniquement utiliser les 18 années les plus stables avec une régression linéaire.

L'équation utilisée pour l'extrapolation est la suivante :

$$y = 12663x + 368748$$

Une fois les facteurs $\{\hat{\lambda}_{n+1}, \dots, \hat{\lambda}_{2n}\}$ déterminées, nous pouvons appliquer la méthode. Voici par exemple le triangle obtenu en appliquant la régression linéaire (VT1) :

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1996	8 994	44 671	23 411	30 537	59 697	47 256	34 613	33 772	34 265	24 834	14 722	18 555	8 685	7 408	13 328	10 552	10 799	2 793	4 695	2 975	13 202	15 216	17 402	20 704	264	85
1997	8 229	22 073	28 743	54 011	41 151	39 009	35 819	35 538	30 960	27 695	14 864	18 891	9 160	7 401	13 046	16 886	8 162	2 458	7 087	3 473	9 111	13 202	18 303	20 449	276	88
1998	4 066	27 100	50 809	37 231	33 958	40 402	37 723	35 111	34 527	27 962	15 333	19 993	9 151	7 244	20 876	12 763	7 183	3 710	8 273	3 997	7 995	13 865	18 078	21 391	288	90
1999	4 992	47 933	35 044	30 733	35 182	42 514	34 085	35 810	34 859	28 467	15 960	19 994	8 957	11 592	15 779	11 231	10 843	4 331	5 710	2 080	8 314	13 715	18 910	22 321	294	92
2000	8 830	33 041	28 928	31 831	37 021	38 414	38 012	36 154	35 489	30 023	15 944	19 482	14 333	8 761	13 885	16 954	12 657	2 989	4 954	2 187	8 212	14 346	19 732	22 785	300	94
2001	6 087	27 274	29 961	33 495	33 451	42 839	38 377	36 808	37 429	29 993	15 606	21 175	10 833	7 710	20 960	19 790	8 735	2 593	5 210	2 160	8 590	14 970	20 142	23 249	306	96
2002	5 025	28 248	31 228	30 265	37 394	43 231	39 071	38 820	37 392	29 358	24 973	23 563	9 534	11 658	24 467	13 658	7 579	2 727	5 146	2 260	9 963	15 281	20 552	23 712	312	97
2003	5 204	29 726	28 287	33 751	37 663	44 033	41 207	38 782	36 600	46 979	18 975	20 736	14 391	13 386	16 885	11 850	7 971	2 694	5 383	3 358	9 149	15 592	20 962	24 175	318	99
2004	5 476	26 859	31 769	34 076	38 344	46 440	41 166	37 960	38 567	35 508	16 610	31 300	16 799	9 376	14 650	12 463	7 873	2 818	5 617	2 407	9 336	15 903	21 372	24 640	324	101
2005	4 948	29 953	32 074	34 692	40 440	46 394	40 294	60 744	44 266	31 247	25 073	36 538	11 593	8 135	15 408	12 310	8 235	2 940	5 734	2 456	9 522	16 214	21 783	25 104	330	103
2006	5 518	30 241	32 654	36 388	40 400	45 412	64 479	45 911	38 955	47 158	29 069	25 215	10 059	8 556	15 219	12 877	8 593	3 001	5 851	2 595	9 708	16 525	22 193	26 568	336	105
2007	5 574	30 788	34 439	35 552	39 545	72 668	48 734	40 493	58 802	55 060	20 199	21 878	10 579	8 451	15 920	13 437	8 772	3 063	5 967	2 554	8 904	16 826	22 603	26 032	342	107
2008	5 672	32 470	34 405	35 778	63 279	54 924	42 887	60 988	68 641	37 998	17 525	23 010	10 449	8 840	16 612	13 716	8 950	3 124	6 084	2 603	10 081	17 147	23 013	26 496	348	108
2009	5 982	32 438	33 677	57 252	47 827	48 334	64 738	71 192	47 371	32 968	18 432	22 727	10 930	9 224	16 957	13 995	9 129	3 185	6 201	2 652	10 267	17 458	23 423	26 959	354	110
2010	5 976	31 752	53 889	43 272	48 089	72 959	75 569	49 131	41 100	34 674	18 206	23 774	11 495	9 416	17 392	14 274	9 308	3 246	6 317	2 701	10 453	17 769	23 833	27 423	360	112
2011	5 949	50 809	40 739	38 080	63 533	85 167	59 152	42 608	43 227	24 248	19 044	24 807	11 642	9 607	17 447	14 553	9 486	3 907	6 434	2 750	10 640	18 081	24 243	27 887	366	114
2012	9 360	38 402	35 843	57 481	74 163	58 776	45 249	44 834	42 696	35 825	19 872	25 322	11 879	9 799	17 992	14 833	9 665	3 368	6 551	2 799	10 826	18 392	24 653	28 351	372	116
2013	7 075	33 794	54 105	67 099	51 182	50 995	47 591	44 283	44 662	37 382	20 285	25 838	12 116	9 991	18 338	15 112	9 843	3 429	6 668	2 848	11 012	18 703	25 063	28 815	378	118
2014	6 226	51 012	63 158	46 307	44 407	53 635	47 066	46 322	46 603	38 159	20 698	26 353	12 353	10 182	18 683	15 391	10 022	3 490	6 784	2 897	11 198	19 014	26 473	29 279	384	119
2015	9 398	59 548	43 597	40 177	46 705	59 975	49 170	48 335	47 571	38 306	21 111	26 869	12 590	10 374	19 028	15 670	10 200	3 551	6 901	2 946	11 385	19 325	25 883	29 743	390	121
2016	10 970	41 095	37 817	42 256	46 131	55 415	51 307	49 340	48 540	39 713	21 523	27 384	12 827	10 566	19 373	15 950	10 379	3 612	7 018	2 995	11 571	19 636	26 293	30 206	396	123
2017	7 571	35 655	39 775	41 737	48 255	57 823	52 373	50 344	49 508	40 490	21 936	27 900	13 064	10 757	19 718	16 229	10 558	3 674	7 135	3 044	11 757	19 947	26 703	30 670	402	125
2018	6 559	37 501	39 286	43 669	50 352	59 025	53 440	51 349	50 477	41 267	22 249	28 415	13 301	10 949	20 064	16 508	10 736	3 735	7 251	3 093	11 943	20 258	27 113	31 134	408	127
2019	6 909	37 049	41 095	45 555	51 399	60 227	54 596	52 353	51 445	42 043	22 762	28 931	13 538	11 441	20 409	16 787	10 915	3 796	7 268	3 147	12 130	20 569	27 523	31 598	414	129
2020	6 824	38 746	42 881	46 503	52 445	61 428	55 572	53 358	52 414	42 820	23 175	29 447	13 775	11 332	20 754	17 067	11 093	3 857	7 485	3 191	12 316	20 880	27 933	32 062	420	130
2021	7 138	40 429	43 772	47 450	53 491	62 630	56 638	54 362	53 382	43 597	23 588	29 962	14 012	11 524	21 099	17 346	11 272	3 918	7 601	3 240	15 502	21 191	28 343	32 526	426	132

Figure 3.41 : Triangle incrémental obtenu grâce à la méthode par séparation – Extrapolation linéaire

L'ensemble du triangle est réévalué par la méthode et la première étape est de vérifier que les estimations sont cohérentes, notamment sur les valeurs observées. En effet, comme nous supposons que les y_j ne sont pas reliés aux années de survenance, il ne faut donc pas qu'il y ait des années significativement différentes des autres pour ne pas avoir d'inconsistance. Ainsi, si pour toutes les années le montant des incréments augmente entre la période 1 et 2 mais que pour certaines, il décroît fortement, en fonction des proportions, la méthode va reproduire ce motif sur toutes les années.

Une fois le triangle incrémental validé, nous pouvons calculer les provisions à partir des montants $\hat{m}_{i,j} = \hat{N}_i \hat{y}_j \hat{\lambda}_k$ où :

- \hat{N}_i est le nombre ultime de sinistres par année calculé à partir de la méthode Chain Ladder (calculé dans la section 3.3),
- \hat{y}_j est la cadence de développement calculée à partir de la méthode par séparation,
- $\hat{\lambda}_k$ est l'impact des effets calendaires calculés à partir de la méthode par séparation et extrapolés à l'aide d'une régression linéaire.

Voici les résultats obtenus pour les deux extrapolations retenues :

	PSAP CM	Coût Moyen ultime	Nombre ultime	PSAP	IBNR
1996	-	350 077	0,0003821	-	0
1997	88	697 558	0,0007122	0	8
1998	379	445 460	0,0005906	0	47
1999	22 707	385 094	0,0004516	10	4
2000	42 911	488 712	0,0006774	29	54
2001	58 762	522 898	0,0006253	37	34
2002	68 918	449 471	0,0007296	50	36
2003	72 656	340 085	0,0004690	34	31
2004	79 701	835 262	0,0005732	46	15
2005	84 186	638 740	0,0006601	56	46
2006	94 385	581 493	0,0005732	54	36
2007	109 606	756 254	0,0005385	59	3
2008	128 282	546 964	0,0005943	76	58
2009	139 915	632 696	0,0006642	93	56
2010	153 921	635 120	0,0008040	124	108
2011	181 565	636 039	0,0008040	146	37
2012	204 790	623 604	0,0009263	190	91
2013	245 938	839 294	0,0005943	146	26
2014	297 084	579 622	0,0010137	301	143
2015	350 931	574 410	0,0005768	202	99
2016	408 754	628 841	0,0008215	336	74
2017	474 160	621 019	0,0006878	326	62
2018	533 296	645 740	0,0007088	378	37
2019	588 682	694 737	0,0007108	418	5
2020	642 340	667 159	0,0004996	321	13
2021	694 437	696 022	0,0004879	339	23
Total	5 678 394	15 512 371	0,016877	3 771	246

Figure 3.42 : Résultats de la méthode par séparation – Extrapolation linéaire

	PSAP CM	Coût Moyen ultime	PSAP	IBNR
1996	-	350 077	-	0
1997	87	697 557	0	8
1998	373	445 454	0	47
1999	22 324	384 711	10	4
2000	42 296	488 097	29	55
2001	58 100	522 236	36	33
2002	68 406	448 959	50	36
2003	72 485	339 915	34	31
2004	79 872	835 433	46	15
2005	84 818	639 372	56	47
2006	95 449	582 556	55	35
2007	111 114	757 762	60	3
2008	130 303	548 985	77	59
2009	142 724	635 506	95	54
2010	157 610	638 809	127	105
2011	186 052	640 526	150	41
2012	210 375	629 189	195	85
2013	252 537	845 892	150	30
2014	304 849	587 387	309	151
2015	360 209	583 689	208	104
2016	419 917	640 003	345	83
2017	487 566	634 425	335	71
2018	549 612	662 057	390	26
2019	608 542	714 596	433	19
2020	666 357	691 176	333	1
2021	723 251	724 836	353	37
Total	5 835 226	15 669 203	3 874	349

Figure 3.43 : Résultats de la méthode par séparation – Extrapolation moyenne passée

Les charges ultimes sont ici très proches de celles calculées à partir de la méthode du coût moyen (15 626 613 € pour la version avec pondération et 15 686 908 € pour la version sans pondération). Les montants d'IBNR sont également proches de ceux obtenus en appliquant les différentes méthodes (Chain Ladder : 402 €, Bornhuetter-Ferguson : 375 €, coût moyen : 336 € et coût moyen pondéré : 280 €). L'extrapolation des lambdas à partir de la moyenne passée donne un montant d'IBNR plus important que l'extrapolation linéaire (349 € vs 246 €).

Conclusion sur la méthode de séparation

Si la modélisation de l'inflation était moins importante dans les dernières décennies du fait d'une inflation stable, dans le contexte actuel d'inflation croissante, il est primordial de la modéliser afin de ne pas sous-estimer les charges futures. Les méthodes basées sur les cadences projettent mécaniquement l'inflation passée sans la modéliser ce qui n'est pas le cas de la méthode de Verbeek-Taylor. Ainsi, en cas d'augmentation soudaine de l'inflation, la méthode par séparation permet d'appliquer les nouvelles tendances directement dans le calcul des provisions alors que pour les autres méthodes, il est nécessaire de calculer à posteriori l'impact d'une sur inflation ou bien d'utiliser des triangles déflatés pour ensuite appliquer l'inflation réelle.

Nous avons choisi une tendance linéaire basée sur les années les plus stables pour extrapoler nos lambdas mais il aurait été possible d'extrapoler indépendamment des données du triangle, par exemple en suivant les prévisions de la Banque de France (pour prendre en compte l'évolution future possible de l'inflation).

4. Evaluation de l'incertitude dans l'estimation des provisions techniques

Les méthodes déterministes permettent d'estimer le montant « Best Estimate » des provisions mais ne permettent pas de quantifier l'incertitude associée. En particulier, elles ne fournissent pas de loi de probabilité de la provision calculée (pas de calcul de volatilité ou de VaR possible). Pour pallier ce problème, des méthodes stochastiques ont été développées (England et Verrall (2002)).

Comme le développement des sinistres est estimé, il convient de savoir l'erreur qui est commise lors de l'utilisation de ces estimations. Une mesure qui permet de quantifier l'erreur commise est la MSEP (Mean Square Error of Prediction). Elle permet de calculer, selon une certaine métrique, l'écart entre une variable aléatoire et son prédicteur.

En notant θ la variable aléatoire dont on souhaite prédire l'issue (par exemple la somme des paiements futurs associés aux sinistres survenus) et $\hat{\theta}$ son prédicteur, la MSEP conditionnelle se définit comme :

$$MSEP(\hat{\theta}) = E[(\theta - \hat{\theta})^2 | D]$$

où D représente l'information disponible à la date de l'estimation, c'est-à-dire les données observées (partie supérieure du triangle).

La MSEP inconditionnelle s'écrit :

$$MSEP(\hat{\theta}) = Var(\theta) + E[(\hat{\theta} - E[\theta])^2] - 2E[(\hat{\theta} - E[\theta])(E[\theta|D] - E[\theta])]$$

Dans le cas où $\hat{\theta}$ est un estimateur sans biais :

$$MSEP(\hat{\theta}) = Var(\theta) + Var(\hat{\theta}) - 2Cov(\hat{\theta}, E[\theta|D])$$

Nous allons chercher à évaluer cette erreur pour les méthodes présentées précédemment.

4.1 Extension stochastique de la méthode Chain Ladder

Thomas Mack a développé des modèles stochastiques permettant d'évaluer les erreurs de prédiction des provisions calculées avec la méthode Chain Ladder au moyen de formules fermées. La version stochastique de Chain Ladder donne le même montant de provision estimé qu'avec la méthode déterministe et inclue l'erreur de prédiction.

4.1.1 Chain Ladder stochastique (Mack (1993))

4.1.1.1 Présentation de la méthode

La méthode proposée par Mack en 1993 est la première méthode permettant de mesurer l'incertitude associée à la prédiction du montant des provisions grâce à une formule fermée. L'idée est de considérer les observations du triangle supérieur comme des réalisations de variables aléatoires d'espérance $f_{i,j}$ inconnue estimée par \hat{f}_j . Les estimateurs \hat{f}_j sont sans biais et indépendants.

Pour pouvoir utiliser la méthode, il est nécessaire de vérifier trois hypothèses :

(H1) Les années de survenance sont indépendantes

$$\{C_{i,1}, \dots, C_{i,n}\} \text{ est indépendant de } \{C_{j,1}, \dots, C_{j,n}\} \text{ pour } i \neq j$$

(H2) L'espérance conditionnelle est sans biais

$$E(C_{i,j+1} | C_{i,1}, \dots, C_{i,j}) = C_{i,j} \cdot f_j \text{ pour } 1 \leq i \leq n \text{ et } 1 \leq j \leq n-1$$

(H3) La variance conditionnelle est donnée par

$$Var(C_{i,j+1} | C_{i,1}, \dots, C_{i,j}) = C_{i,j} \cdot \sigma_j^2 \text{ pour } 1 \leq i \leq n \text{ et } 1 \leq j \leq n-1$$

L'estimateur du paramètre σ_j^2 est :

$$\hat{\sigma}_j^2 = \frac{1}{n-j-1} \sum_{i=1}^{n-j} C_{i,j} \left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j \right)^2 \text{ pour } 1 \leq j \leq n-2$$

Pour $\hat{\sigma}_{n-1}^2$ il y a deux possibilités :

- Le facteur $\hat{f}_{n-1} = 1$ et dans ce cas $\hat{\sigma}_{n-1}^2 = 0$
- $\hat{\sigma}_{n-1}^2 = \min\left(\frac{\hat{\sigma}_{n-2}^4}{\hat{\sigma}_{n-3}^2}, \min(\hat{\sigma}_{n-3}^2, \hat{\sigma}_{n-2}^2)\right)$

Les deux premières hypothèses sont les hypothèses du modèle classique de Chain Ladder et la dernière hypothèse vient du fait que :

$$\begin{aligned} MSEP(\hat{R}_i) &= E\left((\hat{R}_i - R_i)^2 | D\right) = E\left((\hat{C}_{i,n} - C_{i,n})^2 | D\right) \\ &= Var(C_{i,n} | D) + (E(C_{i,n} | D) - \hat{C}_{i,n})^2 \end{aligned}$$

La MSEP peut donc se décomposer en deux termes :

- L'erreur de processus qui ne peut pas être réduite et qui correspond à l'écart entre la valeur espérée et la valeur réelle de la charge ultime (il s'agit de la variabilité du modèle stochastique) ;
- L'erreur d'estimation qui correspond à l'écart entre la valeur espérée de la charge ultime et son estimateur.

L'estimateur de la MSEP pour une année de survenance i est :

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

L'erreur de prédiction globale est :

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

Cette erreur globale n'est pas identique à la somme des erreurs par année car les \hat{R}_i sont corrélés aux estimateurs $\hat{\sigma}_j^2$ et \hat{f}_j .

4.1.1.2 Application aux données

Nous n'allons appliquer cette méthode qu'aux données des polices françaises de la classe « Large » car c'est la seule méthode où les deux premières hypothèses étaient vérifiées dans la version déterministe. Pour vérifier la dernière hypothèse, il faut que le graphique des résidus ne fasse apparaitre aucune structure non aléatoire.

Les résidus sont donnés par la formule :

$$r_{i,j} = \sqrt{C_{i,j}} \frac{\hat{f}_{i,j} - \hat{f}_j}{\hat{\sigma}_j}$$

Voici l'illustration de nos résidus pour les trois premières périodes :

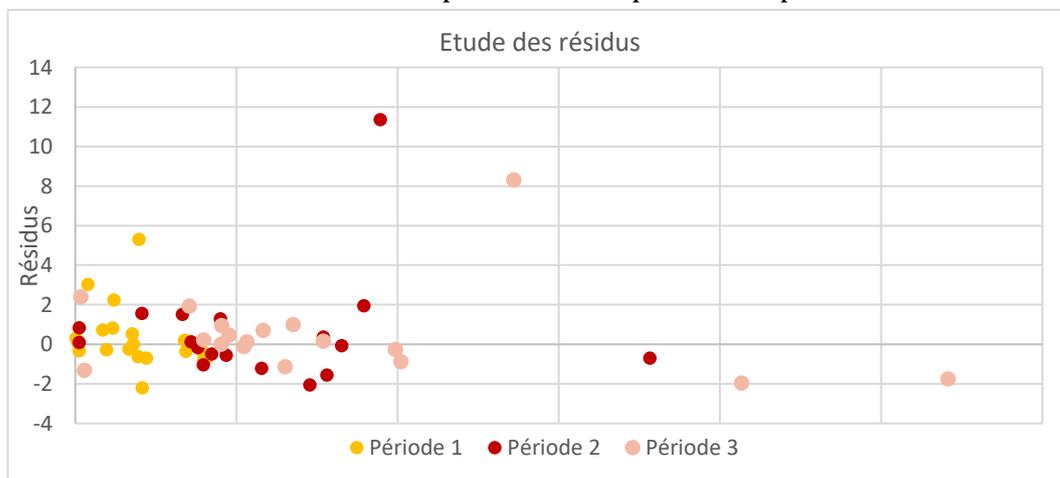


Figure 4.1 : Etude graphique des résidus

Le graphique des résidus est un bon indicateur lorsque beaucoup de données sont disponibles. La dernière hypothèse est ici vérifiée puisque les résidus ne laissent apparaître aucune structure non aléatoire. De plus, le modèle est bien spécifié puisque l'espérance des résidus est proche de 0 malgré la présence de quelques points atypiques.

Après application de la méthode, voici les résultats obtenus :

	Charge Ultime	Provisions D/D	Montant des IBNR	MSE	RMSE	% d'erreur
1996	120	2	-	-	-	0%
1997	148	10	1	6	2,3	23%
1998	221	16	6	40	6,3	29%
1999	221	4	7	46	6,8	58%
2000	211	1	9	43	6,6	67%
2001	189	6	10	56	7,5	46%
2002	187	3	9	63	7,9	67%
2003	197	10	10	75	8,6	43%
2004	114	12	6	40	6,3	35%
2005	211	12	15	98	9,9	37%
2006	180	13	14	92	9,6	36%
2007	141	1	12	78	8,8	70%
2008	252	18	22	169	13,0	33%
2009	186	12	16	126	11,2	40%
2010	238	39	20	195	14,0	24%
2011	232	47	20	226	15,0	22%
2012	300	57	23	326	18,1	23%
2013	240	49	17	316	17,8	27%
2014	253	102	17	386	19,6	17%
2015	274	98	11	487	22,1	20%
2016	263	128	7	564	23,7	17%
2017	292	166	3	741	27,2	16%
2018	301	218	5	924	30,4	14%
2019	318	243	11	1 111	33,3	13%
2020	288	240	24	1 212	34,8	13%
2021	352	247	104	2 374	48,7	14%
Total	5 928	1 753	402	31 844	178,4	8,3%

Figure 4.2 : Résultats du Chain Ladder stochastique

L'extension stochastique proposée par Mack donne la même estimation des provisions que la méthode déterministe et permet d'observer l'erreur commise. Nous pouvons calculer le rapport entre le montant des erreurs et le montant des provisions défini comme :

$$\frac{RMSE}{PSAP D/D + IBNR}$$

Nous trouvons que l'erreur représente environ 8% du montant total des provisions.

Conclusion de la méthode Chain Ladder stochastique de Mack 1993

Nous avons vu que pour enlever les tendances sur certaines années, il peut être nécessaire d'utiliser des poids afin d'exclure certaines tendances. Il est nécessaire d'utiliser cette technique si l'évolution d'un facteur est trop distincte de l'évolution générale et/ou ne serait pas amenée à se reproduire. La formule de Mack 1993 ne permet pas de prendre en compte ces poids ni dans le calcul des provisions ni dans celui de la MSE.

4.1.2 Chain Ladder stochastique pondéré (Mack (1999))

4.1.2.1 Présentation de la méthode

L'extension proposée par Mack en 1999 consiste à inclure des poids pour permettre de valider toutes les hypothèses de la méthode Chain Ladder et répliquer les estimations de la méthode Chain Ladder pondérée présentée dans la section 3.1.2. Les facteurs de développement vont prendre en compte cette modification :

$$\forall j \in \{1, \dots, n-1\}, \quad \hat{f}_j = \frac{\sum_{i=1}^{n-j} w_{i,j} \cdot C_{i,j}^\alpha \cdot f_{i,j}}{\sum_{i=1}^{n-j} w_{i,j} \cdot C_{i,j}^\alpha}$$

Cette extension modifie légèrement la dernière hypothèse précédente :

(H3b) : La variance conditionnelle est donnée par

$$\text{Var}\left(\frac{C_{i,j+1}}{C_{i,j}} \mid C_{i,1}, \dots, C_{i,j}\right) = \frac{\sigma_j^2}{w_{i,j} \cdot C_{i,j}^\alpha} \text{ pour } 1 \leq i \leq n \text{ et } 1 \leq j \leq n-1$$

L'estimation du paramètre $\widehat{\sigma}_j^2$ est elle aussi modifiée pour incorporer les poids :

$$\hat{\sigma}_j^2 = \frac{1}{n-j-1} \sum_{i=1}^{n-j} w_{i,j} \cdot C_{i,j}^\alpha \left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j\right)^2 \text{ pour } 1 \leq j \leq n-2$$

La nouvelle erreur de prédiction est donnée par la formule :

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

Avec

$$\widehat{MSEP}(\hat{f}_j) = \hat{\sigma}_j^2 / \sum_{k=1}^{n-j} w_{k,j} \cdot C_{k,j}^\alpha$$

et

$$\widehat{MSEP}\left(\frac{C_{i,j+1}}{C_{i,j}}\right) = \frac{\hat{\sigma}_j^2}{w_{i,j} \cdot C_{i,j}^\alpha}$$

4.1.2.2 Application aux données

Nous commençons par appliquer la version pondérée aux données précédentes :

	Charge Ultime	Provisions D/D	Montant des IBNR	MSE	RMSE	% d'erreur
Total	6 035	1 753	509	48 754	220,8	9,8%

Figure 4.3 : Résultats du Chain Ladder stochastique pour les données françaises de la classe "Large"

L'erreur commise en appliquant le Chain Ladder pondéré est plus élevée qu'avec l'application du Chain Ladder classique (9,8% vs 8,3%).

Ceci s'explique par le fait que si les hypothèses sont toutes vérifiées, ajouter des pondérations (*i.e.* sélectionner arbitrairement certaines années) enlève de l'information. L'incertitude est donc plus élevée.

Concernant les autres jeux de données, nous obtenons les résultats suivants :

	Charge Ultime	Provisions D/D	Montant des IBNR	MSE	RMSE	% d'erreur
1996	84,4	0	-	-	-	0%
1997	69,5	1	0	0	0,00	0%
1998	79,6	2	0	0	0,44	-22%
1999	84,0	0	0	0	0,46	-266%
2000	81,2	3	0	0	0,55	-19%
2001	86,2	0	0	1	1,08	404%
2002	85,5	0	0	1	1,15	-575%
2003	77,8	0	0	2	1,47	-451%
2004	66,3	0	0	2	1,35	-349%
2005	64,1	0	0	2	1,34	-545%
2006	62,1	0	0	2	1,33	-566%
2007	62,7	0	0	2	1,35	-720%
2008	62,5	1	1	2	1,36	714%
2009	65,0	1	1	2	1,39	-1630%
2010	69,4	1	1	2	1,46	2330%
2011	69,2	1	1	2	1,46	571%
2012	69,7	2	1	2	1,47	147%
2013	66,5	1	1	2	1,43	-595%
2014	63,0	4	1	2	1,43	64%
2015	63,7	4	2	2	1,46	68%
2016	66,5	5	2	2	1,53	55%
2017	69,0	8	3	3	1,63	32%
2018	68,3	11	4	3	1,65	23%
2019	70,8	19	7	3	1,79	15%
2020	58,0	24	8	3	1,63	10%
2021	70,6	43	4	5	2,28	5%
Total	1 835	118	28	180	13,42	15,0%

Figure 4.4 : Résultats du Chain Ladder stochastique pour les données françaises de la classe "Attritional"

	Charge Ultime	Provisions D/D	Montant des IBNR	MSE	RMSE	% d'erreur
1996	134	0	-	-	-	0%
1997	505	8	1	0	0,01	0%
1998	310	47	0	0	0,57	1%
1999	170	6	0	0	0,44	7%
2000	379	83	7	27	5,22	7%
2001	287	3	6	22	4,69	-140%
2002	291	14	0	128	11,33	84%
2003	133	3	5	313	17,70	224%
2004	483	31	19	1 412	37,57	75%
2005	389	9	14	1 089	33,00	141%
2006	383	90	14	1 073	32,76	32%
2007	413	56	8	1 228	35,04	54%
2008	273	18	6	794	28,17	118%
2009	492	149	16	1 698	41,20	25%
2010	606	232	13	4 081	63,89	29%
2011	476	109	2	3 116	55,82	50%
2012	667	280	1	5 549	74,49	27%
2013	465	120	7	5 159	71,83	64%
2014	452	158	7	5 568	74,62	45%
2015	238	103	6	2 877	53,64	49%
2016	455	262	12	6 402	80,01	29%
2017	368	264	4	5 251	72,46	27%
2018	500	415	6	7 775	88,17	21%
2019	507	414	18	8 137	90,21	21%
2020	379	334	32	6 603	81,26	22%
2021	421	315	105	11 190	105,78	25%
Total	10 179	3 525	241	250 443	500,44	13,3%

Figure 4.5 : Résultats du Chain Ladder stochastique pondéré sur les données françaises de la classe "Exceptionnelle"

	Charge Ultime	Provisions D/D	Montant des IBNR		MSE	RMSE	% d'erreur
2002	467	14	-		-	-	0%
2003	1 121	527	-	3	20	4,45	1%
2004	1 076	449	-	13	172	13,10	3%
2005	642	320	-	9	165	12,84	4%
2006	453	194	-	5	524	22,88	12%
2007	925	622	-	21	1 559	39,49	7%
2008	834	416	-	23	2 848	53,37	14%
2009	659	291	-	27	2 879	53,66	20%
2010	894	612	-	37	4 225	65,00	11%
2011	1 039	593	-	2	14 598	120,82	20%
2012	1 460	887	-	7	23 929	154,69	17%
2013	1 698	625	-	5	34 047	184,52	30%
2014	2 195	1 184	-	10	62 494	249,99	21%
2015	2 181	1 527	-	12	79 513	281,98	18%
2016	1 733	1 580	-	14	69 119	262,90	17%
2017	566	511	-	9	27 942	167,16	32%
2018	60	55	-	1	3 401	58,32	105%
2019	29	28	-	1	2 286	47,82	163%
2020	32	25	-	7	3 118	55,84	174%
2021	710	214	-	496	198 600	445,65	63%
Total	18 774	10 675		388	1 020 910	1 010,40	9,1%

Figure 4.6 : Résultats du Chain Ladder stochastique pondéré sur les données d'acceptations des polices anglaises

Les erreurs se situent toutes aux alentours de 10% ce qui n'est pas une valeur jugée aberrante pour des sinistres RC automobile.

Conclusion de la méthode du Chain Ladder stochastique pondéré de Mack 1999

Le modèle de Mack permet d'obtenir très simplement une évaluation du montant des provisions et de calculer la variance associée grâce à une formule fermée. Cependant, cette méthode ne permet pas d'avoir une description complète de la distribution des provisions.

De plus, comme cette méthode est multiplicative, plus il y a d'années de développement plus les erreurs peuvent être importantes. En effet, le montant à l'ultime est estimé à partir de la multiplication entre la dernière observation et le produit des facteurs de développement. Si un facteur est faux, l'erreur est répercutée sur tout le processus de développement.

4.2 Méthode de Bornhuetter-Ferguson stochastique³

4.2.1 Modèle de Mack (2008)

Après la méthode Chain Ladder, la méthode Bornhuetter-Ferguson est probablement la méthode déterministe la plus utilisée en pratique. La première version stochastique de Chain Ladder est parue en 1993 mais jusqu'en 2008, il n'existait rien d'équivalent pour la méthode Bornhuetter-Ferguson. T. Mack en a alors développé une version stochastique qui repose sur la version déterministe présentée dans la section 3.2.

La formule utilisée pour calculer les provisions est $\hat{R}_i^{BF} = (1 - z_{n+1-i}^*)\hat{U}_i$.
A partir de cette formule, l'espérance s'écrit donc comme :

$$E(C_{i,j}) = x_i z_j \text{ pour } 1 \leq i \leq n \text{ et } 1 \leq j \leq n+1 \text{ ou de manière équivalente, } E(S_{i,j}) = x_i y_j \\ \text{si } z_j = \sum_{k=1}^j y_k$$

La variable y_j s'interprète comme la cadence de règlements non cumulés et la variable x_i comme une mesure du volume de l'année de survenance i .

Trois hypothèses nécessitent d'être vérifiées pour pouvoir utiliser le modèle stochastique :

(BF1) : Les montants incrémentaux sont indépendants

(BF2) : $\exists x_i, y_j$ tels que $E(S_{i,j}) = x_i y_j$ et $y_1 + \dots + y_n + y_{n+1} = 1$.

(BF3) : $\exists s_j^2$ tel que $Var(S_{i,j}) = x_i \cdot s_j^2$

Si ces hypothèses sont vérifiées, il est possible d'écrire :

$$\begin{aligned} E(R_i) &= E(S_{i,n+2-i} + \dots + S_{i,n+1}) \\ &= E(S_{i,n+2-i}) + \dots + E(S_{i,n+1}) \\ &= x_i y_{n+2-i} + \dots + x_i y_{n+1} \\ &= x_i (y_{n+2-i} + \dots + y_{n+1}) \\ &= x_i (1 - z_{n+1-i}) \end{aligned}$$

$$\begin{aligned} V(R_i) &= Var(S_{i,n+2-i} + \dots + S_{i,n+1}) \\ &= Var(S_{i,n+2-i}) + \dots + Var(S_{i,n+1}) \\ &= x_i s_{n+2-i}^2 + \dots + x_i s_{n+1}^2 \\ &= x_i (s_{n+2-i}^2 + \dots + s_{n+1}^2) \end{aligned}$$

³ Comme mentionné dans la présentation de la méthode Bornhuetter-Ferguson en section 3.2.1, les développements proposés dans cette section sont applicables à la méthode du loss ratio, cas particulier de la méthode Bornhuetter-Ferguson.

Les paramètres y_j et s_j^2 ne sont pas connus et il est nécessaire de les estimer.

En supposant que $x_i = E(U_i) = \hat{U}_i$, les estimateurs suivants peuvent être utilisés :

$$\hat{y}_j = \frac{\sum_{i=1}^{n+1-j} S_{i,j}}{\sum_{i=1}^{n+1-j} \hat{U}_i} \text{ pour } 1 \leq j \leq n$$

puis \hat{y}_{n+1} se déduit de la condition $\sum_{k=1}^{n+1} y_k = 1$.

$$\hat{s}_j^2 = \frac{1}{n-j} \sum_{i=1}^{n+1-j} \frac{(S_{i,j} - \hat{U}_i \hat{y}_j)^2}{\hat{U}_i} \text{ pour } 1 \leq k \leq n-1$$

puis \hat{s}_n^2 est calculé par extrapolation et \hat{s}_{n+1}^2 par interpolation.

L'ajustement et le lissage de ces paramètres (noté s^* et y^*) permettent d'obtenir une estimation de la provision identique à celle calculée dans la version déterministe :

$$\hat{R}_i^{BF} = (1 - z_{n+1-i}^*) \hat{U}_i = \hat{U}_i (y_{n+2-i}^* + \dots + y_{n+1}^*)$$

Le modèle stochastique nous donne bien la même espérance que le modèle déterministe et nous permet ensuite de calculer l'erreur de prédiction. L'erreur de prédiction par année de survenance est définie comme :

$$MSEP(\hat{R}_i^{BF}) = E \left((\hat{R}_i^{BF} - R_i^{BF})^2 \mid S_{i,1}, \dots, S_{i,n+1-i} \right)$$

Or, d'après l'hypothèse 1, $R_i^{BF} = S_{i,n+2-i} + \dots + S_{i,n+1}$ est indépendant de $S_{i,1} \dots S_{i,n+1}$ donc \hat{R}_i^{BF} est aussi indépendant de $S_{i,1} \dots S_{i,n+1}$ et nous avons :

$$\begin{aligned} MSEP(\hat{R}_i^{BF}) &= E \left((\hat{R}_i^{BF} - R_i^{BF})^2 \right) \\ &= \text{Var}(\hat{R}_i^{BF} - R_i^{BF}) + \left(E(\hat{R}_i^{BF}) - E(R_i^{BF}) \right)^2 \\ &= \text{Var}(\hat{R}_i^{BF}) + \text{Var}(R_i^{BF}) \end{aligned}$$

L'erreur de prédiction est obtenue en sommant l'erreur d'estimation et l'erreur de processus.

- L'erreur de processus

Elle est définie comme :

$$\text{Var}(R_i^{BF}) = \text{Var}(S_{i,n+2-i}) + \dots + \text{Var}(S_{i,n+1}) = x_i (s_{n+2-i}^2 + \dots + s_{n+1}^2)$$

Elle peut donc être estimée par :

$$\widehat{Var}(R_i^{BF}) = \widehat{U}_i (s_{n+2-i}^{*2} + \dots + s_{n+1}^{*2})$$

- L'erreur d'estimation

Elle est définie comme :

$$Var(\widehat{R}_i^{BF}) = \left(E(\widehat{U}_i)^2 + Var(\widehat{U}_i) \right) Var(z_{n+1-i}^*) + Var(\widehat{U}_i)(1 - z_{n+1-i}^*)^2$$

Pour estimer cette erreur, il faut calculer $Var(\widehat{U}_i)$ et $Var(z_{n+1-i}^*)$.

$$Var(\widehat{U}_i) = \frac{v_i}{corr} \sum_{j=1}^n v_j \left(\frac{\widehat{U}_i}{v_j} - \widehat{q} \right)^2 \text{ avec } \widehat{q} = \frac{\sum_{j=1}^n \widehat{U}_j}{\sum_{j=1}^n v_j}$$

Dans son article de 2008, Mack suggère de corriger l'éventuelle corrélation positive entre les ULR. Par défaut, la corrélation entre les \widehat{U}_i est supposée nulle et $corr = n - 1$. Dans le cas d'une corrélation constante, $corr = n - \sqrt{n}$ et dans le cas d'une corrélation décroissante, $corr = n - \sqrt{2n}$. Nous avons retenu l'hypothèse par défaut.

$Var(z_{n+1-i}^*) = \min(Var(y_1^*) + \dots + Var(y_{n+1-i}^*); Var(y_{n+2-i}^*) + \dots + Var(y_{n+1}^*))$ où

$$Var(y_k^*) = \frac{s_k^{*2}}{\sum_{j=1}^{n+1-k} \widehat{U}_j} \text{ pour } 1 \leq k \leq n$$

Pour le dernier coefficient, Mack suppose que $Var(y_{n+1}^*) = 0.5y_{n+1}^*$

Finalement,

$$\widehat{MSEP}(\widehat{R}_i^{BF}) = \left(\widehat{U}_i^2 + Var(\widehat{U}_i) \right) Var(z_{n+1-i}^*) + Var(\widehat{U}_i)(1 - z_{n+1-i}^*)^2 + \widehat{U}_i (s_{n+2-i}^{*2} + \dots + s_{n+1}^{*2})$$

Pour l'erreur de prédiction globale, nous ne pouvons pas juste sommer les erreurs de prédiction par année de survenance puisqu'il faut tenir compte des covariances entre les variables aléatoires.

Soit $\widehat{R}^{BF} = \sum_i \widehat{R}_i^{BF}$, l'erreur de prédiction globale est toujours définie comme la somme de l'erreur de processus et d'estimation :

$$MSEP(\widehat{R}^{BF}) = Var(R^{BF}) + Var(\widehat{R}^{BF})$$

Pour l'erreur de processus, nous pouvons faire la somme des erreurs par année de survenance :

$$\widehat{Var}(R^{BF}) = \sum_{i=1}^n \widehat{Var}(R_i^{BF})$$

Pour l'erreur d'estimation en revanche, ce n'est pas possible à cause du fait que les \hat{R}_i^{BF} sont positivement corrélés via les y_j^* :

$$Var(\hat{R}^{BF}) = \sum_{i=1}^n Var(\hat{R}_i^{BF}) + 2 \sum_{i<j} Cov(\hat{R}_i^{BF}, \hat{R}_j^{BF})$$

Pour calculer le terme de covariance, il faut d'abord remarquer que :

$$Cov(\hat{R}_i^{BF}, \hat{R}_j^{BF}) = Cov(\hat{U}_i(1 - z_{n+1-i}^*), \hat{U}_j(1 - z_{n+1-j}^*))$$

Ensuite, nous utilisons la formule générale :

$$Cov(XY, WZ) = Cov(X, W)E(Y)E(Z) + Cov(X, W)Cov(Y, Z) + E(X)E(W)Cov(Y, Z)$$

Nous obtenons ainsi, en négligeant le terme central :

$$\begin{aligned} Cov(\hat{R}_i^{BF}, \hat{R}_j^{BF}) &= \rho_{i,j}^U \sqrt{Var(\hat{U}_i) Var(\hat{U}_j) E(1 - z_{n+1-i}^*) E(1 - z_{n+1-j}^*)} \\ &+ \rho_{i,j}^Z \sqrt{Var(z_{n+1-i}^*) Var(z_{n+1-j}^*) E(\hat{U}_i) E(\hat{U}_j)} \end{aligned}$$

Les coefficients de corrélations $\rho_{i,j}^U$ et $\rho_{i,j}^Z$ sont approximés pour $i < j$ par :

$$\hat{\rho}_{i,j}^U = \frac{1}{1 + |i - j|} \quad \text{et} \quad \hat{\rho}_{i,j}^Z = \sqrt{\frac{z_{n+1-j}^*(1 - z_{n+1-i}^*)}{z_{n+1-i}^*(1 - z_{n+1-j}^*)}} \text{ pour } z_1^* \leq \dots \leq z_{n+1}^*$$

Ainsi, l'erreur de prédiction globale est donnée par :

$$\widehat{MSEP}(\hat{R}^{BF}) = \sum_{i=1}^n \widehat{MSEP}(\hat{R}_i^{BF}) + 2 \sum_{i<j} \widehat{Cov}(\hat{R}_i^{BF}, \hat{R}_j^{BF})$$

$$\begin{aligned} \text{Avec } \widehat{Cov}(\hat{R}_i^{BF}, \hat{R}_j^{BF}) &= \hat{\rho}_{i,j}^U \sqrt{Var(\hat{U}_i) Var(\hat{U}_j) (1 - z_{n+1-i}^*)(1 - z_{n+1-j}^*)} + \\ &\hat{\rho}_{i,j}^Z \sqrt{Var(z_{n+1-i}^*) Var(z_{n+1-j}^*)} \hat{U}_i \hat{U}_j \end{aligned}$$

Conclusion sur la méthode Bornhuetter-Ferguson stochastique de Mack (2008)

Ce modèle permet d'évaluer l'erreur de prédiction de la méthode Bornhuetter-Ferguson et ainsi d'évaluer l'impact des choix faits par l'actuaire lors de la sélection des paramètres. Néanmoins, le modèle tel qu'il est défini ne permet pas de prendre en compte des pondérations dans le cas où toutes les hypothèses du modèle ne seraient pas vérifiées. Or, nous l'avons vu dans la version déterministe, le modèle n'est pas toujours applicable. Nous nous proposons dans la suite d'étendre le modèle au cas pondéré.

4.2.2 Extension pondérée

Similairement à l'extension pondérée de Chain Ladder, nous allons modifier la dernière hypothèse pour pouvoir intégrer des poids :

$$(BF3b) : Var(S_{i,k}) = x_i s_k^2 w_{i,k} \text{ avec } w_{i,k} \in \{0,1\}$$

En prenant $w_{i,k} = 1$, nous retrouvons le cas standard présenté par Mack.

Cependant, il n'y a pas que les hypothèses qui doivent être modifiées pour étendre le modèle au cas pondéré. Les estimateurs de la cadence et des erreurs doivent aussi être modifiés pour intégrer ces pondérations. Ils sont maintenant définis comme :

$$\begin{aligned} \dot{y}_k &= \frac{\sum_{i=1}^{n+1-k} w_{i,k} S_{i,k}}{\sum_{i=1}^{n+1-k} w_{i,k} U_i} \\ \ddot{s}_k^2 &= \frac{1}{n-k} \sum_{i=1}^{n+1-k} \frac{w_{i,k} (S_{i,k} - U_i \dot{y}_k)^2}{U_i} \end{aligned}$$

Cela va modifier le calcul de l'espérance puisque la nouvelle cadence ne prendra pas en compte toutes les observations mais aussi la variance puisque les poids sont intégrés à l'hypothèse 3.

$$\begin{aligned} E(R_i) &= E(S_{i,n+2-i} + \dots + S_{i,n+1}) & V(R_i) &= Var(S_{i,n+2-i} + \dots + S_{i,n+1}) \\ &= E(S_{i,n+2-i}) + \dots + E(S_{i,n+1}) & &= Var(S_{i,n+2-i}) + \dots + Var(S_{i,n+1}) \\ &= x_i \dot{y}_{n+2-i} + \dots + x_i \dot{y}_{n+1} & &= s_{n+2-i}^2 w_{i,n+2-i} x_i + \dots + s_{n+1}^2 w_{i,n+1} x_i \\ &= x_i (\dot{y}_{n+2-i} + \dots + \dot{y}_{n+1}) & &= x_i (s_{n+2-i}^2 w_{i,n+2-i} + \dots + s_{n+1}^2 w_{i,n+1}) \\ &= x_i (1 - \ddot{z}_{n+1-i}) \end{aligned}$$

Il s'en suit également que le calcul de l'erreur de prédiction $MSEP(\hat{R}_i^{BF}) = Var(\hat{R}_i^{BF}) + Var(R_i^{BF})$ évolue puisque l'erreur d'estimation et de processus sont modifiées.

En effet :

$$Var(R_i^{BF}) = U_i (s_{n+2-i}^2 w_{i,n+2-i} + \dots + s_{n+1}^2 w_{i,n+1})$$

$$Var(\hat{R}_i^{BF}) = (E(\hat{U}_i)^2 + Var(\hat{U}_i)) Var(\ddot{z}_{n+1-i}^*) + Var(\hat{U}_i) (1 - \ddot{z}_{n+1-i}^*)^2$$

Le changement pour l'erreur de processus est bien visible puisque les poids apparaissent clairement dans la nouvelle formule. Pour l'erreur d'estimation, comme le changement intervient au niveau des cadences, il n'est pas clairement visible. En réalité :

$$Var(\ddot{z}_{n+1-i}^*) = \min(Var(\ddot{y}_1) + \dots + Var(\ddot{y}_{n+1-i}); Var(\ddot{y}_{n+2-i}) + \dots + Var(\ddot{y}_{n+1}))$$

Avec :

$$\begin{aligned} Var(\ddot{y}_k) &= Var\left(\frac{\sum_{i=1}^{n+1-k} w_{i,k} S_{i,k}}{\sum_{i=1}^{n+1-k} w_{i,k} x_i}\right) \\ &= \frac{1}{(\sum_{i=1}^{n+1-k} w_{i,k} x_i)^2} Var\left(\sum_{i=1}^{n+1-k} w_{i,k} S_{i,k}\right) \\ &= \frac{1}{(\sum_{i=1}^{n+1-k} w_{i,k} x_i)^2} \sum_{i=1}^{n+1-k} w_{i,k}^2 \cdot Var(S_{i,k}) \\ &= \frac{1}{(\sum_{i=1}^{n+1-k} w_{i,k} x_i)^2} \sum_{i=1}^{n+1-k} w_{i,k} \cdot (x_i s_k^2) \\ &= \frac{s_k^2}{(\sum_{i=1}^{n+1-k} x_i w_{i,k})} \end{aligned}$$

Comme les poids sont soit des 0 soit des 1, il est possible de simplifier

4.2.3 Application aux données

4.2.3.1 Données d'acceptations anglaises

Voici le résultat obtenu en appliquant la méthode aux données du portefeuille d'acceptations des polices anglaises :

	Estimations			Calcul de l'erreur de prédiction			
	ULR	Montant ultime	IBNR	Erreur de processus	Erreur d'estimation	Erreur de prédiction	% Erreur
2002	73,7%	391	0	5 520	-	5 520	521%
2003	86,4%	936	0	13 196	0	13 196	22%
2004	55,2%	880	0	12 501	64	12 565	25%
2005	59,9%	521	0	7 445	33	7 478	27%
2006	54,0%	358	0	5 635	109	5 744	39%
2007	71,9%	710	0	11 682	610	12 292	18%
2008	73,5%	677	0	11 455	699	12 155	26%
2009	47,4%	488	0	9 127	611	9 739	34%
2010	47,6%	625	0	13 280	1 318	14 598	20%
2011	42,1%	697	1	15 950	1 860	17 810	22%
2012	53,7%	958	2	23 943	3 826	27 769	19%
2013	45,5%	1 089	4	33 542	6 316	39 859	32%
2014	51,6%	1 353	10	56 557	12 803	69 360	22%
2015	54,4%	1 355	18	68 301	14 906	83 206	19%
2016	53,9%	1 037	25	65 039	10 451	75 490	17%
2017	48,5%	325	14	28 369	1 038	29 407	33%
2018	11,2%	32	3	3 276	119	3 396	102%
2019	4,9%	16	2	2 015	188	2 203	155%
2020	2,2%	17	5	2 377	1 662	4 039	214%
2021	21,4%	364	253	53 658	24 648	78 307	60%
Total	50%	12 830	336	442 868	261 523	704 391,20	7,6%

Figure 4.7 : Calcul de l'erreur de prédiction du Bornhuetter-Ferguson pondéré

L'erreur de prédiction est largement inférieure à celle de la méthode Chain Ladder (9,1%) bien que le montant d'IBNR soit assez proche (388 €).

Pour la version avec les ULR exogènes, l'erreur est plus importante :

	Estimations			Calcul de l'erreur de prédiction			
	ULR	Montant ultime	IBNR	Erreur de processus	Erreur d'estimation	Erreur de prédiction	% Erreur
2002	82,0%	435	0	5 751	-	5 751	532%
2003	105,8%	1 146	0	15 139	0	15 139	23%
2004	74,3%	1 184	0	15 735	74	15 810	28%
2005	83,1%	723	0	9 653	40	9 693	31%
2006	79,7%	529	0	7 647	132	7 778	45%
2007	114,6%	1 132	0	16 987	867	17 854	21%
2008	121,9%	1 122	0	17 157	1 001	18 158	32%
2009	81,7%	841	0	13 787	829	14 616	42%
2010	83,0%	1 089	1	19 529	1 779	21 307	24%
2011	80,1%	1 327	2	25 222	2 963	28 185	28%
2012	108,7%	1 938	4	39 648	7 051	46 699	24%
2013	94,2%	2 255	9	54 313	11 783	66 096	41%
2014	108,5%	2 846	20	88 245	23 987	112 232	28%
2015	108,5%	2 704	35	98 129	24 615	122 743	22%
2016	117,8%	2 263	54	96 513	19 495	116 008	21%
2017	123,4%	827	36	47 362	2 200	49 562	41%
2018	87,5%	254	21	16 857	244	17 101	173%
2019	88,1%	293	36	22 998	317	23 315	238%
2020	70,1%	544	144	47 377	2 057	49 433	132%
2021	64,2%	1 095	761	109 850	26 398	136 248	38%
Total	96%	24 549	1 123	767 899	457 176	1 225 074,58	9,4%

Figure 4.8 : Calcul de l'erreur de prédiction du Bornhuetter-Ferguson pondéré avec ULR exogènes

En comparant les différentes méthodes appliquées sur les données d'acceptations anglaises, nous remarquons que la méthode Bornhuetter-Ferguson pondérée est celle qui semble le plus adaptée car l'erreur de prédiction associée à cette méthode est plus faible que pour les autres.

4.2.3.2 Données françaises

- Classe « Attritional »

	Estimations		Calcul de l'erreur de prédiction			
	ULR	IBNR	Erreur de processus	Erreur d'estimation	Erreur de prédiction	% Erreur
1996	66,0%	0	1	-	1	3189%
1997	49,0%	0	1	0	1	-121%
1998	58,3%	0	1	0	1	-55%
1999	61,5%	0	1	0	1	605%
2000	57,9%	0	1	0	1	-35%
2001	59,3%	0	2	0	2	1343%
2002	53,6%	0	2	0	2	-3430%
2003	48,5%	0	3	0	3	593%
2004	44,6%	0	2	0	3	797%
2005	45,3%	0	2	0	2	420%
2006	45,4%	0	2	0	2	308%
2007	47,6%	0	2	0	3	264%
2008	48,3%	1	3	0	3	143%
2009	49,7%	1	3	0	3	154%
2010	51,6%	1	3	0	4	127%
2011	50,2%	1	3	0	4	113%
2012	48,9%	1	3	0	4	79%
2013	46,4%	0	3	0	3	170%
2014	50,0%	0	3	0	3	52%
2015	53,1%	0	3	0	3	53%
2016	54,9%	-1	3	0	4	46%
2017	57,5%	-1	4	1	4	31%
2018	55,4%	-3	4	1	4	25%
2019	57,2%	-5	5	1	6	18%
2020	45,3%	-7	5	1	6	14%
2021	53,7%	5	8	0	9	6%
Total	52%	-6	71	72	143	10,7%

Figure 4.9 : Calcul de l'erreur de prédiction du Bornhuetter-Ferguson pondéré pour la classe "Attritional"

Le montant d'IBNR obtenu par la méthode Bornhuetter-Ferguson indépendante pondérée est proche du montant trouvé grâce à la méthode Chain Ladder pondérée mais l'erreur de prédiction est largement inférieure (10,7% vs 15%). Nous concluons que la méthode la plus adaptée est la méthode Bornhuetter-Ferguson pondérée.

- Classe « Large »

	Estimations		Calcul de l'erreur de prédiction			
	ULR	IBNR	Erreur de processus	Erreur d'estimation	Erreur de prédiction	% Erreur
1996	4,0%	0	8	-	8	143%
1997	4,4%	2	10	1	11	30%
1998	6,8%	5	31	16	47	33%
1999	6,8%	7	34	21	55	64%
2000	6,3%	10	33	25	58	74%
2001	5,4%	11	44	33	77	51%
2002	4,9%	13	48	46	95	61%
2003	5,2%	17	57	62	119	40%
2004	3,2%	10	35	53	88	43%
2005	6,2%	22	73	80	153	36%
2006	5,5%	20	71	83	154	38%
2007	4,4%	17	63	82	145	70%
2008	8,0%	30	120	105	225	31%
2009	5,9%	23	97	94	190	40%
2010	7,3%	28	143	105	248	23%
2011	7,0%	28	168	112	279	22%
2012	8,7%	33	231	123	354	21%
2013	6,9%	25	237	104	341	25%
2014	8,3%	25	293	98	391	16%
2015	9,5%	20	376	88	464	18%
2016	9,0%	16	446	82	528	16%
2017	10,1%	12	583	93	676	15%
2018	10,1%	14	753	98	851	13%
2019	10,6%	21	908	110	1 018	12%
2020	9,3%	33	979	129	1 108	12%
2021	11,1%	114	1 563	592	2 155	13%
Total	7%	556	7 404	24 623	32 027	7,8%

Figure 4.10 : Calcul de l'erreur de prédiction du Bornhuetter-Ferguson pondéré pour la classe "Large"

Les montants d'IBNR calculés par les deux méthodes sont très proches mais l'erreur de prédiction de la méthode Bornhuetter-Ferguson est légèrement inférieure à celle de Chain Ladder (l'erreur de la méthode Chain Ladder est de 8,3%). Nous concluons que la méthode la plus adaptée pour les données de la classe « Large » est aussi la méthode Bornhuetter-Ferguson.

- Classe « Exceptionnelle »

	Estimations		Calcul de l'erreur de prédiction			
	ULR	IBNR	Erreur de processus	Erreur d'estimation	Erreur de prédiction	% Erreur
1996	4,1%	0	200	-	200	3115%
1997	13,9%	0	753	0	753	338%
1998	8,9%	0	463	0	463	46%
1999	4,8%	0	253	0	254	253%
2000	10,6%	1	589	8	597	29%
2001	7,7%	1	449	6	455	564%
2002	7,1%	1	542	26	568	158%
2003	3,2%	1	444	52	496	631%
2004	12,5%	3	1 624	327	1 951	133%
2005	10,6%	2	1 316	223	1 539	337%
2006	10,8%	3	1 299	218	1 517	42%
2007	12,1%	4	1 449	262	1 711	69%
2008	8,1%	3	993	137	1 130	160%
2009	14,5%	6	1 852	397	2 249	31%
2010	17,4%	8	3 854	1 108	4 962	29%
2011	13,4%	8	3 119	731	3 850	53%
2012	18,1%	13	4 993	1 571	6 564	28%
2013	12,6%	10	4 913	1 096	6 009	59%
2014	13,9%	12	5 275	1 128	6 402	47%
2015	7,7%	8	2 949	408	3 357	52%
2016	14,6%	18	5 964	1 175	7 139	30%
2017	11,9%	18	5 052	782	5 833	27%
2018	15,8%	31	7 059	1 313	8 372	20%
2019	15,8%	44	7 324	1 359	8 684	20%
2020	11,5%	52	6 047	949	6 996	22%
2021	12,7%	129	8 257	2 139	10 396	23%
Total	11%	375	77 031	130 582	207 613	11,7%

Figure 4.11 : Calcul de l'erreur de prédiction du Bornhuetter-Ferguson pondéré pour la classe "Exceptionnelle"

Le montant d'IBNR est supérieur à celui calculé par la méthode Chain Ladder et l'erreur de prédiction est plus faible (11,7% vs 13,3%). Nous concluons une fois encore que la méthode la plus adaptée est la méthode Bornhuetter-Ferguson.

Conclusion sur la méthode du Bornhuetter-Ferguson stochastique

Cette méthode est similaire à la méthode Chain Ladder stochastique dans sa structure mais reflète bien la différence fondamentale entre les méthodes Chain Ladder et Bornhuetter-Ferguson : les provisions estimées par la méthode Chain Ladder sont proportionnelles aux montants déjà connus alors que la méthode Bornhuetter-Ferguson n'en dépend pas. La méthode Bornhuetter-Ferguson stochastique suppose en effet que les différences observées entre les montants $C_{i,n+1-i}$ et $\hat{z}_{n+1-i}\hat{U}_i$ sont purement aléatoires et ne sont pas indicatives des développements futurs.

Pour nos données, les résultats obtenus à partir de la méthode Bornhuetter-Ferguson sont meilleurs que ceux calculés grâce à la méthode Chain Ladder.

4.3 Méthode du coût moyen stochastique

Nous allons chercher à calculer l'erreur de prédiction (MSEP) de la méthode du coût moyen à l'aide d'une formule fermée adaptée de la méthode de Mack (1993, 1999). Dans sa version déterministe, la méthode du coût moyen consiste à combiner les estimations Chain Ladder pour le triangle des nombres d'une part et pour le triangle des coûts moyens d'autre part. Pour calculer l'erreur de prédiction, nous allons exploiter la décomposition de la charge de sinistres entre fréquence et sévérité et reprendre les hypothèses de Mack utilisées dans la version stochastique de la méthode Chain Ladder.

Soit $N_{i,j}$ les informations contenues dans le triangle de nombres cumulés et $C_{i,j}$ les informations contenues dans le triangle de charges cumulées. On pose $\bar{X}_{i,j}$ les informations contenues dans le triangle des coûts moyens définis comme $\bar{X}_{i,j} = \frac{C_{i,j}}{N_{i,j}}$.

On suppose que les $N_{i,j}$ et les $\bar{X}_{i,j}$ vérifient les hypothèses du modèle de Mack 1999 pour le calcul de l'erreur de prédiction de la méthode Chain Ladder pondérée.

Ainsi, on suppose que pour les $N_{i,j}$:

(H1) Les années de survenance sont indépendantes :

$$\{N_{i,1}, \dots, N_{i,n}\} \text{ est indépendant de } \{N_{l,1}, \dots, N_{l,n}\} \text{ pour } i \neq l$$

(H2) Il existe des facteurs $f_j^{(N)}$:

$$E(N_{i,j+1} | N_{i,1}, \dots, N_{i,j}) = N_{i,j} \cdot f_j^{(N)} \text{ pour } 1 \leq i \leq n \text{ et } 1 \leq j \leq n-1$$

(H3) Il existe $\sigma_j^{2(N)}$:

$$Var(N_{i,j+1} | N_{i,1}, \dots, N_{i,j}) = \frac{\sigma_j^{2(N)} \cdot N_{i,j}^2}{w_{i,j}^{(N)} \cdot N_{i,j}^{\alpha(N)}} \text{ pour } 1 \leq i \leq n \text{ et } 1 \leq j \leq n-1$$

De la même façon, on suppose que pour les $\bar{X}_{i,j}$:

(H1) : $\{\bar{X}_{i,1}, \dots, \bar{X}_{i,n}\}$ est indépendant de $\{\bar{X}_{l,1}, \dots, \bar{X}_{l,n}\}$ pour $i \neq l$

(H2) : Il existe des facteurs $f_j^{(X)}$:

$$E(\bar{X}_{i,j+1} | \bar{X}_{i,1}, \dots, \bar{X}_{i,j}) = \bar{X}_{i,j} \cdot f_j^{(X)} \text{ pour } 1 \leq i \leq n \text{ et } 1 \leq j \leq n-1$$

(H3) : Il existe $\sigma_j^{2(X)}$:

$$Var(\bar{X}_{i,j+1} | \bar{X}_{i,1}, \dots, \bar{X}_{i,j}) = \frac{\sigma_j^{2(X)} \cdot \bar{X}_{i,j}^2}{w_{i,j}^{(X)} \cdot \bar{X}_{i,j}^{\alpha(X)}} \text{ pour } 1 \leq i \leq n \text{ et } 1 \leq j \leq n-1$$

Les $w_{i,j}^{(X)} \in [0, 1]$ et les $w_{i,j}^{(N)} \in [0, 1]$ sont des poids et $\alpha(X) = \alpha(N) \in \{0,1,2\}$.

On remarque qu'il est possible d'écrire $C_{i,j} = \sum_{k_{i,j}}^{N_{i,j}} X_{i,j,k_{i,j}} = N_{i,j} \bar{X}_{i,j}$ en utilisant le modèle collectif de la théorie du risque. Dans ce cadre, les $N_{i,j}$ sont indépendants des coûts individuels $X_{i,j,k_{i,j}}$ et les $X_{i,j,k_{i,j}}$ sont indépendants et identiquement distribués.

Le but est de calculer l'erreur de prédiction commise par la méthode du coût moyen. Le but est donc de calculer :

$$MSEP(\hat{R}_i) = E \left((\hat{R}_i - R_i)^2 \middle| D \right)$$

Dans la méthode Chain Ladder, les montants des provisions sont définis comme la différence entre la charge finale et la dernière charge observée donc :

$$MSEP(\hat{R}_i) = E \left((\hat{C}_{i,n} - C_{i,n})^2 \middle| D \right) = \text{Var}(C_{i,n}|D) + (E(C_{i,n}|D) - \hat{C}_{i,n})^2$$

où D est l'ensemble de l'information connue $D = \{X_{i,j,k_{i,j}}, N_{i,j} : i + j \leq n + 1, k_{i,j} \leq N_{i,j}\}$

On utilisera dans la suite les notations suivantes :

$$\begin{aligned} E_D(X) &= E(X|D) \\ \text{Var}_D(X) &= \text{Var}(X|D) \end{aligned}$$

4.3.1 Calcul des termes d'espérance

Dans le cadre du modèle collectif, on rappelle que l'on peut écrire :

$$E(C_{i,n}|D) = E(N_{i,n}|D) \times E(\bar{X}_{i,n}|D)$$

Le calcul de $E(N_{i,n}|D)$ donne :

$$\begin{aligned} E(N_{i,n}|D) &= E_D(N_{i,n}) \\ &= E_D \left(E(N_{i,n} | N_{i,1}, \dots, N_{i,n-1}) \right) \\ &= E_D \left(N_{i,n-1} f_{n-1}^{(N)} \right) \\ &= f_{n-1}^{(N)} E_D(N_{i,n-1}) \\ &= \dots \\ &= E_D(N_{i,n-i+1}) f_{n-i+1}^{(N)} \dots f_{n-1}^{(N)} \\ &= N_{i,n-i+1} f_{n-i+1}^{(N)} \dots f_{n-1}^{(N)} \end{aligned}$$

De façon analogue, le calcul de $E(\bar{X}_{i,n}|D)$ donne :

$$\begin{aligned}
E(\bar{X}_{i,n}|D) &= E_D(\bar{X}_{i,n}) \\
&= E_D\left(E(\bar{X}_{i,n}|\bar{X}_{i,1}, \dots, \bar{X}_{i,n-1})\right) \\
&= E_D\left(\bar{X}_{i,n-1}f_{n-1}^{(X)}\right) \\
&= f_{n-1}^{(X)}E_D(\bar{X}_{i,n-1}) \\
&= \dots \\
&= E_i(\bar{X}_{i,n-i+1})f_{n-i+1}^{(X)} \dots f_{n-1}^{(X)} \\
&= \bar{X}_{i,n-i+1}f_{n-i+1}^{(X)} \dots f_{n-1}^{(X)}
\end{aligned}$$

Finalemment :

$$E(C_{i,n}|D) = E(N_{i,n}|D) \times E(\bar{X}_{i,n}|D) = N_{i,n-i+1} \bar{X}_{i,n-i+1} \times f_{n-i+1}^{(X)} f_{n-i+1}^{(N)} \dots f_{n-1}^{(X)} f_{n-1}^{(N)}$$

4.3.2 Calcul des termes de variance

Dans le cadre du modèle collectif, on peut montrer que :

$$Var(C_{i,n}|D) = E(N_{i,n}|D) \frac{Var(\bar{X}_{i,n}|D)}{E\left(\frac{1}{N_{i,n}}|D\right)} + Var(N_{i,n}|D) E(\bar{X}_{i,n}|D)^2$$

La démonstration est reportée en [annexe 2](#).

Le calcul de $Var(\bar{X}_{i,n}|D)$ donne :

$$\begin{aligned}
Var(\bar{X}_{i,n}|D) &= Var_i(\bar{X}_{i,n}) \\
&= E_D\left(Var(\bar{X}_{i,n}|\bar{X}_{i,1}, \dots, \bar{X}_{i,n-1})\right) + Var_D\left(E(\bar{X}_{i,n}|\bar{X}_{i,1}, \dots, \bar{X}_{i,n-1})\right) \\
&= E_D\left(\frac{\sigma_{n-1}^{2(X)} \bar{X}_{i,n-1}^2}{w_{i,n-1}^{(X)} \bar{X}_{i,n-1}^\alpha}\right) + Var_D\left(\bar{X}_{i,n-1}f_{n-1}^{(X)}\right) \\
&= \frac{\sigma_{n-1}^{2(X)}}{w_{i,n-1}^{(X)}} E_D\left(\bar{X}_{i,n-1}^{2-\alpha}\right) + \left(f_{n-1}^{(X)}\right)^2 Var_D(\bar{X}_{i,n-1})
\end{aligned}$$

Il faut distinguer les cas en fonction de la valeur du paramètre α .

- **Cas $\alpha = 1$:**

$$\begin{aligned}
\text{Var}(\bar{X}_{i,n}|D) &= \frac{\sigma_{n-1}^{2(X)}}{w_{i,n-1}^{(X)}} E_D(\bar{X}_{i,n-1}) + (f_{n-1}^{(X)})^2 \text{Var}_D(\bar{X}_{i,n-1}) \\
&= \frac{\sigma_{n-1}^{2(X)}}{w_{i,n-1}^{(X)}} \bar{X}_{i,n-i+1} f_{n-i+1}^{(X)} \cdots f_{n-2}^{(X)} + (f_{n-1}^{(X)})^2 \left[E_D(\text{Var}(\bar{X}_{i,n-1}|\bar{X}_{i,1}, \dots, \bar{X}_{i,n-2})) + \right. \\
&\quad \left. \text{Var}_D(E(\bar{X}_{i,n-1}|\bar{X}_{i,1}, \dots, \bar{X}_{i,n-2})) \right] \\
&= \frac{\sigma_{n-1}^{2(X)}}{w_{i,n-1}^{(X)}} \bar{X}_{i,n-i+1} f_{n-i+1}^{(X)} \cdots f_{n-2}^{(X)} + (f_{n-1}^{(X)})^2 \frac{\sigma_{n-2}^{2(X)}}{w_{i,n-2}^{(X)}} E_D(\bar{X}_{i,n-2}) + (f_{n-1}^{(X)})^2 (f_{n-2}^{(X)})^2 \text{Var}_D(\bar{X}_{i,n-2}) = \dots \\
&= \bar{X}_{i,n-i+1} \sum_{k=n+1-i}^{n-1} f_{n-i+1}^{(X)} \cdots f_{k-1}^{(X)} \times \frac{\sigma_k^{2(X)}}{w_{i,k}^{(X)}} \times (f_{k+1}^{(X)})^2 \cdots (f_{n-1}^{(X)})^2
\end{aligned}$$

- **Cas $\alpha = 2$:**

$$\begin{aligned}
\text{Var}(\bar{X}_{i,n}|D) &= \frac{\sigma_{n-1}^{2(X)}}{w_{i,n-1}^{(X)}} + (f_{n-1}^{(X)})^2 \text{Var}_D(\bar{X}_{i,n-1}) \\
&= \frac{\sigma_{n-1}^{2(X)}}{w_{i,n-1}^{(X)}} + (f_{n-1}^{(X)})^2 \left[E_D(\text{Var}(\bar{X}_{i,n-1}|\bar{X}_{i,1}, \dots, \bar{X}_{i,n-2})) + \text{Var}_D(E(\bar{X}_{i,n-1}|\bar{X}_{i,1}, \dots, \bar{X}_{i,n-2})) \right] \\
&= \frac{\sigma_{n-1}^{2(X)}}{w_{i,n-1}^{(X)}} + (f_{n-1}^{(X)})^2 \frac{\sigma_{n-2}^{2(X)}}{w_{i,n-2}^{(X)}} + (f_{n-1}^{(X)})^2 (f_{n-2}^{(X)})^2 \text{Var}_D(\bar{X}_{i,n-2}) = \dots \\
&= \sum_{k=n+1-i}^{n-1} \frac{\sigma_k^{2(X)}}{w_{i,k}^{(X)}} \times (f_{k+1}^{(X)})^2 \cdots (f_{n-1}^{(X)})^2
\end{aligned}$$

- **Cas $\alpha = 0$:**

$$\begin{aligned}
\text{Var}(\bar{X}_{i,n}|D) &= \frac{\sigma_{n-1}^{2(X)}}{w_{i,n-1}^{(X)}} E_D(\bar{X}_{i,n-1}^2) + (f_{n-1}^{(X)})^2 \text{Var}_D(\bar{X}_{i,n-1}) \\
&= \frac{\sigma_{n-1}^{2(X)}}{w_{i,n-1}^{(X)}} \left[\text{Var}_D(\bar{X}_{i,n-1}) + E_D(\bar{X}_{i,n-1})^2 \right] + (f_{n-1}^{(X)})^2 \text{Var}_D(\bar{X}_{i,n-1}) \\
&= \frac{\sigma_{n-1}^{2(X)}}{w_{i,n-1}^{(X)}} (\bar{X}_{i,n-i+1} f_{n-i+1}^{(X)} \cdots f_{n-2}^{(X)})^2 + \\
&\quad \left((f_{n-1}^{(X)})^2 + \frac{\sigma_{n-1}^{2(X)}}{w_{i,n-1}^{(X)}} \right) \left[E_D(\text{Var}(\bar{X}_{i,n-1}|\bar{X}_{i,1}, \dots, \bar{X}_{i,n-2})) + \text{Var}_D(E(\bar{X}_{i,n-1}|\bar{X}_{i,1}, \dots, \bar{X}_{i,n-2})) \right] = \dots \\
&= \bar{X}_{i,n-i+1}^2 \sum_{k=n+1-i}^{n-1} (f_{n-i+1}^{(X)})^2 \cdots (f_{k-1}^{(X)})^2 \times \frac{\sigma_k^{2(X)}}{w_{i,k}^{(X)}} \times \prod_{l=k+1}^{n-1} \left((f_l^{(X)})^2 + \frac{\sigma_l^{2(X)}}{w_{i,l}^{(X)}} \right)
\end{aligned}$$

Similairement, le calcul de $Var(N_{i,n} | D)$ donne en fonction de la valeur du paramètre α :

- **Cas $\alpha = 1$:**

$$Var(N_{i,n} | D) = N_{i,n-i+1} \sum_{k=n+1-i}^{n-1} f_{n-i+1}^{(N)} \dots f_{k-1}^{(N)} \times \frac{\sigma_k^{2(N)}}{w_{i,k}^{(N)}} \times (f_{k+1}^{(N)})^2 \dots (f_{n-1}^{(N)})^2$$

- **Cas $\alpha = 2$:**

$$Var(N_{i,n} | D) = \sum_{k=n+1-i}^{n-1} \frac{\sigma_k^{2(N)}}{w_{i,k}^{(N)}} \times (f_{k+1}^{(N)})^2 \dots (f_{n-1}^{(N)})^2$$

- **Cas $\alpha = 0$:**

$$Var(N_{i,n} | D) = (N_{i,n-i+1})^2 \sum_{k=n+1-i}^{n-1} (f_{n-i+1}^{(N)})^2 \dots (f_{k-1}^{(N)})^2 \times \frac{\sigma_k^{2(N)}}{w_{i,k}^{(N)}} \times \prod_{l=k+1}^{n-1} \left((f_l^{(N)})^2 + \frac{\sigma_l^{2(N)}}{w_{i,l}^{(N)}} \right)$$

Le terme $\left(E \left(\frac{1}{N_{i,n}} \middle| D \right) \right)^{-1}$ ne peut pas être évalué sans hypothèse supplémentaire. On peut cependant utiliser l'inégalité de Jensen pour en déterminer un majorant :

Soit g une fonction convexe définie sur un intervalle réel I et Y une variable aléatoire à valeurs dans I dont $E(g(Y))$ existe. L'inégalité de Jensen nous assure que :

$$g(E[Y]) \leq E[g(Y)].$$

L'inégalité de Jensen reste valable pour l'espérance conditionnelle.

Comme la fonction $g(x) = \frac{1}{x}$ est convexe, nous pouvons appliquer l'inégalité de Jensen :

$$\frac{1}{E(Y)} \leq E\left(\frac{1}{Y}\right)$$

Avec $Y = 1/N_{i,n}$ et en conditionnant par D , nous obtenons :

$$\frac{1}{E\left(\frac{1}{N_{i,n}} \middle| D\right)} \leq E(N_{i,n} | D) \Rightarrow E(N_{i,n} | D) \frac{Var(\bar{X}_{i,n} | D)}{E\left(\frac{1}{N_{i,n}} \middle| D\right)} \leq Var(\bar{X}_{i,n} | D) \left(E(N_{i,n} | D) \right)^2$$

D'où

$$Var(C_{i,n} | D) \leq \left(E(\bar{X}_{i,n} | D) \right)^2 Var(N_{i,n} | D) + \left(E(N_{i,n} | D) \right)^2 Var(\bar{X}_{i,n} | D)$$

4.3.3 Calcul et estimation de la MSEP

Une fois les termes d'espérance et de variance calculés, il est possible de déterminer un estimateur de l'erreur de prédiction par année de survenance grâce à la décomposition suivante de la MSEP :

$$MSEP(\hat{R}_i) = Var(C_{i,n}|D) + (E(C_{i,n}|D) - \hat{C}_{i,n})^2$$

Pour cela, nous allons remplacer les quantités inconnues par des estimateurs.

$$\hat{C}_{i,n} = \hat{N}_{i,n-i+1} \hat{X}_{i,n-i+1} = N_{i,n-i+1} \bar{X}_{i,n-i+1} \hat{f}_{n-i+1}^X \hat{f}_{n-i+1}^N \dots \hat{f}_{n-1}^X \hat{f}_{n-1}^N$$

Il est possible d'estimer la variance conditionnelle $Var(C_{i,n}|D)$ par :

$$\hat{N}_{i,n}^2 \hat{X}_{i,n}^2 \sum_{k=n+1-i}^{n-1} \left(\frac{\hat{\sigma}_k^{2(X)}/(\hat{f}_k^{(X)})^2}{w_{i,k}^{(X)} \bar{X}_{i,k}^\alpha} \prod_{l=k+1}^{n-1} \left(1 + \frac{\hat{\sigma}_l^{2(X)} \mathbf{1}_{\{\alpha=0\}}}{w_{i,l}^{(X)} (f_l^{(X)})^2} \right) + \frac{\hat{\sigma}_k^{2(N)}/(\hat{f}_k^{(N)})^2}{w_{i,k}^{(N)} N_{i,k}^\alpha} \prod_{l=k+1}^{n-1} \left(1 + \frac{\hat{\sigma}_l^{2(N)} \mathbf{1}_{\{\alpha=0\}}}{w_{i,l}^{(N)} (f_l^{(N)})^2} \right) \right)$$

Les facteurs de développement sont remplacés par les estimations données par la méthode Chain Ladder. Les termes $\sigma_k^{2(X)}$ et $\sigma_k^{2(N)}$ sont eux aussi remplacés par leurs estimateurs :

$$\hat{\sigma}_k^{2(X)} = \frac{1}{n-k-1} \sum_{i=1}^{n-k} w_{i,k}^{(X)} \bar{X}_{i,k}^\alpha \left(\frac{\bar{X}_{i,k+1}}{\bar{X}_{i,k}} - \hat{f}_k^{(X)} \right)^2$$

$$\hat{\sigma}_k^{2(N)} = \frac{1}{n-k-1} \sum_{i=1}^{n-k} w_{i,k}^{(N)} N_{i,k}^\alpha \left(\frac{N_{i,k+1}}{N_{i,k}} - \hat{f}_k^{(N)} \right)^2$$

Pour estimer le terme $(E(C_{i,n}|D) - \hat{C}_{i,n})^2$, nous ne pouvons pas juste remplacer les f_j par leurs estimateurs car sinon cela donnerait 0. En effet :

$$\begin{aligned} (E(C_{i,n}|D) - \hat{C}_{i,n})^2 &= \left(N_{i,n-i+1} \bar{X}_{i,n-i+1} \times f_{n-i+1}^{(X)} f_{n-i+1}^{(N)} \dots f_{n-1}^{(X)} f_{n-1}^{(N)} - \hat{C}_{i,n} \right)^2 \\ &= \left(N_{i,n-i+1} \bar{X}_{i,n-i+1} \times (f_{n-i+1}^{(X)} f_{n-i+1}^{(N)} \dots f_{n-1}^{(X)} f_{n-1}^{(N)} - \hat{f}_{n-i+1}^{(X)} \hat{f}_{n-i+1}^{(N)} \dots \hat{f}_{n-1}^{(X)} \hat{f}_{n-1}^{(N)}) \right)^2 \end{aligned}$$

Nous allons poser :

$$F = f_{n-i+1}^{(X)} f_{n-i+1}^{(N)} \dots f_{n-1}^{(X)} f_{n-1}^{(N)} - \hat{f}_{n-i+1}^{(X)} \hat{f}_{n-i+1}^{(N)} \dots \hat{f}_{n-1}^{(X)} \hat{f}_{n-1}^{(N)} = \sum_{k=n+1-i}^{n-1} S_k$$

où

$$S_k = \hat{f}_{n-i+1}^{(X)} \hat{f}_{n-i+1}^{(N)} \dots \hat{f}_{k-1}^{(X)} \hat{f}_{k-1}^{(N)} (f_k^{(X)} f_k^{(N)} - \hat{f}_k^{(X)} \hat{f}_k^{(N)}) f_{k+1}^{(X)} f_{k+1}^{(N)} \dots \hat{f}_{n-1}^{(X)} \hat{f}_{n-1}^{(N)}$$

Ainsi :

$$(E(C_{i,n}|D) - \hat{C}_{i,n})^2 = N_{i,n-i+1}^2 \bar{X}_{i,n-i+1}^2 F^2$$

Nous cherchons un estimateur pour F^2 . Afin d'alléger les notations, on pose $f_k = f_k^{(X)} f_k^{(N)}$.

$$F^2 = \left(\sum_{k=n+1-i}^{n-1} S_k \right)^2 = \sum_{k=n+1-i}^{n-1} S_k^2 + 2 \sum_{j < k} S_j S_k \approx \sum_{k=n+1-i}^{n-1} E(S_k^2 | B_k) + 2 \sum_{j < k} E(S_j S_k | B_k)$$

$$\text{avec } B_k = \{N_{i,j}, \bar{X}_{i,j} | j \leq k, i + j \leq n + 1\}$$

Après calculs ([en annexe 3](#)), nous obtenons :

- $E(S_j S_k | B_k) = 0$
- $E(S_k^2 | B_k) = \hat{f}_{n+1-i}^2 \dots \hat{f}_{k-1}^2 E((f_k - \hat{f}_k)^2 | B_k) \hat{f}_{k+1}^2 \dots \hat{f}_{n-1}^2$
 $= \hat{f}_{n+1-i}^2 \dots \hat{f}_{k-1}^2 \text{Var}(\hat{f}_k | B_k) \hat{f}_{k+1}^2 \dots \hat{f}_{n-1}^2$
- $\text{Var}(\hat{f}_k | B_k) = \left(\frac{\hat{\sigma}_k^{2(X)} / (\hat{f}_k^{(X)})^2}{\sum_{j=1}^{n-k} w_{j,k}^{(X)} \bar{X}_{j,k}^\alpha} + \frac{\hat{\sigma}_k^{2(X)} \hat{\sigma}_k^{2(N)} / (\hat{f}_k)^2}{(\sum_{j=1}^{n-k} w_{j,k}^{(X)} \bar{X}_{j,k}^\alpha) (\sum_{j=1}^{n-k} w_{j,k}^{(N)} N_{j,k}^\alpha)} + \frac{\hat{\sigma}_k^{2(N)} / (\hat{f}_k^{(N)})^2}{\sum_{j=1}^{n-k} w_{j,k}^{(N)} N_{j,k}^\alpha} \right)$

Finalement, $(E(C_{i,n}|D) - \hat{C}_{i,n})^2$ peut s'estimer par :

$$\hat{N}_{i,n}^2 \hat{X}_{i,n}^2 \sum_{k=n+1-i}^{n-1} \left(\frac{\hat{\sigma}_k^{2(X)} / (\hat{f}_k^{(X)})^2}{\sum_{j=1}^{n-k} w_{j,k}^{(X)} \bar{X}_{j,k}^\alpha} + \frac{\hat{\sigma}_k^{2(X)} \hat{\sigma}_k^{2(N)} / (\hat{f}_k)^2}{(\sum_{j=1}^{n-k} w_{j,k}^{(X)} \bar{X}_{j,k}^\alpha) (\sum_{j=1}^{n-k} w_{j,k}^{(N)} N_{j,k}^\alpha)} + \frac{\hat{\sigma}_k^{2(N)} / (\hat{f}_k^{(N)})^2}{\sum_{j=1}^{n-k} w_{j,k}^{(N)} N_{j,k}^\alpha} \right)$$

L'estimateur de $mse(\hat{R}_i)$ peut donc s'écrire comme :

$$\hat{N}_{i,n}^2 \hat{X}_{i,n}^2 \sum_{k=n+1-i}^{n-1} \left(\frac{\hat{\sigma}_k^{2(X)} / (\hat{f}_k^{(X)})^2}{w_{i,k}^{(X)} \bar{X}_{i,k}^\alpha} \prod_{l=k+1}^{n-1} \left(1 + \frac{\hat{\sigma}_k^{2(X)} \mathbf{1}_{\{\alpha=0\}}}{w_{i,l}^{(X)} (f_l^{(X)})^2} \right) + \frac{\hat{\sigma}_k^{2(N)} / (\hat{f}_k^{(N)})^2}{w_{i,k}^{(N)} N_{i,k}^\alpha} \prod_{l=k+1}^{n-1} \left(1 + \frac{\hat{\sigma}_k^{2(N)} \mathbf{1}_{\{\alpha=0\}}}{w_{i,l}^{(N)} (f_l^{(N)})^2} \right) \right)$$

+

$$\hat{N}_{i,n}^2 \hat{X}_{i,n}^2 \sum_{k=n+1-i}^{n-1} \left(\frac{\hat{\sigma}_k^{2(X)} / (\hat{f}_k^{(X)})^2}{\sum_{j=1}^{n-k} w_{j,k}^{(X)} \bar{X}_{j,k}^\alpha} + \frac{\hat{\sigma}_k^{2(X)} / (\hat{f}_k^{(X)})^2}{\sum_{j=1}^{n-k} w_{j,k}^{(X)} \bar{X}_{j,k}^\alpha} \times \frac{\hat{\sigma}_k^{2(N)} / (\hat{f}_k^{(N)})^2}{\sum_{j=1}^{n-k} w_{j,k}^{(N)} N_{j,k}^\alpha} + \frac{\hat{\sigma}_k^{2(N)} / (\hat{f}_k^{(N)})^2}{\sum_{j=1}^{n-k} w_{j,k}^{(N)} N_{j,k}^\alpha} \right)$$

Nous remarquons que l'erreur de prédiction se décompose bien comme une somme de l'erreur de processus et de l'erreur d'estimation :

- L'erreur de processus peut se voir comme la somme des erreurs de processus sur les nombres (en rouge) et sur les coûts moyens (en bleu) ;
- L'erreur d'estimation est quant à elle la somme des erreurs d'estimations plus un terme d'erreur croisé. Le terme croisé reflète le fait qu'une erreur d'estimation sur les nombres (respectivement les coûts moyens) influe sur l'estimation du nombre de tardifs (respectivement du coût moyen) mais aussi le résultat global qui est un produit des deux estimations.

Il est également possible de calculer l'erreur globale :

$$MSEP(\hat{R}) = MSEP\left(\sum_{i=1}^n \hat{R}_i\right) = Var\left(\sum_{i=1}^n C_{i,n} | D\right) + \left(E\left(\sum_{i=1}^n C_{i,n} | D\right) - \sum_{i=1}^n \hat{C}_{i,n}\right)^2$$

L'indépendance des années de survenance entraîne :

$$Var\left(\sum_{i=1}^n C_{i,n} | D\right) = \sum_{i=1}^n Var(C_{i,n} | D).$$

D'autre part

$$\begin{aligned} \left(E\left(\sum_{i=1}^n C_{i,n} | D\right) - \sum_{i=1}^n \hat{C}_{i,n}\right)^2 &= \sum_{i,l} (E(C_{i,n} | D) - \hat{C}_{i,n}) (E(C_{l,n} | D) - \hat{C}_{l,n}) \\ &= \sum_{i,l} N_{i,n-i+1} \bar{X}_{i,n-i+1} (f_{n+1-i} \dots f_{n-1} - \hat{f}_{n+1-i} \dots \hat{f}_{n-1}) \\ &\quad \times N_{l,n-l+1} \bar{X}_{l,n-l+1} (f_{n+1-l} \dots f_{n-1} - \hat{f}_{n+1-l} \dots \hat{f}_{n-1}) \end{aligned}$$

D'où

$$\begin{aligned} MSEP(\hat{R}) &= \sum_{i=1}^n Var(C_{i,n} | D) + \sum_{i,l} N_{i,n-i+1} \bar{X}_{i,n-i+1} (f_{n+1-i} \dots f_{n-1} - \hat{f}_{n+1-i} \dots \hat{f}_{n-1}) \\ &\quad \times N_{l,n-l+1} \bar{X}_{l,n-l+1} (f_{n+1-l} \dots f_{n-1} - \hat{f}_{n+1-l} \dots \hat{f}_{n-1}) \\ &= \sum_{i=1}^n MSEP(\hat{R}_i) + 2 \sum_{i < l} N_{i,n-i+1} \bar{X}_{i,n-i+1} F_i \times N_{l,n-l+1} \bar{X}_{l,n-l+1} F_l \end{aligned}$$

en utilisant le fait que $MSEP(\hat{R}_i) = Var(C_{i,n} | D) + N_{i,n-i+1}^2 \bar{X}_{i,n-i+1}^2 F_i^2$

avec $F_i = f_{n+1-i} \dots f_{n-1} - \hat{f}_{n+1-i} \dots \hat{f}_{n-1}$ et $F_l = f_{n+1-l} \dots f_{n-1} - \hat{f}_{n+1-l} \dots \hat{f}_{n-1}$.

Comme précédemment, il n'est pas possible de juste remplacer les f_j par leurs estimateurs car sinon cela donnerait 0. Similairement aux calculs précédents pour calculer F^2 , nous allons donner un estimateur de $F_i F_l$. Pour $i < l$:

$$\begin{aligned}\widehat{F}_i \widehat{F}_l &= \sum_{k=n+1-i}^{n-1} \hat{f}_{n+1-l} \dots \hat{f}_{n-i} \hat{f}_{n+1-i}^2 \dots \hat{f}_{k-1}^2 \text{Var}(\hat{f}_k | B_k) \hat{f}_{k+1}^2 \dots \hat{f}_{n-1}^2 \\ &= \sum_{k=n+1-i}^{n-1} \hat{f}_{n+1-l} \dots \hat{f}_{n-i} \frac{\text{Var}(\hat{f}_k | B_k)}{\hat{f}_k^2} \hat{f}_{n+1-i}^2 \dots \hat{f}_{k-1}^2 \hat{f}_k^2 \hat{f}_{k+1}^2 \dots \hat{f}_{n-1}^2 \\ &= \sum_{k=n+1-i}^{n-1} \hat{f}_{n+1-l} \dots \hat{f}_{n-i} \frac{\text{Var}(\hat{f}_k | B_k)}{\hat{f}_k^2} \hat{f}_{n+1-i} \dots \hat{f}_{n-1}\end{aligned}$$

où

$$\text{Var}(\hat{f}_k | B_k) = \left(\frac{\hat{\sigma}_k^{2(X)} / (\hat{f}_k^{(X)})^2}{\sum_{j=1}^{n-k} w_{j,k}^{(X)} \bar{X}_{j,k}^\alpha} + \frac{\hat{\sigma}_k^{2(X)} / (\hat{f}_k^{(X)})^2}{\sum_{j=1}^{n-k} w_{j,k}^{(X)} \bar{X}_{j,k}^\alpha} \times \frac{\hat{\sigma}_k^{2(N)} / (\hat{f}_k^{(N)})^2}{\sum_{j=1}^{n-k} w_{j,k}^{(N)} N_{j,k}^\alpha} + \frac{\hat{\sigma}_k^{2(N)} / (\hat{f}_k^{(N)})^2}{\sum_{j=1}^{n-k} w_{j,k}^{(N)} N_{j,k}^\alpha} \right)$$

Ainsi :

$$MSEP(\widehat{R}) = \sum_{i=1}^n MSEP(\widehat{R}_i) + 2 \sum_{i < l} N_{i,n-i+1} \bar{X}_{i,n-i+1} N_{l,n-l+1} \bar{X}_{l,n-l+1} \widehat{F}_i \widehat{F}_l$$

4.3.4 Application aux données

Dans la section [3.3](#), nous avons calculé les provisions relatives aux données françaises de la classe « Exceptionnelle » grâce à la méthode des coûts moyens. Nous avons proposé deux versions, l'une prenant en compte l'ensemble des données du triangle des coûts moyens et l'autre uniquement une sélection des 6 dernières années calendaires.

Le calcul de l'erreur de prédiction va nous permettre de choisir la version la plus appropriée du point de vue de la volatilité.

Pour la première version, nous obtenons une erreur de 10,38% alors que pour la seconde, l'erreur est de 13,21%. La principale différence se situe sur l'erreur d'estimation. En effet, tout comme pour la méthode Chain Ladder stochastique, enlever de l'information a pour conséquence d'augmenter l'erreur d'estimation. Lorsqu'il n'est pas nécessaire d'utiliser des pondérations pour valider les hypothèses des méthodes, il vaut donc mieux garder toutes les années pour conserver le maximum d'information.

Voici les détails des résultats obtenus :

	Charge Ultime	Charge connue	Montant des IBNR	Erreur de processus	Erreur d'estimation	MSE	RMSE	% d'erreur
1996	134	134	-	-	-	-	-	0%
1997	505	505	1	0	0	0	0,64	8%
1998	310	310	1	0	0	1	0,77	2%
1999	170	170	0	0	0	0	0,51	8%
2000	380	385	6	29	8	37	6,08	8%
2001	288	293	5	22	5	27	5,24	-223%
2002	292	292	0	119	18	137	11,69	83%
2003	131	128	3	159	14	173	13,14	218%
2004	481	464	17	770	200	970	31,14	65%
2005	388	375	13	726	132	858	29,28	132%
2006	383	369	14	636	130	766	27,68	27%
2007	411	405	7	738	163	900	30,00	48%
2008	271	267	4	732	87	818	28,61	129%
2009	490	476	13	1 592	293	1 885	43,42	27%
2010	625	619	6	3 444	578	4 022	63,42	27%
2011	484	474	10	2 796	357	3 153	56,15	47%
2012	684	668	16	5 112	758	5 870	77	26%
2013	486	473	13	2 739	414	3 153	56	42%
2014	457	445	12	4 910	391	5 301	73	43%
2015	241	232	9	1 753	119	1 873	43	39%
2016	459	443	16	5 277	460	5 737	76	27%
2017	381	365	16	4 210	349	4 559	68	24%
2018	516	495	21	6 944	702	7 646	87	20%
2019	519	489	30	7 494	751	8 245	91	20%
2020	374	347	27	4 962	474	5 437	74	20%
2021	413	316	97	10 235	1 172	11 407	107	26%
Total	10 274	9 938	336	65 401	95 145	160 546	401	10,38%

Figure 4.12 : Calcul de l'erreur de prédiction pour la méthode des coûts moyens sans pondération

	Charge Ultime	Charge connue	Montant des IBNR	Erreur de processus	Erreur d'estimation	MSE	RMSE	% d'erreur
1996	134	134	-	-	-	-	-	0%
1997	505	505	1	0	0	0	0,64	8%
1998	310	310	1	0	0	1	0,77	2%
1999	170	170	0	0	0	0	0,51	8%
2000	380	385	6	29	8	37	6,08	8%
2001	288	293	5	22	5	27	5,24	-223%
2002	292	292	1	135	22	157	12,53	85%
2003	133	128	5	209	25	235	15,32	191%
2004	483	464	19	942	340	1 282	35,81	72%
2005	389	375	14	879	222	1 102	33,19	143%
2006	383	369	14	755	216	971	31,16	30%
2007	412	405	8	802	261	1 063	32,61	51%
2008	273	267	6	626	122	748	27,35	111%
2009	494	476	18	1 356	414	1 770	42,07	25%
2010	609	619	10	3 793	1 046	4 839	69,56	31%
2011	479	474	5	3 066	663	3 728	61,06	54%
2012	667	668	2	5 744	1 459	7 203	85	30%
2013	467	473	6	3 709	978	4 686	68	60%
2014	456	445	11	6 880	1 027	7 908	89	52%
2015	241	232	9	2 213	305	2 518	50	45%
2016	461	443	18	6 351	1 173	7 525	87	31%
2017	371	365	6	4 610	821	5 431	74	27%
2018	504	495	10	6 699	1 569	8 268	91	21%
2019	511	489	22	6 956	1 641	8 597	93	21%
2020	379	347	33	4 138	980	5 118	72	19%
2021	425	316	108	8 739	1 829	10 568	103	24%
Total	10 218	9 938	280	68 655	184 095	252 749	503	13,21%

Figure 4.13 : Calcul de l'erreur de prédiction pour la méthode des coûts moyens avec pondération

Conclusion sur la méthode du coût moyen stochastique

Cette méthode donne des résultats très proche des résultats de ceux de la méthode Chain Ladder et cela n'est pas surprenant puisqu'elle se base sur les mêmes cadences. Les erreurs des deux méthodes sont du même ordre de grandeur et, dans notre cas, la version pondérée est moins adaptée que la version non pondérée puisque, sans nécessité du fait du respect des hypothèses Chain Ladder, enlever de l'information conduit à augmenter l'erreur d'estimation.

4.4 Modèle factoriel stochastique

4.4.1 Une version stochastique de la méthode de De Vylder : le modèle de Christofides

Pour faciliter la résolution de la méthode proposée par De Vylder et pour pouvoir calculer une erreur de prédiction, Christofides propose en 1989 d'étudier le modèle linéaire log-normal. Dans ce modèle, sous la condition que les $S_{i,j}$ soient indépendants, les logarithmes des incréments sont supposés suivre une loi normale :

$$\log(S_{i,j}) \sim \mathcal{N}(x_i + y_j; \sigma^2)$$

Une fois les paramètres estimés à la suite d'une régression linéaire sur les deux facteurs x_i et y_j , il est possible d'estimer les provisions. Cependant, les estimations faites à partir de la réécriture ci-dessus ne peuvent néanmoins pas être utilisées directement.

En effet, dans le cas où les données suivent une loi log-normale :

$$E(S_{i,j}) = \exp\left(x_i + y_j + \frac{\sigma^2}{2}\right)$$

Or, dans le cas du passage au logarithme, l'espérance de la loi normale donne :

$$\exp(E[\log(S_{i,j})]) = \exp(x_i + y_j)$$

Pour que les estimations obtenues à partir du modèle sur les logarithmes des incréments convergent vers le modèle log-normal, il faut multiplier les estimations par $\exp\left(\frac{\sigma^2}{2}\right)$.

Nous avons utilisé le triangle de charges des données françaises de la classe « Large » pour la version déterministe mais du fait de l'utilisation d'un logarithme, il n'est pas possible de regarder des montants incrémentaux négatifs. Comme le triangle de charges que nous utilisons contiennent plus de 70% d'incrémentaux négatifs, il faut appliquer la méthode sur le triangle des règlements. Le triangle des règlements ne contient lui que 4 valeurs négatives proches de 0. Pour continuer d'utiliser la méthode même en présence d'incrémentaux négatifs, plusieurs solutions peuvent être envisagées :

- Ajouter une constante suffisamment grande à l'ensemble de nos données pour ne plus avoir d'incrémentaux négatifs,
- Décaler les charges positives vers une période adjacente qui contient des incréments négatifs,
- Mettre les valeurs négatives à 1 pour que lors du passage au logarithme cela donne 0 (méthode qui a été retenue dans notre approche).

Aucune de ces méthodes n'est idéale et dans le cas d'une surreprésentation d'incrémentaux négatifs, il est préférable de choisir une autre méthode qui n'impose pas ces restrictions.

Voici le triangle incrémental des paiements où les valeurs en rouge ont été retraitées et remplacées par des 1.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1996	7,83	13,05	11,57	16,87	4,49	22,15	13,01	5,89	5,87	3,93	1,93	2,19	0,80	0,15	3,27	0,58	0,02	0,05	0,50	0,48	0,00	0,45	0,06	1,17	0,18	1,50
1997	8,13	13,62	12,41	19,96	15,56	15,97	7,14	4,61	6,82	4,30	3,61	7,69	0,36	1,87	0,00	0,28	1,61	1,10	6,29	0,69	2,12	0,23	0,03	1,59	2,11	
1998	7,55	31,94	21,75	25,46	29,49	13,88	12,03	7,88	10,08	8,98	4,98	0,38	6,71	3,25	1,54	3,14	2,39	0,54	0,09	0,05	0,18	1,36	4,55	0,73		
1999	6,30	28,11	26,93	25,91	23,15	9,22	25,39	9,27	10,01	10,01	4,45	5,90	8,43	1,05	2,63	3,19	1,05	3,15	0,69	0,04	2,48	0,36	1,17			
2000	5,98	31,78	16,65	23,89	22,59	19,77	14,64	13,61	13,15	8,64	0,71	6,58	2,51	4,71	5,08	1,09	1,68	0,79	3,29	3,19	4,66	0,07				
2001	11,06	28,87	25,94	19,74	27,80	16,31	6,75	12,11	2,59	0,22	1,87	2,11	1,32	2,02	3,50	0,02	2,16	5,86	1,75	0,14	0,86					
2002	13,07	15,31	13,88	17,24	11,14	23,06	20,26	14,21	5,81	5,48	4,07	3,20	8,55	6,09	1,61	2,13	0,29	3,15	4,63	1,67						
2003	3,20	15,59	23,46	26,11	22,85	6,12	22,91	14,86	6,52	10,13	4,31	1,22	1,80	2,80	4,10	0,87	2,31	3,85	4,06							
2004	4,02	9,33	15,74	9,76	6,38	14,98	12,78	5,67	0,48	3,76	4,70	0,43	0,28	3,55	1,96	0,00	0,16	2,10								
2005	5,56	17,95	26,00	27,55	28,30	10,38	13,25	9,92	7,78	16,72	9,39	3,01	0,81	1,02	1,36	3,60	1,78									
2006	7,68	19,01	13,04	15,94	12,77	10,12	18,37	8,40	19,14	7,74	8,25	2,71	6,96	2,55	0,02	0,83										
2007	6,12	17,58	24,55	10,57	17,02	13,26	10,72	7,43	8,79	6,96	2,41	1,86	0,98	0,00	0,03											
2008	6,88	41,85	17,49	23,95	15,34	25,84	29,29	11,56	11,83	9,57	10,82	4,97	1,87	1,14												
2009	5,90	31,37	15,20	23,50	22,73	23,95	9,64	8,12	4,75	7,12	3,09	2,84	0,10													
2010	5,08	20,47	26,45	15,75	24,93	21,54	19,20	12,94	9,46	14,37	5,13	3,94														
2011	7,58	23,53	24,42	22,51	21,55	12,97	17,36	4,20	11,98	7,39	11,18															
2012	9,17	43,29	27,22	32,39	35,20	25,39	17,25	10,36	7,71	11,88																
2013	5,65	20,37	20,74	25,50	17,64	9,19	24,89	20,81	28,19																	
2014	8,64	20,05	16,83	22,05	23,31	16,59	16,96	9,49																		
2015	2,48	39,96	34,50	25,05	31,46	15,90	15,68																			
2016	5,01	26,47	23,98	30,89	20,69	20,05																				
2017	4,73	28,89	28,08	35,10	25,79																					
2018	2,73	25,84	24,18	24,84																						
2019	2,34	22,94	38,48																							
2020	4,46	18,89																								
2021	0,40																									

Figure 4.14 : Triangle des paiements incrémentaux

Une fois ce retraitement effectué, nous avons appliqué le modèle pour expliquer les log-incréments à partir des périodes de développement et des années de survenance.

Voici un extrait du résultat de la régression :

```

call:
lm(formula = log(X_paid) ~ as.factor(AY) + as.factor(DY), data = dtf_paid)

Residuals:
    Min       1Q   Median       3Q      Max
-7.8107 -0.2798  0.0405  0.4851  2.5575

Coefficients:
(Intercept)           0.6884      0.3242  2.051 0.041152 *
as.factor(AY)1997     0.7906      0.3243  2.438 0.013362 *
as.factor(AY)1998     1.2818      0.3294  3.892 0.000123 ***
as.factor(AY)1999     1.4268      0.3341  4.270 2.63e-05 ***
as.factor(AY)2000     1.4872      0.3390  4.387 1.59e-05 ***
as.factor(AY)2001     1.0840      0.3440  3.151 0.001789 **
as.factor(AY)2002     1.4456      0.3492  4.139 4.53e-05 ***
as.factor(AY)2003     1.3374      0.3549  3.769 0.000197 ***
as.factor(AY)2004     0.1959      0.3609  0.543 0.587738
as.factor(AY)2005     1.3197      0.3675  3.591 0.000384 ***
as.factor(AY)2006     1.0266      0.3746  2.740 0.006507 **
as.factor(AY)2007     0.1475      0.3825  0.386 0.700128
as.factor(AY)2008     1.3868      0.3913  3.544 0.000457 ***
as.factor(AY)2009     0.8073      0.4011  2.013 0.045045 *
as.factor(AY)2010     1.2693      0.4122  3.079 0.002266 **
as.factor(AY)2011     1.1864      0.4249  2.792 0.005572 **
as.factor(AY)2012     1.4420      0.4396  3.280 0.001159 **
as.factor(AY)2013     1.2958      0.4568  2.836 0.004873 **
as.factor(AY)2014     1.1285      0.4775  2.363 0.018742 *
as.factor(AY)2015     1.2025      0.5027  2.392 0.017368 *
as.factor(AY)2016     1.2129      0.5344  2.270 0.023942 *
as.factor(AY)2017     1.2965      0.5758  2.252 0.025061 *
as.factor(AY)2018     0.9854      0.6326  1.558 0.120396
as.factor(AY)2019     0.9733      0.7173  1.357 0.175835
as.factor(AY)2020     0.8452      0.8619  0.981 0.327581
as.factor(AY)2021    -1.6126      1.1944 -1.350 0.177991
as.factor(DY)2        1.3715      0.3243  4.229 3.12e-05 ***
as.factor(DY)3        1.2875      0.3294  3.909 0.000115 ***
as.factor(DY)4        1.3136      0.3341  3.931 0.000105 ***
as.factor(DY)5        1.1786      0.3390  3.477 0.000582 ***
as.factor(DY)6        0.9814      0.3440  2.853 0.004629 **
as.factor(DY)7        0.9766      0.3492  2.796 0.005503 **
as.factor(DY)8        0.4890      0.3549  1.378 0.169253
as.factor(DY)9        0.2644      0.3609  0.733 0.464430
as.factor(DY)10       0.1279      0.3675  0.348 0.728110
as.factor(DY)11      -0.2684      0.3746 -0.716 0.474251 *
as.factor(DY)12      -0.7835      0.3825 -2.048 0.041403 *
as.factor(DY)13      -1.2729      0.3913 -3.253 0.001272 **
as.factor(DY)14      -1.7088      0.4011 -4.250 2.86e-05 ***
as.factor(DY)15      -1.9974      0.4122 -4.846 2.09e-06 ***
as.factor(DY)16      -2.0641      0.4249 -4.858 1.91e-06 ***
as.factor(DY)17      -2.0576      0.4396 -4.681 4.34e-06 ***
as.factor(DY)18      -1.4323      0.4568 -3.135 0.001887 **
as.factor(DY)19      -1.4139      0.4775 -2.961 0.003309 **
as.factor(DY)20      -2.4122      0.5027 -4.798 2.53e-06 ***
as.factor(DY)21      -3.1908      0.5344 -5.971 6.66e-09 ***
as.factor(DY)22      -2.8122      0.5758 -4.884 1.69e-06 ***

```

Figure 4.15 : Résultats de la régression

Nous pouvons voir que certaines années ne sont pas significatives comme c'est le cas pour les années récentes par exemple. En effet, la p-valeur associée à ces années (dernière colonne des résultats ci-dessus) est supérieure à 5%. Cela signifie que l'hypothèse nulle (les coefficients associés à ces années sont égaux à 0) n'est pas rejetée.

Le test réalisé pour tester la significativité des coefficients est le test du t de Student. Comme la loi de Student est symétrique, on cherche à savoir si la t -valeur (avant dernière colonne dans une figure ci-dessus) est comprise dans l'intervalle $[-t_{\alpha/2}; t_{\alpha/2}]$. Si elle est effectivement dans l'intervalle, on ne rejette pas l'hypothèse nulle. Cet intervalle contient 95% des valeurs d'une loi de Student car dans notre cas $\alpha = 5\%$.

Prenons l'exemple de l'année 2021 : la valeur du coefficient est de -1.6126 mais la p -valeur vaut 0.17799 ce qui signifie que la probabilité qu'une loi de Student prenne des valeurs supérieures à $|-1.35|$ vaut 18%. La t -valeur est donc bien une réalisation d'une loi de Student et on ne rejette pas l'hypothèse nulle : on ne peut donc pas exclure le fait que le coefficient soit en réalité égal à 0. C'est uniquement lorsque la p -valeur est inférieure à 5% que la t -valeur n'appartient pas aux valeurs que peut prendre la loi de Student et que le coefficient est significatif.

Le R^2 ajusté est de 60%, ce qui montre que la régression est moyennement adaptée à nos données. Pour bien mesurer la qualité d'ajustement de la droite de régression, il faut étudier les résidus. Les résidus doivent :

- Être homoscédastiques
- Être indépendants
- Suivre une loi normale

Voici les graphiques qui permettent d'étudier les résidus :

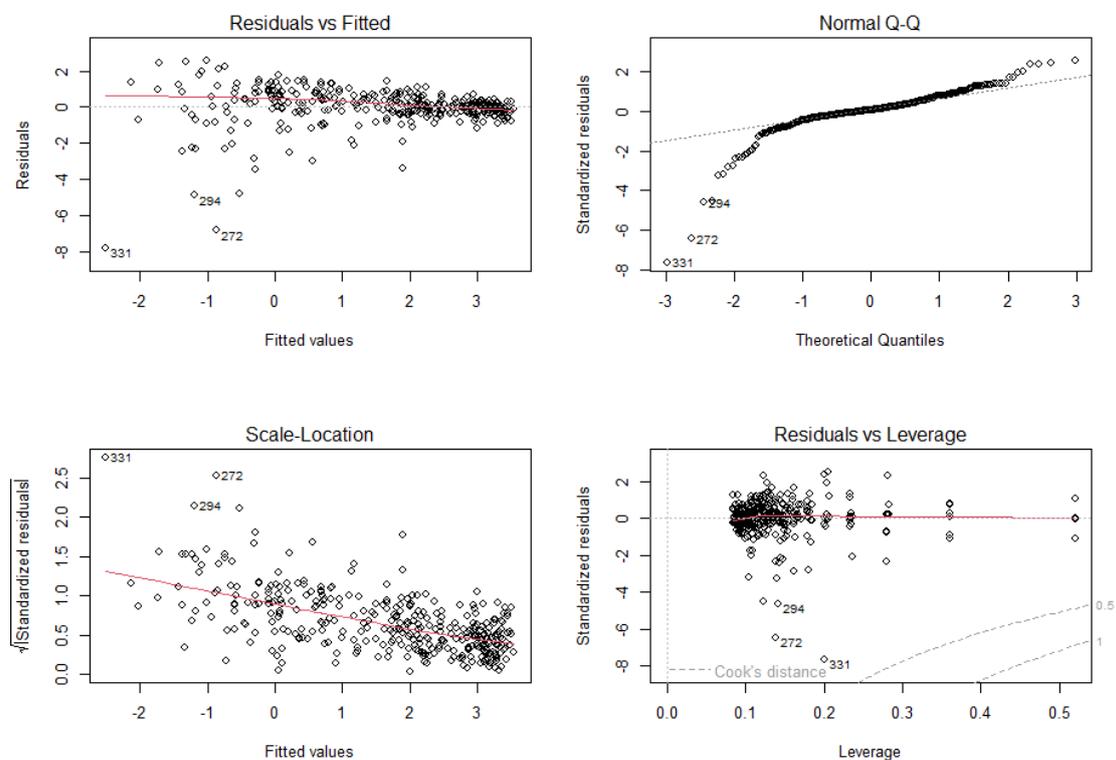


Figure 4.16 : Résidus du modèle

Le premier graphique (en haut à gauche) permet de vérifier l'hypothèse d'homoscédasticité. Comme les points sont sans structure non aléatoire et répartis assez équitablement (sauf les points 294, 272 et 331 qui sont atypiques), l'hypothèse est vérifiée. En moyenne, la droite de régression est bien adaptée aux données et l'hypothèse d'une relation linéaire est acceptable.

Le deuxième graphique (en haut à droite) permet de vérifier l'hypothèse de normalité. Lorsque la distribution des erreurs suit une loi normale, les points sont situés sur la bissectrice du QQ-plot qui représente les quantiles de la distribution des erreurs standardisées en fonction des quantiles d'une loi normale centrée réduite. Ici, l'hypothèse n'est pas vérifiée puisque certains points sont très éloignés de la bissectrice, notamment sur la gauche. La distribution est donc centrée en zéro mais asymétrique. Ce graphique peut parfois être compliqué à interpréter en fonction de l'échelle choisie. En réalisant le test de Shapiro-Wilk, l'hypothèse de normalité peut aussi être testée et dans notre cas, l'hypothèse est toujours rejetée puisque la p-valeur est inférieure à 5%.

Le troisième graphique « Scale - Location » (en bas à gauche) permet de tester l'indépendance. Si les points sont sans structure non aléatoire apparente alors ils peuvent être considérés comme indépendants. Dans notre exemple, la courbe rouge est légèrement descendante donc la dispersion des résidus n'est pas homogène et ceci est dû aux points extrêmes.

Enfin, le dernier graphique (en bas à droite) permet d'évaluer l'influence des données sur les paramètres de régression grâce à la distance de Cook. Cette distance permet de mesurer l'impact du changement dans l'estimation lorsqu'une donnée n'est pas prise en compte et ainsi détecter les points extrêmes (ou *outliers*). Le point 331 est le plus extrême. Il s'agit du paiement d'un sinistre de 1996 réalisé après 21 périodes de développement. Ce paiement est très proche de 0 (il vaut 0,000033) alors que ses voisins sont aux alentours de 0,45.

Notre modèle ne vérifie pas exactement toutes les hypothèses à cause de certains points extrêmes. Utiliser le modèle pour compléter le triangle incrémental des paiements pour déterminer les montants ultimes et les provisions totales ne sera donc pas optimal. Nous ne retiendrons pas cette méthode pour la suite du mémoire.

Conclusion sur l'extension stochastique de la méthode de De Vylder proposée par Christofides

Cette méthode permet d'obtenir facilement des estimations pour la partie basse du triangle mais il n'est pas possible d'appliquer le passage au logarithme dans le cas d'incrément négatifs. Cette limitation est assez importante puisqu'il est très courant d'avoir des valeurs négatives (liées à l'encaissement des recours dans un triangle de règlements ou aux dégagements de PSAP dossier/dossier dans un triangle de charges par exemple).

4.4.2 Méthode de Verbeek-Taylor stochastique : le Bootstrap paramétrique

La méthode de séparation, initialement formulée en 1977 dans un cadre déterministe, n'a connu une extension stochastique que des années plus tard, en 2010, grâce aux travaux de S. Björkwall, O. Hossjer et E. Ohlsson proposant une approche basée sur un Bootstrap paramétrique de la méthode de séparation.

Dans cette section, nous allons utiliser les notations suivantes :

- Δ désigne le triangle inférieur que nous cherchons à compléter
- ∇ désigne le triangle supérieur contenant nos observations
- $\hat{n}_{i,j}$ désigne l'estimateur de Chain Ladder du nombre de sinistres $n_{i,j} = E(N_{i,j})$
- N_i désigne le nombre total de sinistres survenu en i et $\hat{N}_i = \sum_{\nabla} N_{i,j} + \sum_{\Delta} \hat{n}_{i,j}$
- $Z_{i,j,l}$ désigne le montant payé durant l'année calendaire k ($k = i + j$) pour le $l^{\text{ème}}$ sinistre individuel survenu en i et $E(Z_{i,j,l}|N_i) = y_j \lambda_k$.
- $Z_{i,j}$ désigne les paiements incrémentaux payés en k pour des sinistres survenus en i avec $Z_{i,j} = \sum_{l=1}^{N_i} Z_{i,j,l}$

La méthode du Bootstrap paramétrique repose sur deux hypothèses :

- Les $N_{i,j}$ représentant le nombre de sinistres sont indépendants et suivent une loi de Poisson

$$N_{i,j} \in \mathcal{P}(n_{i,j})$$

- Les montants de paiements individuels $Z_{i,j,l}|N_i$ sont indépendants et suivent une loi Gamma

$$Z_{i,j,l}|N_i \in \Gamma\left(\frac{1}{\phi}, y_j \lambda_k \phi\right) \Leftrightarrow Z_{i,j}|N_i \in \Gamma\left(\frac{N_i}{\phi}, y_j \lambda_k \phi\right)$$

Cela permet d'assurer que

$$E(Z_{i,j}|N_i) = N_i y_j \lambda_k$$

$$Var(Z_{i,j}|N_i) = \phi N_i (y_j \lambda_k)^2 = \frac{\phi E(Z_{i,j}|N_i)^2}{N_i}$$

avec ϕ le paramètre de dispersion au sens de la famille exponentielle notamment utilisé dans le cadre des modèles linéaires généralisés.

L'estimateur de ϕ est donné par (estimateur du Chi2) : $\hat{\phi} = \frac{1}{|\nabla| - q} \sum_{\nabla} \hat{N}_i \frac{(Z_{i,j} - \hat{N}_i y_j \hat{\lambda}_k)^2}{(\hat{N}_i y_j \hat{\lambda}_k)^2}$

où $|\nabla| = \frac{(n+1)(n+2)}{2}$ est le nombre d'observations du triangle supérieur et $q = 2n + 1$ est le nombre de paramètres qui devront être estimés par la méthode (n paramètres y_j pour $j = 0, 1, \dots, n - 1$ et $(n + 1)$ paramètres λ_k pour $k = 0, \dots, n$).

Dans la suite, l'index * désigne les variables aléatoires du triangle supérieur et l'index ** désigne les variables aléatoires sur le triangle inférieur. La méthode nécessite de suivre différentes étapes :

- **Etape 1**

Nous disposons d'un triangle de paiements (aucune restriction n'empêche d'utiliser le triangle des charges) et d'un triangle de nombres. Sur le triangle de nombres, la méthode Chain Ladder permet d'obtenir l'estimation $\hat{n}_{i,j}$ du nombre de sinistres par année de survenance et période de développement.

La méthode de séparation classique est ensuite appliquée sur le triangle des coûts moyens (triangle de règlements renormalisé par le nombre ultime de sinistres par an). Cela permet d'obtenir une estimation du montant des provisions totales \hat{R} ainsi qu'une estimation de $\hat{y}_j, \hat{\lambda}_k, \hat{\phi}$.

Début de la boucle avec B simulations

- **Etape 2**

Les $N_{i,j}$ sont simulés pour chaque i et j à partir de tirages de la loi de Poisson $\mathcal{P}(\hat{n}_{i,j})$. Cela nous permet d'obtenir le nombre ultime de sinistres par année de survenance donné par :

$$N_i^{**} = \sum_{j \in \nabla} N_{i,j}^* + \sum_{j \in \Delta} N_{i,j}^{**}$$

- **Etape 3**

Une fois le nombre ultime de sinistres N_i^{**} calculé, les $Z_{i,j}^*$ sont simulés en utilisant le fait que :

$$Z_{i,j}^* | N_i^{**} \in \Gamma\left(\frac{N_i^{**}}{\hat{\phi}}, \hat{y}_j \hat{\lambda}_k \hat{\phi}\right)$$

- **Etape 4**

L'estimation Chain Ladder du nombre ultime de sinistres est calculée comme :

$$\hat{N}_i^* = \sum_{j \in \nabla} N_{i,j}^* + \sum_{j \in \Delta} \hat{n}_{i,j}^*$$

Nous avons également les estimateurs \hat{y}_j^* et $\hat{\lambda}_k^*$ en appliquant le modèle de séparation.

Cela nous permet de calculer les montants futurs $\widehat{M}_{i,j}^* = \widehat{N}_i^* \times \widehat{y}_j^* \times \widehat{\lambda}_k^*$ ainsi que le montant des provisions par année $\widehat{R}_i^* = \sum_{j \in \Delta} \widehat{M}_{i,j}^*$. Cela permet également d'obtenir le montant total des provisions $\widehat{R}^* = \sum_{\Delta} \widehat{M}_{i,j}^*$.

- **Etape 5**

Les flux futurs par année de survenance $R_i^{**} = \sum_{j \in \Delta} Z_{i,j}^{**}$ sont ensuite calculés. Similairement, les flux totaux sont donnés par $R^{**} = \sum_{\Delta} Z_{i,j}^{**}$. L'erreur de prédiction non standardisée est donnée par $pe_i = R_i^{**} - \widehat{R}_i^*$ pour l'erreur par année de survenance et par $pe = R^{**} - \widehat{R}^*$ pour l'erreur globale.

Fin de la boucle

- **Etape 6**

La distribution des provisions est approximée par la distribution empirique de

$$\tilde{R}^{**(b)} = \widehat{R} + pe^{(b)}$$

En appliquant la méthode définie précédemment sur les données françaises de la classe « Exceptionnelle », nous allons pouvoir décider de la méthode à adopter entre la méthode utilisant une régression linéaire pour l'extrapolation des effets calendaires et celle utilisant l'historique de l'inflation des données passées. Nous obtenons les distributions suivantes pour nos provisions :

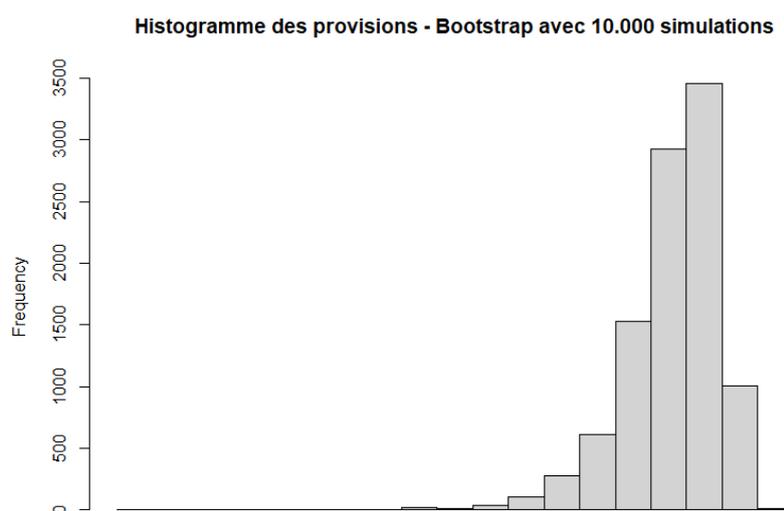


Figure 4.17 : Histogramme des provisions calculées avec l'hypothèse d'inflation historique

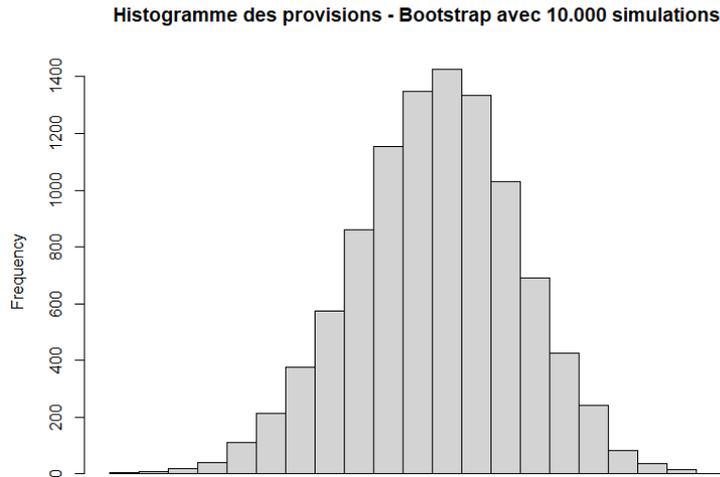


Figure 4.18 : Histogramme des provisions calculées avec l'extrapolation linéaire des lambdas

Le pourcentage d'erreur est de 25% pour les provisions calculées à partir de l'hypothèse d'inflation constante. Cette erreur est largement supérieure au pourcentage d'erreur de 11% obtenu avec la régression linéaire. Il y aura donc plus d'incertitude si nous utilisons l'hypothèse d'inflation constante. Cela ne veut pas forcément dire que l'extrapolation linéaire donnera de meilleurs résultats, simplement que la variabilité de ces résultats sera plus faible.

Dans le cadre de nos données, nous allons retenir la régression linéaire comme méthode d'extrapolation des λ_k .

Conclusion sur la méthode du Bootstrap paramétrique

Le modèle par séparation permet de mesurer l'inflation passée contenue dans le triangle ainsi que la tendance dans laquelle elle s'inscrit. La version stochastique de la méthode, reposant sur le Bootstrap paramétrique, permet par ailleurs de mesurer l'erreur de prédiction ce qui n'était pas possible dans la version déterministe.

Nous avons constaté que l'erreur de prédiction pouvait varier sensiblement selon le choix de la méthode d'extrapolation des λ_k .

De manière générale, l'extrapolation de ces coefficients requiert une attention particulière. L'avantage du modèle est de pouvoir intégrer des scénarios de rupture et ainsi intégrer une hypothèse d'inflation future qui pourrait prendre en compte l'impact de la crise énergétique et de la guerre en Ukraine par exemple.

4.5 Agrégation des méthodes par formule fermée pour le calcul de l'erreur globale

En pratique, les assureurs utilisent souvent des combinaisons de méthodes en associant des méthodes différentes en fonction des exercices de survenance ou d'une segmentation des données (par tranche de sévérité par exemple). S'il est possible d'additionner les montants des provisions calculés à partir de plusieurs méthodes différentes, les MSEP n'étant pas additives, l'erreur de prédiction globale ne peut pas se calculer comme une somme des erreurs de chacune des méthodes. A notre connaissance, aucune formule fermée n'a été développée pour calculer l'erreur de prédiction d'une combinaison de méthodes. Nous nous proposons dans cette section de faire ce calcul.

Nous allons commencer par déterminer l'erreur de prédiction commise dans le cas de l'agrégation de deux méthodes avant d'étendre ce résultat au cas général.

4.5.1 Formule fermée pour le calcul de l'erreur globale pour deux méthodes

L'agrégation la plus courante est la combinaison des méthodes Chain Ladder et Bornhuetter-Ferguson :

- Pour les années récentes (le plus souvent l'exercice courant N et les exercices antérieurs $N-1$ et $N-2$, parfois plus notamment dans le cas d'acceptations en réassurance), les assureurs tendent à préférer utiliser les estimations données par la méthode Bornhuetter-Ferguson (ou par la méthode du loss ratio), notamment sur les branches à développement long, car les estimations sont moins approximatives et ne dépendent pas que de l'information disponible.
- Les estimations données par la méthode Chain Ladder sont généralement utilisées pour les années plus anciennes, là où le développement de la charge connue est suffisant pour conduire à des estimations stables.

De manière plus générale, il est possible d'utiliser une combinaison de deux méthodes au choix. Soit A et B les deux méthodes utilisées. Supposons que sur notre historique de n années, les provisions de k années soient estimées avec la méthode A et les $(n - k)$ restantes avec la méthode B (notés R_A pour la méthode A et R_B pour la méthode B). Pour obtenir le montant global des provisions, il faut calculer $R = R_A + R_B$.

Pour calculer l'erreur globale, nous pouvons utiliser la décomposition suivante :

$$\begin{aligned}MSEP(R) &= MSEP(R_A + R_B) \\&= Var(R_A + R_B|D) + \left(E(R_A + R_B|D) - (\hat{R}_A + \hat{R}_B)\right)^2 \\&= Var(R_A|D) + Var(R_B|D) + 2cov(R_A, R_B|D) + \left(E(R_A|D) + E(R_B|D) - (\hat{R}_A + \hat{R}_B)\right)^2\end{aligned}$$

$$= \text{Var}(R_A|D) + \text{Var}(R_B|D) + 2\text{cov}(R_A, R_B|D) + \\ (E(R_A|D) - \hat{R}_A)^2 + (E(R_B|D) - \hat{R}_B)^2 + 2(E(R_A|D) - \hat{R}_A)(E(R_B|D) - \hat{R}_B)$$

Nous distinguons en bleu le calcul de la MSEP de la méthode Chain Ladder et en rouge la MSEP de la méthode Bornhuetter-Ferguson. Ces deux quantités peuvent être calculées à partir des formules présentées dans les sections 4.1 et 4.2.

$$MSEP(R) = \text{MSEP}(R_A) + \text{MSEP}(R_B) + 2[\text{cov}(R_A, R_B|D) + (E(R_A|D) - \hat{R}_A)(E(R_B|D) - \hat{R}_B)]$$

Il reste à calculer le terme de droite :

- $\text{cov}(R_A, R_B|D) = \rho_{A,B} \sqrt{\text{Var}(R_A|D)} \sqrt{\text{Var}(R_B|D)}$
- $(E(R_A|D) - \hat{R}_A)(E(R_B|D) - \hat{R}_B) \approx E\left(\left(E(R_A|D) - \hat{R}_A\right)\left(E(R_B|D) - \hat{R}_B\right) \middle| D\right) \\ = \text{cov}(\hat{R}_A, \hat{R}_B|D) = \rho_{A,B} \sqrt{\text{Var}(\hat{R}_A|D)} \sqrt{\text{Var}(\hat{R}_B|D)}$

Ainsi :

$$MSEP(R) \approx \text{MSEP}(R_A) + \text{MSEP}(R_B) \\ + 2\rho_{A,B} \left(\sqrt{\text{Var}(\hat{R}_A|D)} \sqrt{\text{Var}(\hat{R}_B|D)} + \sqrt{\text{Var}(R_A|D)} \sqrt{\text{Var}(R_B|D)} \right)$$

La corrélation entre les deux méthodes $\rho_{A,B}$ peut être approximée par les corrélations calculées sur les montant ultimes prédits par chacune des deux méthodes.

4.5.2 Formule fermée pour le calcul de l'erreur globale avec M méthodes

Il est possible d'étendre le résultat précédent dans le cas général où M méthodes sont utilisées.

Nous utilisons le fait que :

$$\text{Var}\left(\sum_{m=1}^M R_m\right) = \sum_{m=1}^M \text{Var}(R_m) + 2 \sum_{1 \leq m < p \leq M} \text{cov}(R_m, R_p)$$

Ainsi :

$$MSEP(R) = MSEP\left(\sum_{m=1}^M R_m\right) = \text{Var}\left(\sum_{m=1}^M R_m | D\right) + \left(E\left(\sum_{m=1}^M R_m | D\right) - \left(\sum_{m=1}^M \hat{R}_m\right)\right)^2 \\ = \sum_{m=1}^M \text{Var}(R_m|D) + 2 \sum_{1 \leq m < l \leq M} \text{cov}(R_m, R_l|D) + \left(\sum_{m=1}^M E(R_m|D) - \left(\sum_{m=1}^M \hat{R}_m\right)\right)^2$$

En utilisant la même décomposition que dans le cas de deux méthodes, nous obtenons dans le cas général :

$$MSEP(R) = \sum_{m=1}^M MSEP(R_m) + 2 \left[\sum_{1 \leq m < l \leq M} cov(R_m, R_l | D) + (E(R_m | D) - \hat{R}_m)(E(R_l | D) - \hat{R}_l) \right]$$

Comme précédemment, la MSEP de chaque méthode peut être calculée et il reste à spécifier le terme de droite :

- Calcul des termes de covariance :

$$\sum_{1 \leq m < l \leq M} cov(R_m, R_l | D) = \sum_{1 \leq m < l \leq M} \rho_{m,l} \sqrt{Var(R_m | D)} \sqrt{Var(R_l | D)}$$

- Calcul des termes d'espérance :

$$\begin{aligned} \sum_{1 \leq m < l \leq M} (E(R_m | D) - \hat{R}_m)(E(R_l | D) - \hat{R}_l) &\approx \sum_{1 \leq m < l \leq M} E \left((E(R_m | D) - \hat{R}_m)(E(R_l | D) - \hat{R}_l) | D \right) \\ &= \sum_{1 \leq m < l \leq M} cov(\hat{R}_l, \hat{R}_m | D) = \sum_{1 \leq m < l \leq M} \rho_{m,l} \sqrt{Var(\hat{R}_l | D)} \sqrt{Var(\hat{R}_m | D)} \end{aligned}$$

Ainsi, $MSEP(R)$ est approximée par :

$$\sum_{m=1}^M MSEP(R_m) + 2 \sum_{1 \leq m < l \leq M} \rho_{l,m} \left(\sqrt{Var(\hat{R}_m | D)} \sqrt{Var(\hat{R}_l | D)} + \sqrt{Var(R_m | D)} \sqrt{Var(R_l | D)} \right)$$

4.5.3 Application aux données

Pour rappel, les données des polices françaises ont été séparées en plusieurs catégories de sinistres :

- La catégorie « Attritional »
- La catégorie « Large »
- La catégorie « Exceptionnelle »

L'avantage de séparer les sinistres est de pouvoir appliquer des méthodes différentes et ainsi tenir compte des spécificités de chacune des catégories dans le développement des charges ultimes. Nous avons calculé pour chaque catégorie le montant de provisions et l'erreur de prédiction associée. Néanmoins, il peut être intéressant de connaître le montant de provisions total et de connaître l'erreur globale commise. Il est aussi pertinent de regarder l'impact de la segmentation.

Nous allons appliquer la formule ci-dessus au cas des données françaises (3 méthodes utilisées). Avant de faire le calcul, il faut sélectionner parmi toutes les méthodes celles que nous souhaitons appliquer sur nos données. Nous retenons comme critère la minimisation de l'erreur de prédiction. Ainsi, nous allons utiliser la méthode du coût moyen pour la classe « Exceptionnelle » et la méthode Bornhuetter-Ferguson pour les classes « Attritional » et « Large ».

D'après la formule ci-dessus, la première étape pour calculer l'erreur globale est d'additionner les MSEP de chacune des méthodes. La deuxième étape est de calculer les coefficients de corrélation. Pour ce faire, nous allons calculer la corrélation entre les montants ultimes de chaque classe. Enfin, la dernière étape est d'identifier les erreurs de prédiction et de processus pour pouvoir appliquer la formule.

- Etape 1 : Calcul de la somme des MSEP

D'après les calculs précédents, nous avons :

	« Attritional »	« Large »	« Exceptionnelle »
MSE	143	32 027	160 546

Ainsi :

$$\sum_{m=1}^3 MSEP(R_m) = 143 + 32\,027 + 160\,546 = 192\,716$$

- Etape 2 : Calcul des corrélations

Les montants ultimes que nous avons prédits peuvent être récapitulés dans le tableau suivant :

	Montant ultime Attritional	Montant ultime Large	Montant ultime Exceptionnelle
1996	85	126	134
1997	70	155	505
1998	80	229	310
1999	84	229	170
2000	82	218	380
2001	87	195	288
2002	86	193	292
2003	78	204	131
2004	66	118	481
2005	64	217	388
2006	62	185	383
2007	63	144	411
2008	63	258	271
2009	65	190	490
2010	70	244	625
2011	70	238	484
2012	70	307	684
2013	67	246	486
2014	64	260	457
2015	64	282	241
2016	67	271	459
2017	70	301	381
2018	69	310	516
2019	72	327	519
2020	59	296	374
2021	71	360	413
Total	1 847	6 105	10 274

Figure 4.19 : Montants ultimes calculés pour chaque jeu de données

Pour étudier la corrélation entre nos trois classes, nous allons utiliser le coefficient de corrélation de Pearson. Pour deux variables X et Y, le coefficient de Pearson se calcule comme :

$$\text{Corr}(X, Y) = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

avec x_i (resp. y_i) les observations de la variable X (resp. Y)
et \bar{x} (resp. \bar{y}) la valeur moyenne des x_i (resp. y_i)

Pour nos données, les coefficients obtenus sont les suivants :

- Corrélation entre les classes « Exceptionnelle » et « Large » : 28,9%
- Corrélation entre les classes « Large » et « Attritional » : -23,3%
- Corrélation entre les classes « Attritional » et « Exceptionnelle » : -46,3%

- Etape 3 : Calcul de l'erreur globale

Les erreurs d'estimation et de processus calculées dans les étapes précédentes sont récapitulées dans le tableau suivant :

	« Attritional »	« Large »	« Exceptionnelle »
Erreur d'estimation	72	24 623	95 145
Erreur de processus	71	7 404	65 401

Ainsi, l'erreur globale est :

$$192\,716 + 2 * \left(-23,3\%(\sqrt{72 * 24\,623} + \sqrt{71 * 7\,404}) \right. \\ \left. + 28,9\%(\sqrt{95\,145 * 24\,623} + \sqrt{65\,401 * 7\,404}) \right. \\ \left. - 46,7\%(\sqrt{72 * 95\,145} + \sqrt{71 * 65\,401}) \right) \approx 227\,980$$

Finalement, le montant des provisions est de 6 281 et l'erreur de prédiction est de 227 980. En pourcentage d'erreur, cela donne 7,6%. Nous profitons de l'effet de diversification, qui dans notre cas vaut 19% :

$$\frac{\sum_{m=1}^3 RMSE(R_m) - RMSE_{globale}}{\sum_{m=1}^3 RMSE(R_m)} = 1 - \frac{\sqrt{227\,980}}{\sqrt{143} + \sqrt{32\,027} + \sqrt{160\,546}} \approx 19\%$$

Si nous avons appliqué la méthode Bornhuetter-Ferguson sur l'ensemble de la base sans séparer nos données, nous aurions obtenu un montant de provisions plus faible car la classe « Attritional » est surreprésentée. Nous en concluons qu'il est important de séparer nos données lorsque celles-ci ne sont pas homogènes pour pouvoir projeter les données au mieux.

5. Comparaison des méthodes

Nous avons pu calculer une estimation de l'erreur de prédiction des méthodes Chain Ladder, Bornhuetter-Ferguson, Coût Moyen et Verbeek-Taylor grâce aux calculs précédents. Voici le récapitulatif des erreurs de prédiction des différentes méthodes utilisées sur chaque jeu de données :

Méthode Chain Ladder :

	Classique	Pondéré
Attritional		15,0%
Large	8,3%	9,8%
Exceptionnelle		13,3%
RC Auto UK		9,1%

Méthode Bornhuetter-Ferguson :

	Pondéré
Attritional	8,2%
Large	7,8%
Exceptionnelle	11,7%
RC Auto UK	6,2%

Méthode du Coût Moyen :

	Classique	Pondéré
Exceptionnelle	10,4%	13,2%

Méthode Verbeek-Taylor :

	Régression Linéaire	Inflation constante
Exceptionnelle	11,1%	25,3%

Nous n'avons pas pu appliquer l'ensemble des méthodes sur toutes nos données car certains triangles ne vérifient pas les hypothèses nécessaires. De plus, certaines méthodes nécessitent d'avoir l'historique du nombre de sinistres (comme la méthode du coût moyen par exemple) mais ces données ne sont pas toujours disponibles.

Nous remarquons que la méthode la plus adaptée (au sens de la minimisation de l'erreur de prédiction) varie en fonction des données mais globalement, la méthode Bornhuetter-Ferguson semble donner de meilleurs résultats. Cette observation est valable sur nos données mais n'est pas forcément toujours vraie. Le but est donc maintenant de comparer les performances des méthodes de manière la plus générale possible.

Les données simulées vont nous permettre de calculer l'erreur réelle puisque nous connaissons déjà les montants ultimes. Nous allons donc pouvoir comparer les méthodes entre

elles et déterminer si certaines méthodes sont plus adaptées que d'autres en fonction des caractéristiques des données.

Nous allons utiliser comme base de référence les données simulées présentées dans le [chapitre 2](#) afin de calculer les erreurs de prédiction associées à chacune des méthodes. Les méthodes étudiées sont :

- Chain Ladder sans pondération (moy_CL_sp)
- Chain Ladder avec pondération (moy_CL_ap)
- Bornhuetter-Ferguson de Mack sans pondération (moy_BF_sp)
- Bornhuetter-Ferguson de Mack avec pondération (moy_BF_ap)
- Coût moyen sans pondération (moy_CM_sp)
- Coût moyen avec pondération (moy_CM_ap)
- Verbeek-Taylor avec extrapolation linéaire (moy_VT)

Voici les erreurs de prédiction observées pour chaque méthode en fonction du nombre de simulations :

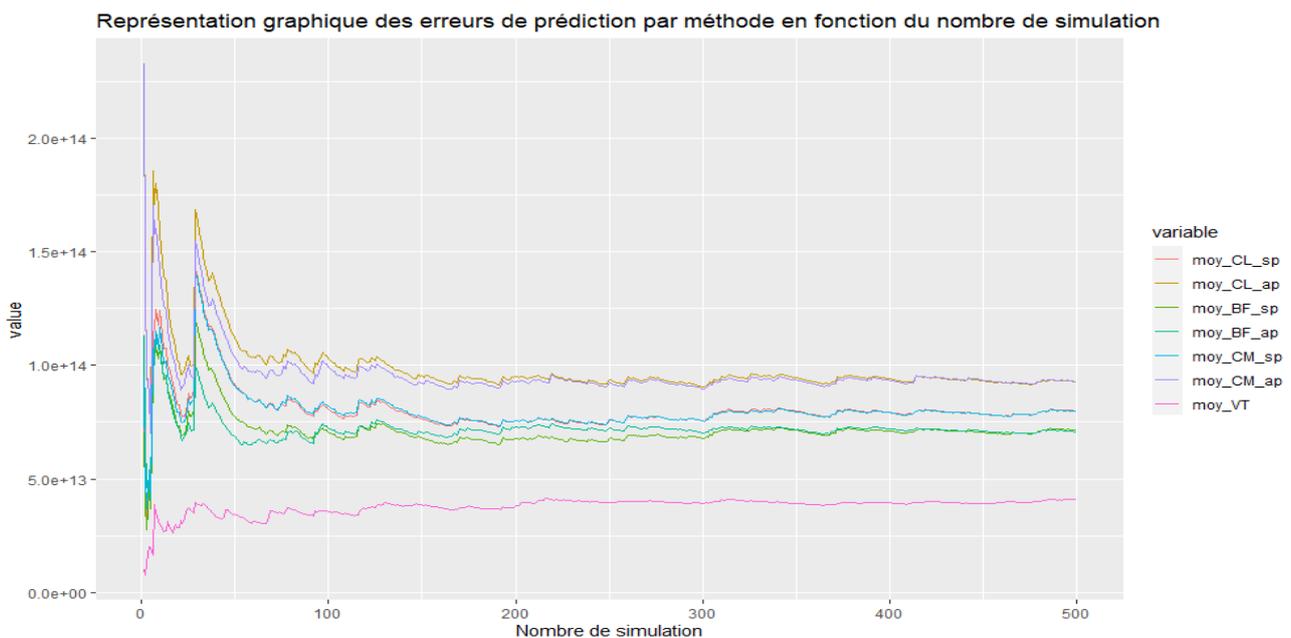


Figure 5.1 : Comparaison des erreurs de prédiction

Sur ce graphique, nous distinguons quatre groupes :

- La méthode de Verbeek-Taylor qui donne une erreur largement plus faible que les autres.
- Les méthodes qui donnent les erreurs de prédiction les plus fortes (Coût Moyen et Chain Ladder pondéré) car, sans nécessité du fait du respect des hypothèses Chain Ladder, enlever de l'information fait augmenter l'erreur.

- Les méthodes qui se situent entre les deux groupes précédents avec les méthodes non pondérées de Chain Ladder et du Coût Moyen d'un côté et les deux méthodes Bornhuetter-Ferguson de Mack de l'autre. Les méthodes Bornhuetter-Ferguson sont légèrement meilleures que la méthode Chain Ladder non pondéré qui se confond avec la méthode du Coût Moyen.

Les observations faites sur les données réelles sont confirmées. Pour conforter nos conclusions, nous allons réaliser différentes analyses de sensibilités. Comme nos données sont simulées, il est possible d'en modifier les caractéristiques pour observer la réaction des méthodes. Les sensibilités que nous allons appliquer vont modifier la fréquence, le coût, les délais de déclaration et de traitement ainsi que l'inflation. Voici les sensibilités qui vont être testées :

Module	Cas central	Sensibilité 1	Sensibilité 2
Fréquence (lambda)	90,6	120,6	290,6
Coût (shape)	0,667	2,667	0,417
Délai de déclaration (moyenne)	2,14	1,14	0,14
Délai de traitement (moyenne)	9,46	7,46	5,46
Inflation future	Constante à 2%	Constante à 5%	Suivant les projections macroéconomiques

Figure 5.2 : Liste des sensibilités testées

5.1 Variation des paramètres de fréquence

Pour faire varier la fréquence (et donc le nombre de sinistres), nous avons tout d'abord modifié le paramètre de la loi de Poisson. Nous avons fait varier notre paramètre initial d'abord de +30 puis de +200. Les courbes obtenues sont les suivantes :

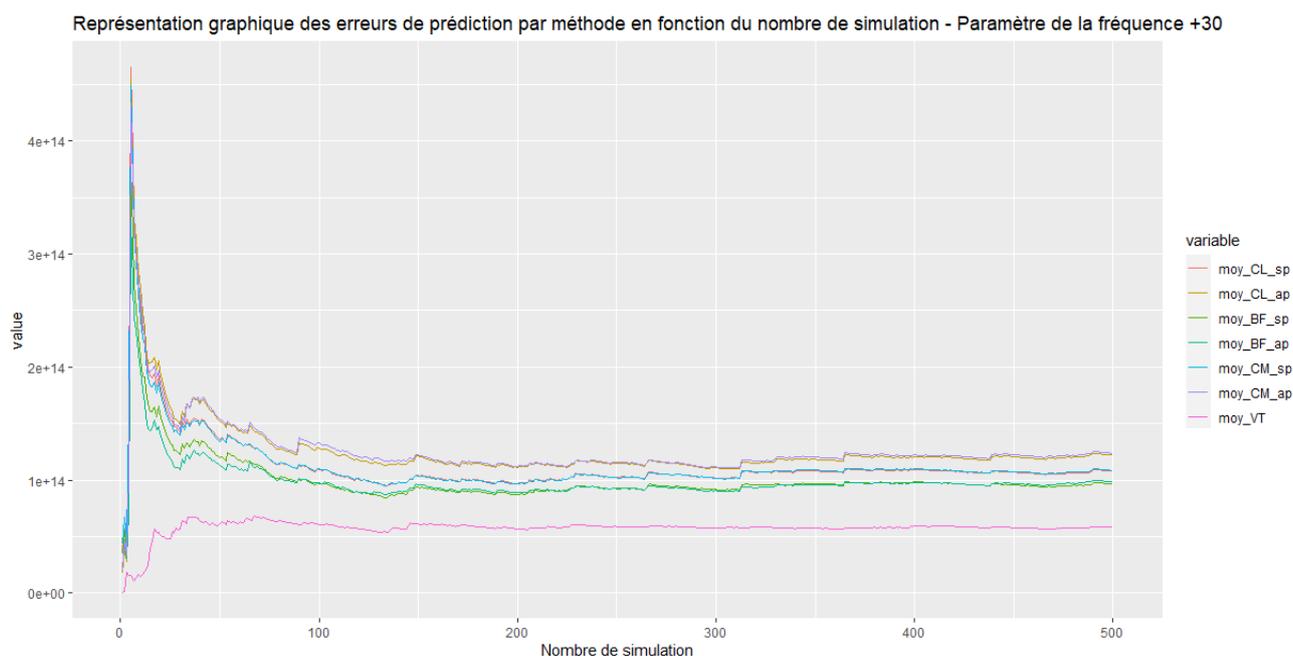


Figure 5.3 : Choc sur la fréquence - Augmentation du nombre de sinistres moyen (+30)

Représentation graphique des erreurs de prédiction par méthode en fonction du nombre de simulation - Paramètre de la fréquence +200

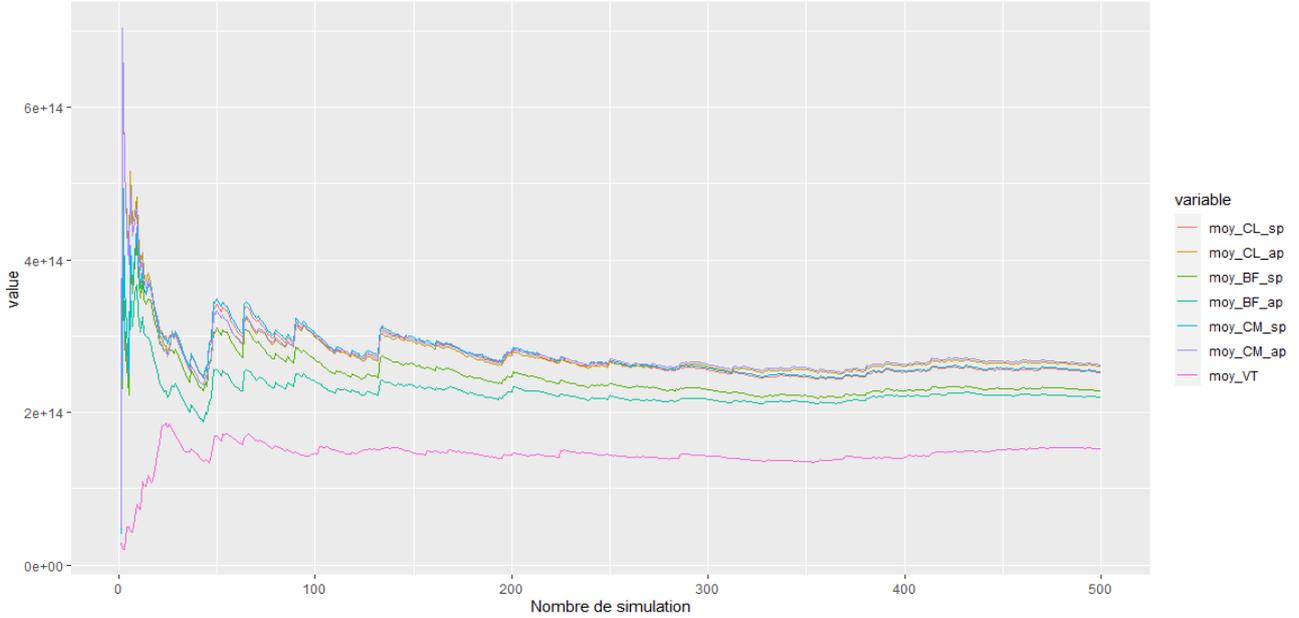


Figure 5.4 : Choc sur la fréquence - Augmentation du nombre de sinistres moyen (+200)

Nous remarquons que l'augmentation du nombre de sinistres conduit à une hausse de toutes les erreurs, augmente la durée de stabilisation (les estimations sont donc plus volatiles) et tend à modifier les groupes identifiés précédemment. En effet, le groupe Bornhuetter-Ferguson (où la version pondérée est cette fois légèrement meilleure que la version non pondérée) se détache plus largement des méthodes Chain Ladder et coût moyen (avec et sans pondération).

Nous concluons que peu importe le paramétrage de la fréquence, la méthode Verbeek-Taylor donne les meilleurs résultats et Bornhuetter-Ferguson arrive en seconde position. De plus, plus la fréquence augmente et plus les erreurs commises sont plus importantes et volatiles.

Cependant, la loi de Poisson a une variance et une espérance identique. Il est donc intéressant d'étudier une sensibilité où ce n'est pas le cas. Nous avons choisi de considérer la loi négative binomiale qui nous permet, avec la même moyenne que dans le cas central, d'avoir une dispersion différente. Voici le récapitulatif des densités qui ont été testées :

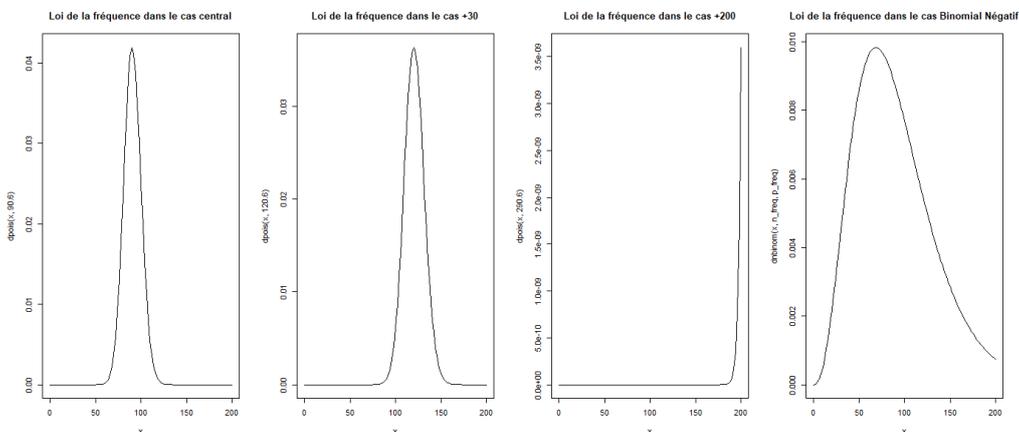


Figure 5.5 : Densités testées pour la fréquence

Nous remarquons que dans le cas d'une dispersion élevée, les méthodes donnent de moins bons résultats car le niveau de l'erreur est plus élevé. De plus, l'écart entre les méthodes se resserre démontrant qu'aucune méthode n'arrive à bien prédire les montants ultimes comme l'illustre le graphique ci-dessous.

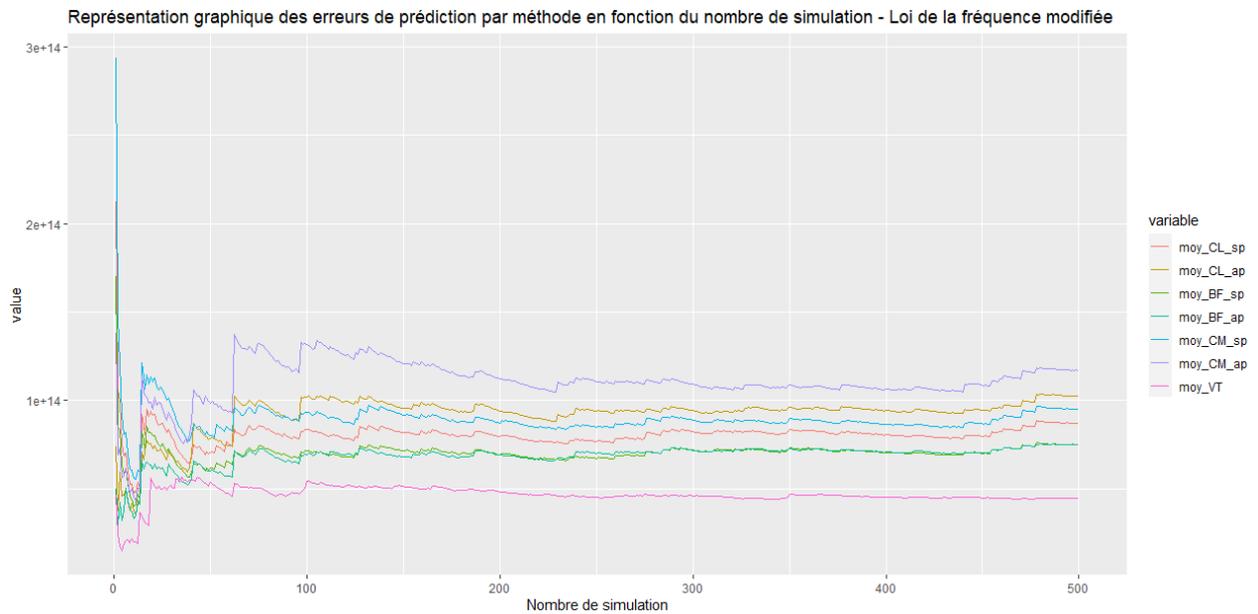


Figure 5.6 : Variation de la loi de la fréquence - cas d'une loi binomiale négative

La fréquence a un impact sur le niveau de l'erreur ainsi que sur la variabilité mais pas sur la hiérarchie des méthodes.

5.2 Variation des paramètres de coûts

Après avoir étudié l'impact des paramètres de la fréquence sur les erreurs des méthodes, nous allons faire de même pour les coûts. Nous allons conserver la même loi et modifier uniquement le paramètre de forme car c'est la variabilité qui influe sur la performance du modèle. Voici une illustration des paramètres de forme étudiés sur l'intervalle [0 ; 300 000] :

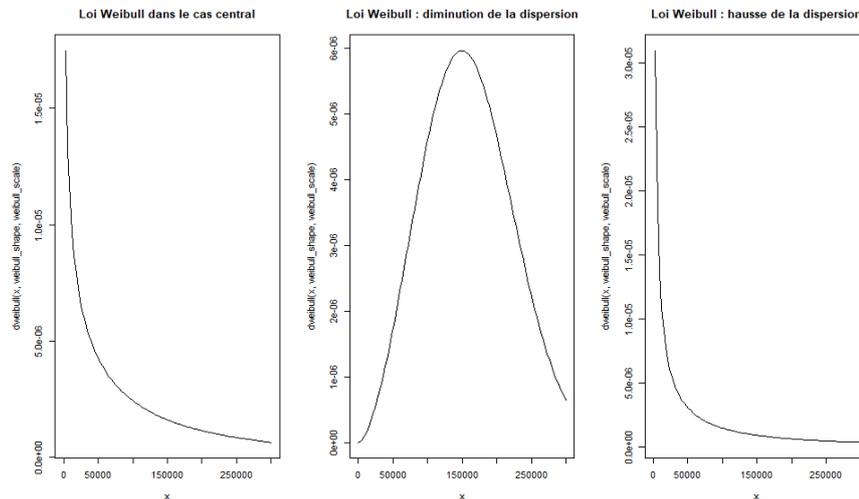


Figure 5.7 : Densité de la loi Weibull en fonction des paramètres

- Dans la version centrale, le paramètre de forme est inférieur à 1 ce qui signifie que la probabilité décroît de manière exponentielle depuis l'infini (figure de gauche). Nous aurons dans ce cas beaucoup de sinistres avec un montant peu élevé et peu de sinistres avec un montant très élevé.
- Dans le cas où le paramètre de forme est proche de 2 (figure du milieu), la loi tend à être presque symétrique et nous observons peu de montant très faible ou très élevé.
- Dans le cas où le paramètre est proche de 0 (figure de droite), les montants de sinistres seront majoritairement très faibles avec une variabilité accrue.

Voici les résultats obtenus :

Représentation graphique des erreurs de prédiction par méthode en fonction du nombre de simulation - Variation du coût : diminution de la dispersion

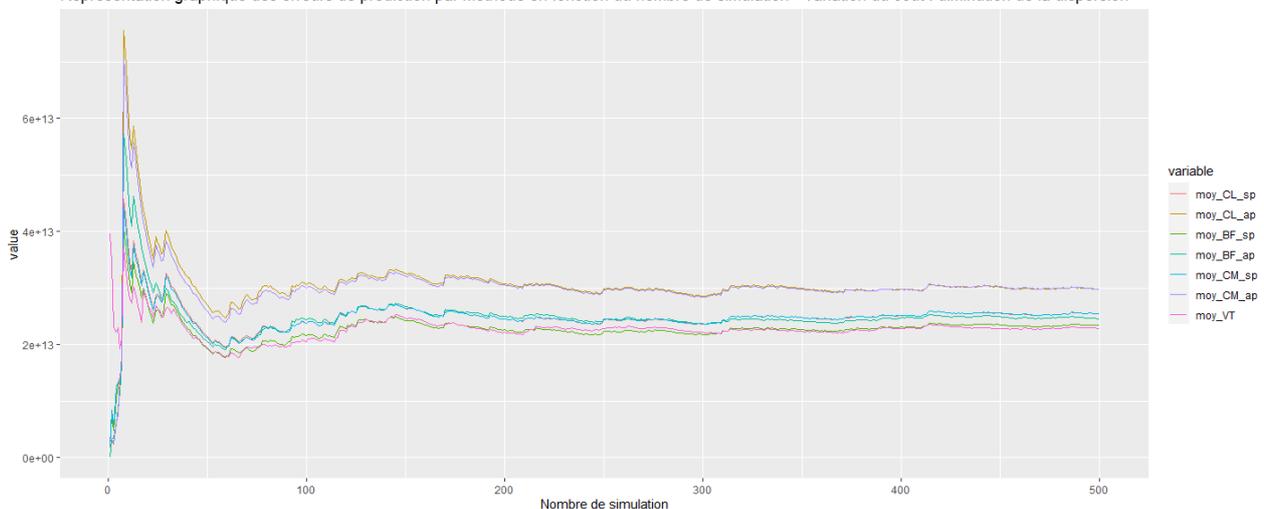


Figure 5.8 : Variation du paramètre de forme - Diminution de la dispersion

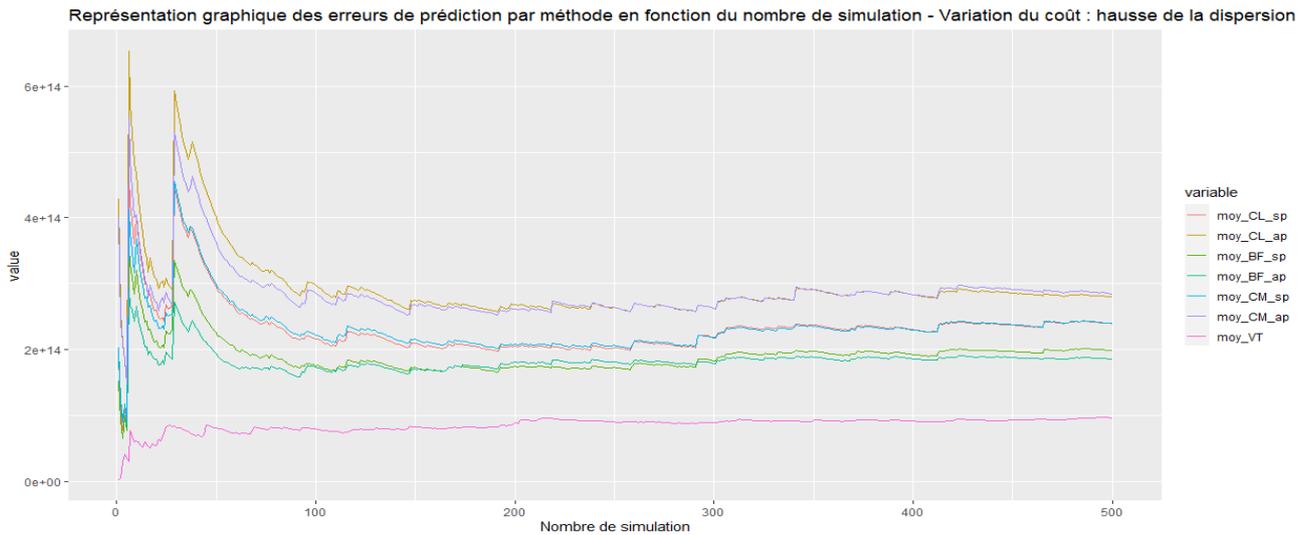


Figure 5.9 : Variation du paramètre de forme - Hausse de la dispersion

Nous remarquons que les méthodes Chain Ladder, Bornhuetter-Ferguson et coût moyen donnent de meilleures approximations lorsque le paramètre de forme est élevé (erreur inférieure à $2e13$ alors que dans la version centrale, les erreurs étaient de l'ordre de $1e14$).

La méthode de Verbeek-Taylor est quant à elle moins sensible puisque qu'une baisse de la dispersion conduit à une erreur d'environ $2e13$ alors que dans la version centrale, l'erreur se situait entre $3e13$ et $4e13$. La diminution de l'erreur n'est donc que de 40% environ alors que pour les autres méthodes, l'erreur est divisée par 5. En cas d'une hausse de la dispersion, la méthode de Verbeek-Taylor va également être un peu moins impactée puisque l'erreur est multipliée par un peu plus de 2 contre environ 3 pour les autres méthodes.

Nous concluons que la méthode de Verbeek-Taylor est moins sensible aux variations du paramètre de forme. La méthode Bornhuetter-Ferguson sans poids donne, dans le cas d'un paramètre de forme élevé, les meilleurs résultats. Cela est en cohérence avec les observations faites sur nos données réelles. Nous ne pouvons pas appliquer la méthode de Verbeek-Taylor sur toutes les classes puisque nous ne disposons pas du nombre de sinistres mais il aurait été possible que Verbeek-Taylor donne les meilleurs résultats sur des jeux de données où les montants de sinistres sont très dispersés.

5.3 Variation de la durée de déclaration

La durée de déclaration est le délai entre la date de survenance d'un sinistre et sa déclaration à l'assureur. Un délai de déclaration long, contribue ainsi à rallonger le développement de la charge de sinistres.

Deux sensibilités ont été testées, l'un en réduisant de 1 an le délai de déclaration moyen et l'autre en le réduisant de 2 ans (dans la version centrale, le délai de déclaration moyen était de 2,41 ans). Voici ce qu'une réduction des délais de déclaration implique sur les erreurs de prédictions :

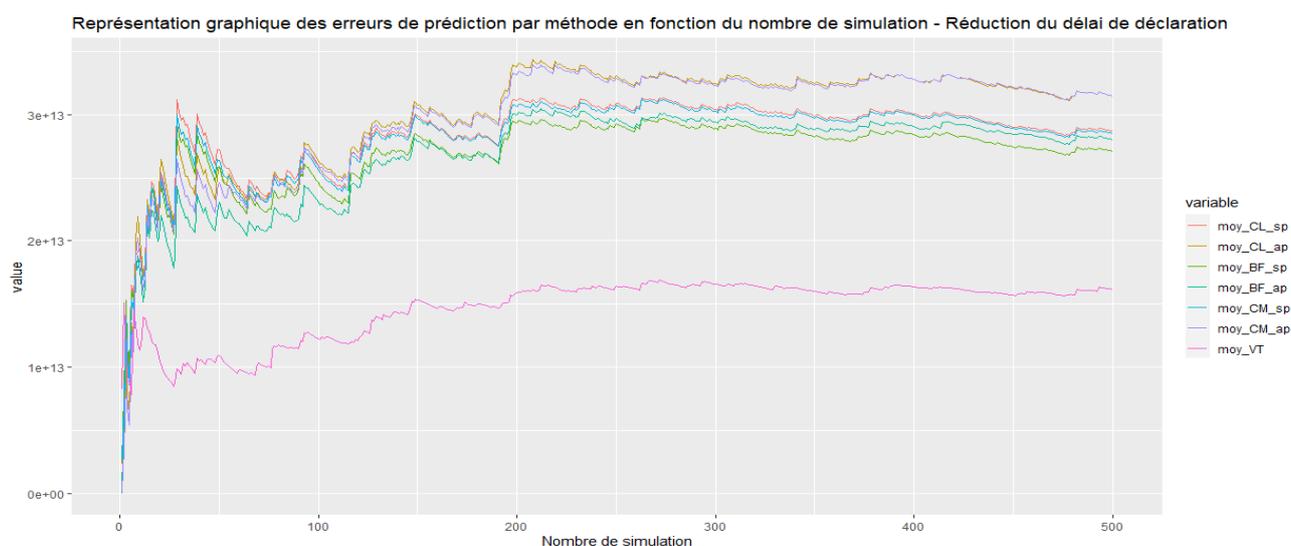


Figure 5.11 : Réduction du délai de déclaration moyen – sensibilité 1

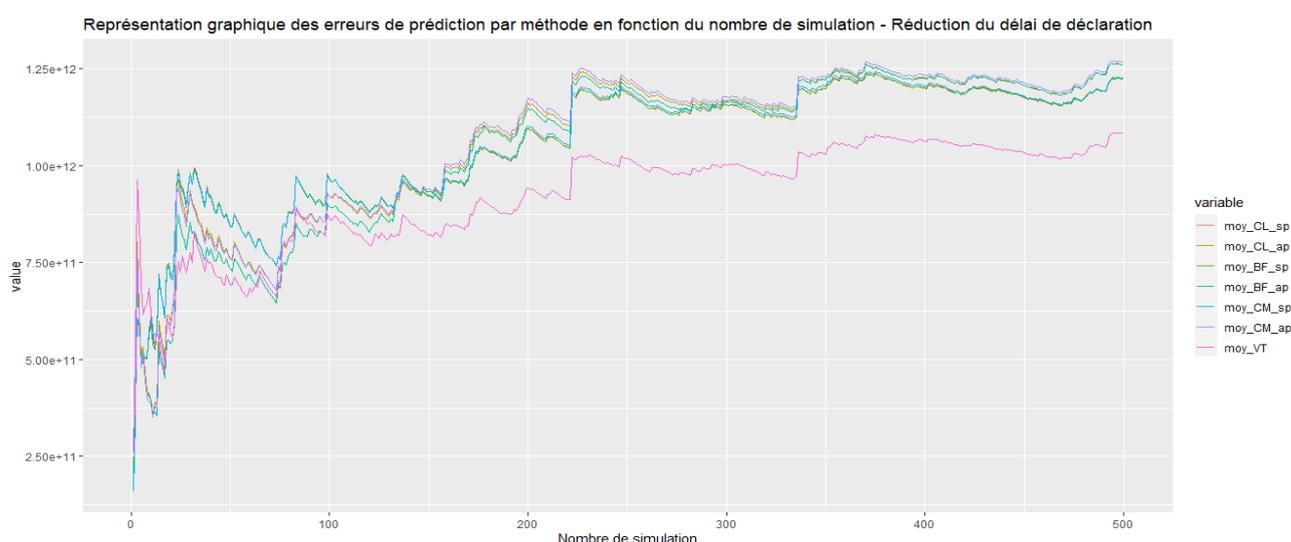


Figure 5.10 : Réduction du délai de déclaration moyen – sensibilité 2

Nous remarquons que plus le délai de déclaration diminue, plus l'erreur de prédiction est faible. Il est en effet cohérent d'avoir une erreur plus faible lorsqu'il y a moins de sinistres tardifs (les IBNR seront équivalent aux IBNER et non pas IBNER + IBNYR). Toutes les méthodes sont impactées et la hiérarchie établie précédemment est conservée.

5.4 Variation de la durée de traitement

Si le délai de déclaration impacte fortement la performance des méthodes, la durée de traitement n'influe presque pas sur les résultats. Nous avons diminué le délai moyen de traitement à 7,5 ans (graphique du haut) puis à 5,5 ans (graphique du bas), mais les erreurs sont similaires (tant sur le niveau et la variabilité que sur la hiérarchie) au cas central.

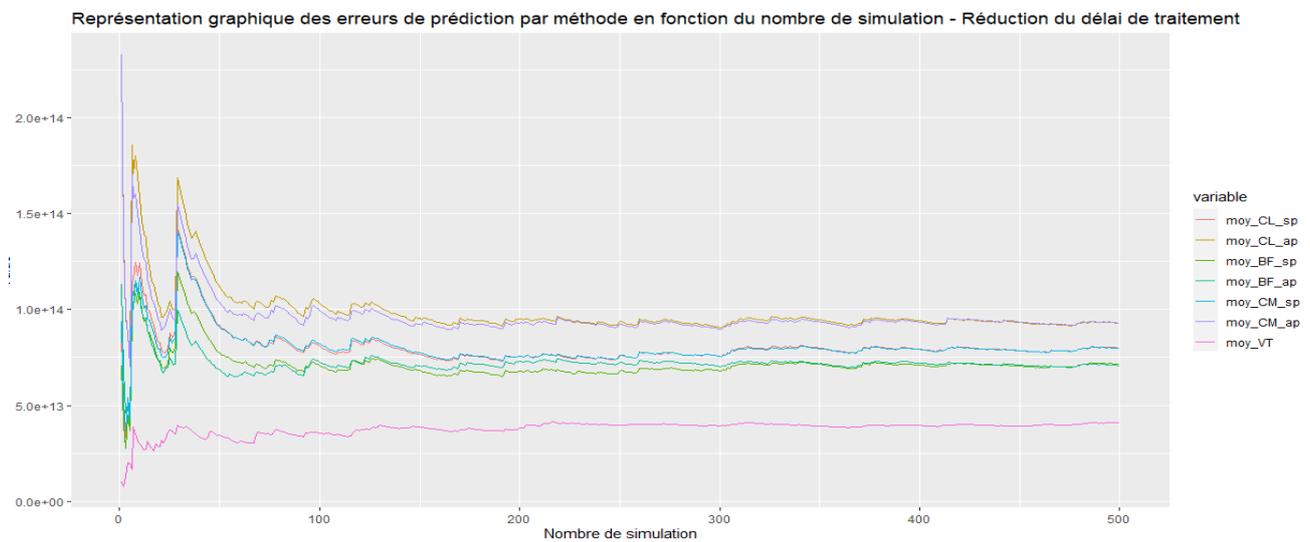


Figure 5.12 : Réduction du délai moyen de traitement - Sensibilité 1

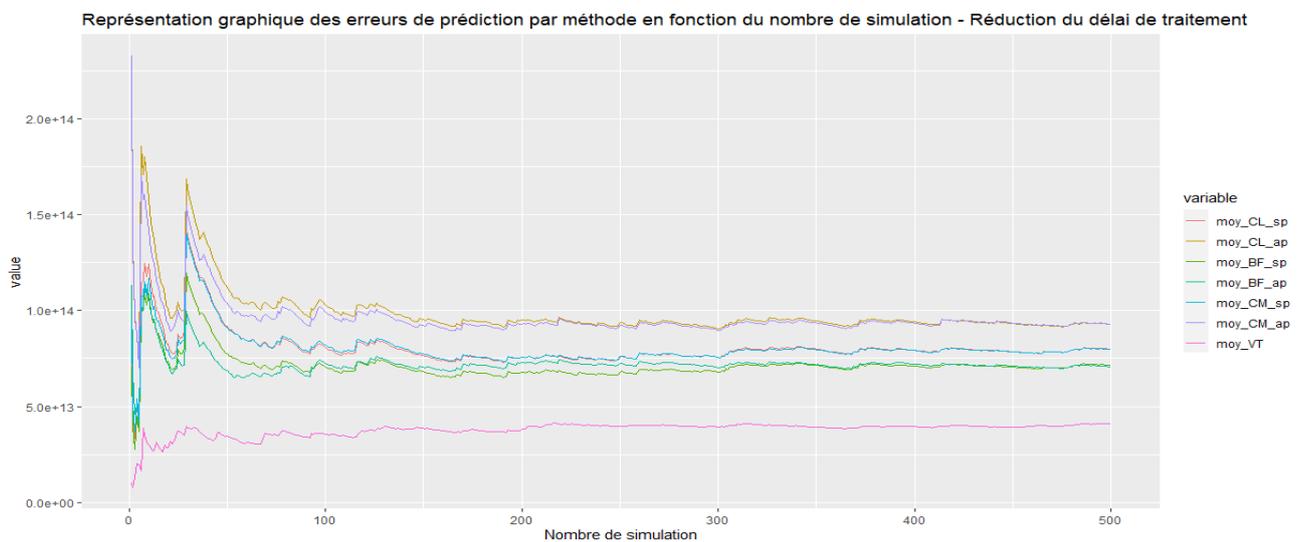


Figure 5.13 : Réduction du délai moyen de traitement : Sensibilité 2

5.5 Variation de l'inflation

L'inflation étant un sujet majeur, nous allons appliquer différents taux d'inflation pour comparer nos méthodes. Dans le cas d'une inflation future qui serait constante à 5%, nous obtenons le graphique suivant :

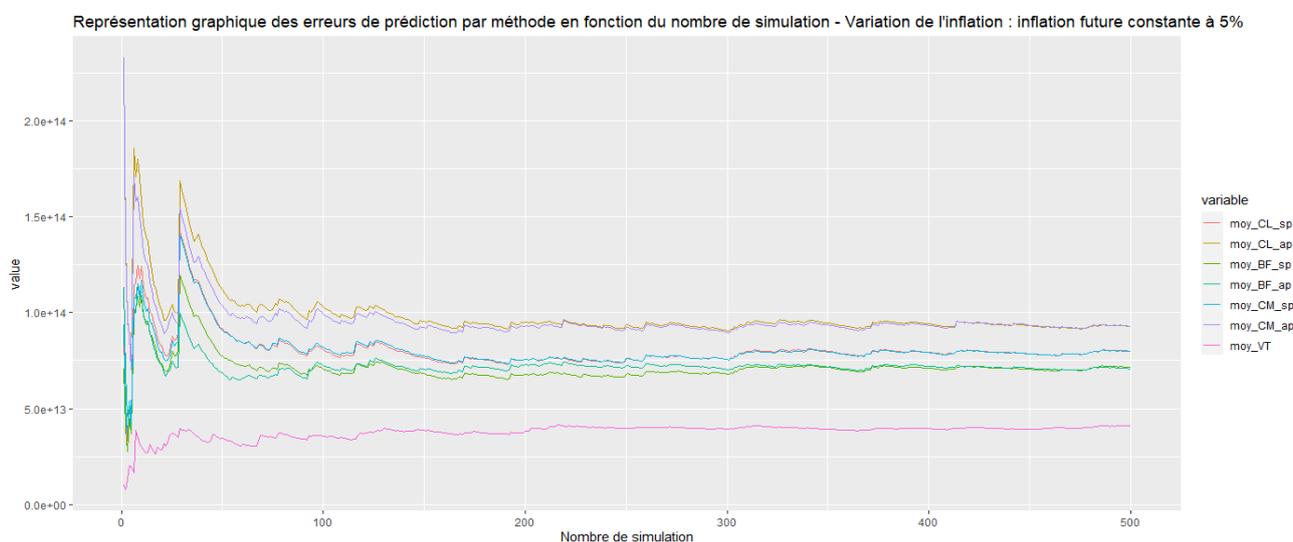


Figure 5.14 : Choc sur l'inflation - Inflation future constante à 5%

La méthode de Verbeek-Taylor est celle qui fonctionne le mieux et la hiérarchie des méthodes ainsi que le niveau d'erreur sont les mêmes que pour le cas central. Ainsi, une inflation constante d'un niveau supérieur n'exercera aucune influence sur la performance des méthodes.

Nous allons maintenant tester le cas où l'inflation future n'est pas constante et suit les projections macroéconomiques de la Banque de France pour les années à venir. Le résultat est très similaire à celui du cas central comme l'illustre le graphique ci-dessous.

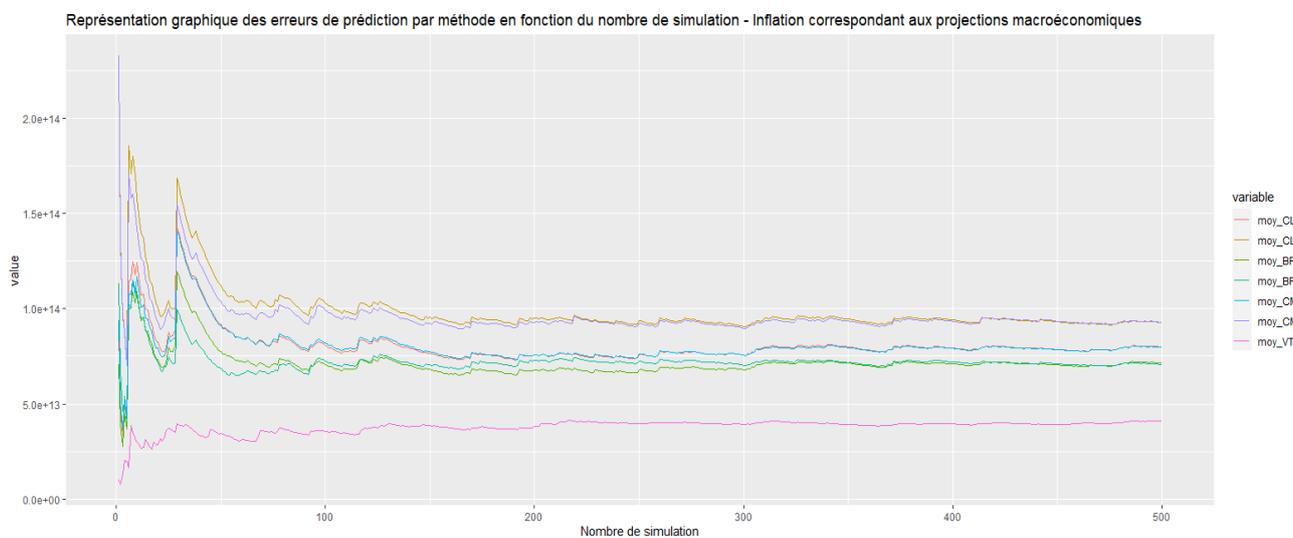


Figure 5.15 : Choc sur l'inflation - Inflation future correspondant aux projections macroéconomiques

Conclusion sur la comparaison des méthodes

Après avoir étudié différentes sensibilités, nous remarquons que les méthodes semblent suivre une certaine hiérarchie.

De manière générale, la méthode Chain Ladder, qui est la méthode de référence, semble toujours avoir une erreur de prédiction plus élevée. Dans certains cas (coûts suivant une loi de Weibull avec un paramètre de forme élevé par exemple), elle est néanmoins très proche des autres méthodes. Etant donné sa simplicité de calcul, elle pourra être préférée à d'autres méthodes dans le cas où les erreurs de prédiction sont proches.

La méthode du coût moyen est légèrement meilleure que Chain Ladder mais la suit de très près. Plus compliquée à mettre en œuvre, elle peut néanmoins être très utilisée si des données sur les nombres de sinistres sont disponibles. En effet, elle peut être plus appropriée que la méthode Chain Ladder notamment dans le cas où les sinistres sont répartis en classe de sévérité.

La méthode de Bornhuetter-Ferguson, qui utilise une cadence différente de Chain Ladder, est meilleure que les deux méthodes précédentes et conforte ce qui était observé sur nos données.

Enfin, la méthode de Verbeek-Taylor, qui nécessite plus d'informations, semble toujours donner les résultats les plus satisfaisants.

Nous n'avons testé que quelques sensibilités et il n'est pas possible d'affirmer avec certitude qu'une méthode aura toujours une erreur de prédiction inférieure à une autre quel que soit le jeu de données, mais nous avons remarqué, sur nos données réelles et simulées, que certaines méthodes semblent donner de meilleurs résultats que d'autres.

6. Application des méthodes stochastiques au calcul de l'ajustement pour risque au titre des sinistres survenus sous IFRS 17

6.1 Méthodes de calcul de l'ajustement pour risque

Pour calculer l'ajustement pour risque (RA), aucune méthode standard n'est prescrite par la norme IFRS 17 et chaque entité est donc libre de la méthode à utiliser du moment que le calcul respecte les principes généraux ci-dessous (paragraphe B91) :

- **Cohérence avec le niveau de risque** : si la distribution des réserves de la branche 1 a une queue épaisse (niveau de risque élevé) et si la distribution des réserves de la branche 2 a une queue fine (niveau de risque plus faible) alors $RA_1 > RA_2$;
- **Cohérence avec la durée des engagements** : l'ajustement pour risque augmente avec la durée de projection des cash flows ;
- **Cohérence avec les données** : s'il y a beaucoup d'incertitude, l'ajustement pour risque doit être plus élevé que si l'incertitude est faible ;
- **Cohérence avec la sévérité** : si les risques sont peu fréquents mais graves, l'ajustement pour risque sera plus élevé que si les risques sont fréquents mais peu graves.

En plus des principes ci-dessus, il doit être adaptatif en fonction des nouvelles anticipations.

Un principe général gouvernant le choix de la méthode est que celui-ci relève de l'exercice du jugement de l'entité et doit contribuer à produire une information financière pertinente permettant aux utilisateurs de l'information financière de comparer les performances de l'entité avec celles de ses pairs (paragraphe B92).

Les deux méthodes les plus utilisées sont l'approche par coût du capital dérivée du calcul de Risk Margin sous Solvabilité 2 et l'approche par quantile. En ce qui concerne le risque non financier associé à la liquidation des sinistres survenus (RA LIC) en assurance non-vie, on constate qu'une majorité d'acteurs ont adopté la seconde approche, celle-ci étant déjà fréquemment mise en œuvre pour fixer un niveau de prudence explicite dans le provisionnement en comptes statutaires.

Nous allons donc nous intéresser à l'approche par quantile où l'ajustement pour risque se calcule comme la différence entre une mesure de risque d'un niveau de confiance α et la moyenne (Best Estimate). Nous introduisons deux mesures de risque :

- **La Value-at-Risk (VaR)** définie comme $\mathbb{P}(X \geq VaR_\alpha(X)) = \alpha$ qui représente le montant associé au risque qui ne sera dépassé qu'avec une probabilité α . L'avantage cette mesure de risque est sa facilité d'interprétation. Néanmoins, elle n'est pas cohérente au sens d'Artzner et al. (1999) puisque la propriété de sous-additivité n'est pas satisfaite. En effet, pour tout risque A et B, la VaR ne vérifie pas l'inégalité suivante : $VaR(A + B) \leq VaR(A) + VaR(B)$

- **La Tail Value-at-Risk (TVaR)** définie comme la moyenne des VaR dépassant $VaR_\alpha(X)$.

$$TVaR = \frac{1}{1 - \alpha} \int_\alpha^1 VaR_\omega(X) d\omega$$

Bien que plus compliquée à calculer, la TVaR a l'avantage d'être cohérente au sens d'Artzner et al. (1999).

Quelque soit la méthode retenue, la norme IFRS 17 précise que (paragraphe 119) :

- La méthode utilisée doit être décrite en annexe ;
- Le niveau de quantile équivalent doit être publié en annexe des états financiers.

6.2 Calcul de l'ajustement pour risque à l'aide de l'approche par quantiles

6.2.1 Définition des paramètres

Les modèles stochastiques que nous avons utilisés donnent une estimation de l'espérance et de la variance mais ne fournissent pas la distribution complète des provisions (sauf pour le cas de la méthode par séparation qui utilise le Bootstrap). Cependant, pour utiliser les mesures de risque définies précédemment, nous avons besoin de formuler une hypothèse sur cette distribution. Nous avons choisi de regarder les lois lognormale, normale et gamma, qui sont les distributions usuellement retenues en pratique par les assureurs non-vie

A partir des espérances et variances calculés dans les chapitres précédents, nous allons calculer les paramètres de ces lois par la méthode des moments. Notons que l'espérance est définie comme la somme des IBNR calculés et des PSAP dossier/dossier.

- **Loi lognormale**

Pour estimer les deux paramètres (μ, σ^2) de la loi lognormale, nous allons utiliser les deux premiers moments : l'espérance et la variance.

$$E(X) = e^{\mu + \frac{\sigma^2}{2}} \quad V(X) = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$$

Par la méthode des moments, nous obtenons :

$$E(X) = e^{\mu + \frac{\sigma^2}{2}} \Leftrightarrow \ln(E(X)) = \mu + \frac{\sigma^2}{2} \Leftrightarrow \mu = \ln(E(X)) - \frac{\sigma^2}{2}$$

En remplaçant la valeur de μ dans la variance, nous obtenons la valeur de σ^2 et en déduisons celle de μ :

$$\begin{aligned}
V(X) &= e^{2\mu+\sigma^2}(e^{\sigma^2} - 1) \\
\Leftrightarrow V(X) &= e^{2\ln(E(X))}(e^{\sigma^2} - 1) \\
\Leftrightarrow \frac{V(X)}{e^{2\ln(E(X))}} &= (e^{\sigma^2} - 1) \\
\Leftrightarrow \frac{V(X)}{e^{2\ln(E(X))}} + 1 &= e^{\sigma^2} \\
\Leftrightarrow \sigma^2 &= \ln\left(\frac{V(X)}{e^{2\ln(E(X))}} + 1\right)
\end{aligned}$$

Ainsi :

$$\sigma^2 = \ln\left(\frac{V(X)}{E(X)^2} + 1\right) \quad \text{et} \quad \mu = \ln(E(X)) - \frac{1}{2} \ln\left(\frac{V(X)}{E(X)^2} + 1\right)$$

- **Loi Normale**

Dans le cas de la loi normale, aucun calcul n'est nécessaire puisque les deux premiers moments sont :

$$E(X) = \mu \quad V(X) = \sigma^2$$

- **Loi gamma**

Pour une loi gamma de paramètre (α, β) les deux premiers moments sont :

$$E(X) = \alpha\beta \quad V(X) = \alpha\beta^2$$

Par la méthode des moments, nous obtenons :

$$\begin{aligned}
\alpha &= \frac{E(X)}{\beta} \\
\beta^2 &= \frac{V(X)}{\alpha} = \frac{V(X)\beta}{E(X)}
\end{aligned}$$

Ainsi :

$$\beta = \frac{V(X)}{E(X)} \quad \text{et} \quad \alpha = \frac{E(X)^2}{V(X)}$$

6.2.2 Application aux données

Pour chaque jeu de données, nous avons sélectionné la méthode qui permet de minimiser l'erreur de prédiction. Voici les estimations retenues :

	RC Auto France « Attritional »	RC Auto France « Large »	RC Auto France « Exceptionnelle »	Acceptations RC Auto UK
Espérance	112	2 309	3 861	11 011
Variance	143	32 027	160 546	704 391

Pour chaque base de données, nous allons présenter les distributions estimées des réserves. Comme mesure de risque, nous allons retenir dans la suite uniquement la VaR pour son explicabilité et son lien immédiat avec le niveau de quantile représentant l'aversion au risque de l'entité.

- RC Auto France : classe « Attritional »

La variance des données est faible donc nous nous attendons à ce que la courbe soit resserrée autour de la moyenne. Voici la représentation des différentes distributions possibles :

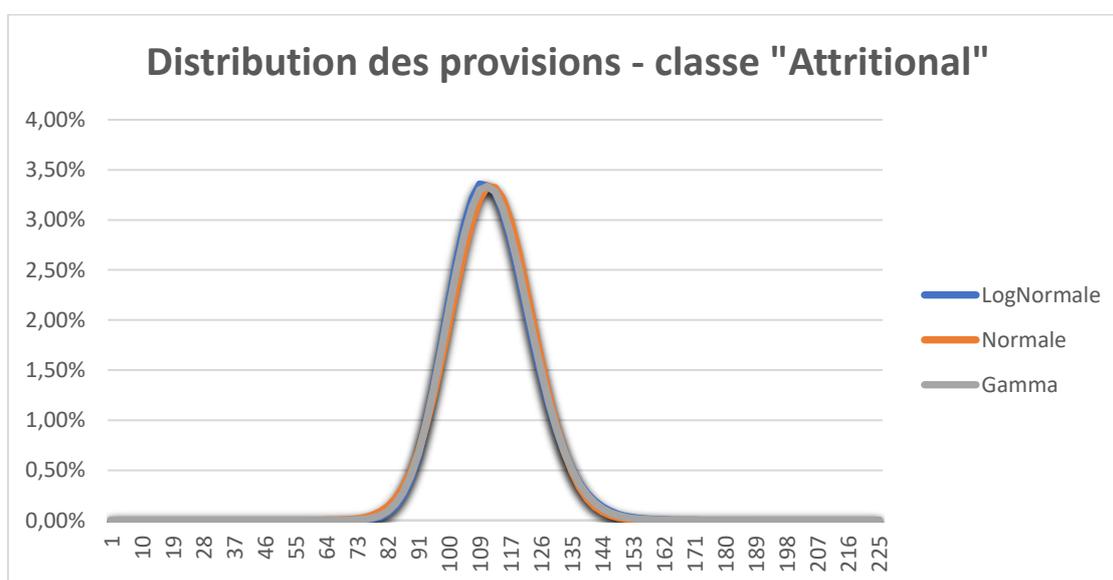


Figure 6.1 : Distribution des provisions - Classe « Attritional »

Les trois courbes se superposent presque et cela se retrouve dans les calculs de la VaR :

Mesure de risque	Niveau de quantile	Loi Lognormale	Loi Normale	Loi Gamma
VaR	70%	117,62	118,13	117,81
	80%	121,67	121,93	121,78

Pour un niveau de quantile à 80%, l'ajustement pour risque vaut en fonction des lois :

Mesure de risque	Niveau de quantile	Loi Lognormale	Loi Normale	Loi Gamma
VaR	80%	121,67 - 111,86 = 9,81	121,93 - 111,86 = 10,06	121,78 - 111,86 = 9,92

Nous n'observons pas de différence significative entre les différentes lois. Nous allons choisir la loi Lognormale et ainsi, la valeur de l'ajustement pour risque vaut 9,81 pour la classe « Attritional ».

En pratique, les assureurs non-vie évaluent d'un côté un « Best Estimate » selon la méthode de leur choix (selon une combinaison de méthodes selon les exercices de survenance ou une segmentation par tranche de sévérité par exemple) et d'un autre côté une volatilité (RMSEP) selon la méthode de Mack (1993 ou 1999). Ils rapprochent ensuite les deux composantes avec le *scaling* en supposant que la volatilité de la méthode utilisée est celle de Chain Ladder.

Ainsi :

$$RMSEP_{\text{méthode choisie}} = \frac{RMSEP_{\text{Chain Ladder}} \times BE_{\text{méthode choisie}}}{BE_{\text{Chain Ladder}}}$$

A titre comparatif, si nous avons appliqué la méthode du *scaling* pour déterminer l'erreur de prédiction associée aux résultats de la méthode Bornhuetter-Ferguson, nous aurions obtenu :

$$RMSEP_{BFAtt} = \frac{13,45 \times 111,86}{89,42} = 16,83$$

La valeur du quantile à 80% aurait été de 125. L'ajustement pour risque aurait donc été égal à : 125,46 – 111,86 = 13,60.

L'ajustement pour risque est légèrement plus élevé mais dans ce cas, l'impact n'est pas significatif car les montants sont très faibles.

- RC Auto France : classe « Large »

Pour la classe « Large », les distributions des provisions sont représentées sur le graphique ci-dessous.

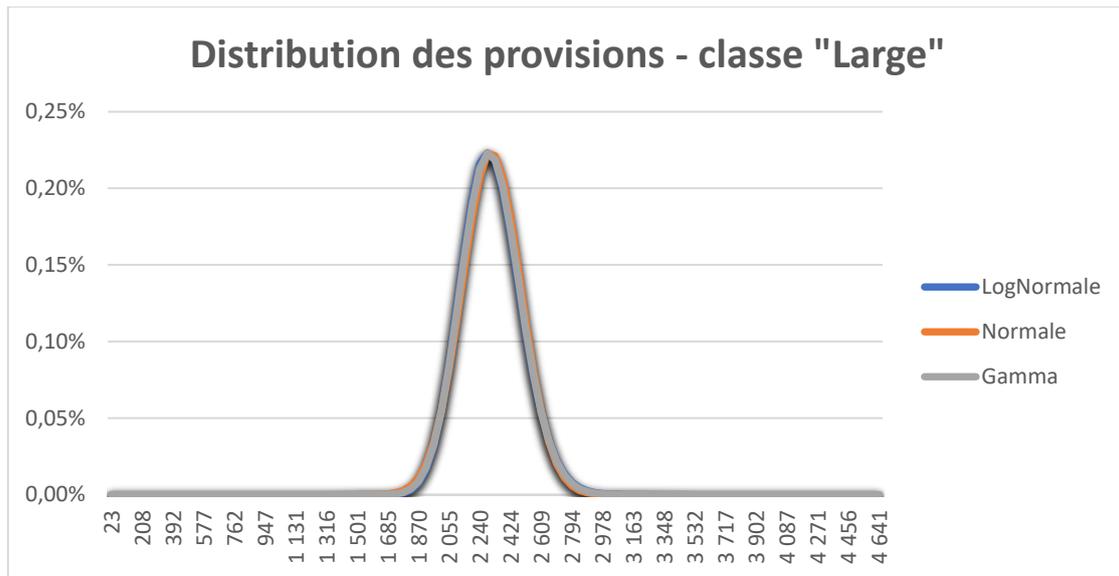


Figure 6.2 : Distribution des provisions – Classe "Large"

Encore une fois, les trois courbes se superposent presque et cela se retrouve dans les calculs de la VaR puisque les différences entre les lois sont minimales :

Mesure de risque	Niveau de quantile	Loi Lognormale	Loi Normale	Loi Gamma
VaR	70%	2 397,21	2 402,62	2 399,16
	80%	2 456,80	2 459,39	2 457,88

Pour un niveau de quantile à 80%, l'ajustement pour risque vaut en fonction des lois :

Mesure de risque	Niveau de quantile	Loi Lognormale	Loi Normale	Loi Gamma
VaR	80%	148,03	150,62	149,11

En choisissant la loi Lognormale, nous obtenons une valeur de l'ajustement pour risque égale à 148,03.

En appliquant la méthode du *scaling*, nous aurions obtenu une valeur de la RMSEP égale à :

$$RMSEP_{BF_{Large}} = \frac{2\,309 \times 221}{2\,262} = 225$$

L'ajustement pour risque aurait donc été évalué à 185,32 avec une loi Lognormale ce qui est largement plus élevé que ce que nous obtenons.

Ceci s'explique par le fait qu'avec la méthode du *scaling*, la RMSEP est évaluée à 225 € alors qu'avec la méthode théorique, elle est seulement de 179 €.

- RC Auto France : classe « Exceptionnelle »

Les distributions obtenues à partir des données de la classe « Exceptionnelle » sont représentées ci-dessous :

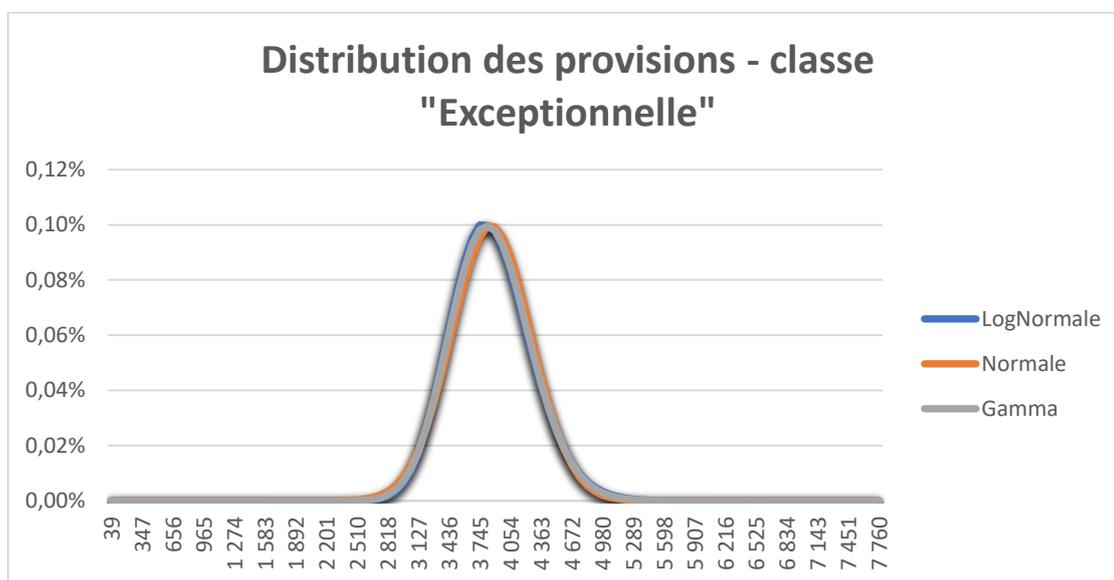


Figure 6.3 : Distribution des provisions – Classe "Exceptionnelle"

Les trois courbes sont ici aussi très proches et nous en déduisons que le choix de la loi a un faible impact sur le calcul de la VaR. En effet, voici les résultats obtenus :

Mesure de risque	Niveau de quantile	Loi Lognormale	Loi Normale	Loi Gamma
VaR	70%	4 054,41	4 070,95	4 060,49
	80%	4 189,74	4 198,06	4 193,39

Pour un niveau de quantile à 80%, l'ajustement pour risque vaut en fonction des lois :

Mesure de risque	Niveau de quantile	Loi Lognormale	Loi Normale	Loi Gamma
VaR	80%	328,90	337,22	332,55

La valeur de l'ajustement pour risque est de 332,55 en utilisant la loi Gamma. En utilisant le *scaling*, la RMSEP aurait été évaluée à :

$$RMSEP_{CM_{Exceptionnelle}} = \frac{3\,861 \times 501}{3\,766} = 513$$

Avec la méthode théorique, la RMSEP était de 401 €. Ceci explique que l'ajustement pour risque calculé à partir de la méthode du *scaling* soit largement supérieur (417,30 vs 332,55).

- Acceptations RC Auto UK

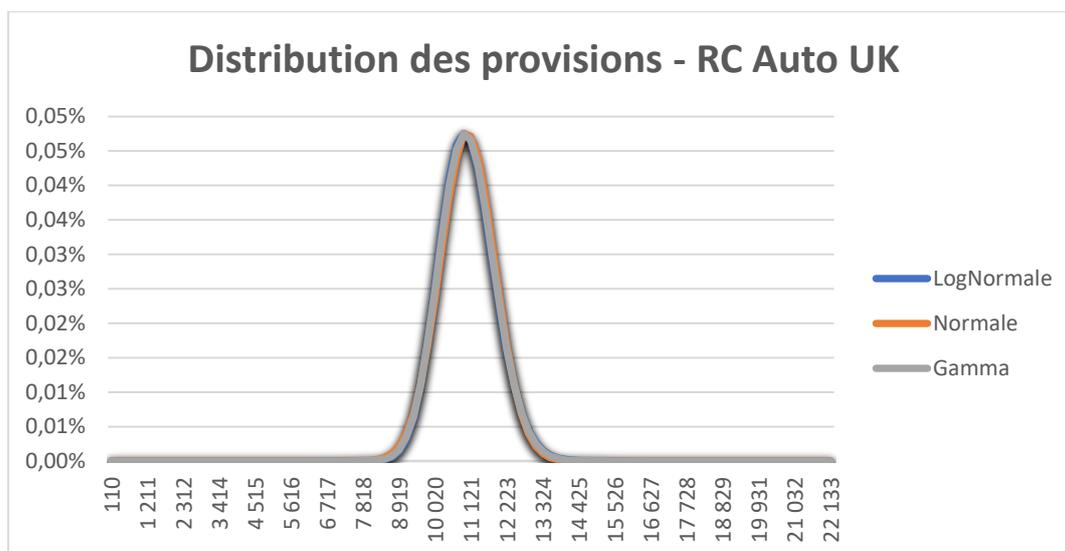


Figure 6.4 : Distribution des provisions pour la base de données RC Auto UK

Comme précédemment, les lois sont proches et le calcul de la VaR donne :

Mesure de risque	Niveau de quantile	Loi Lognormale	Loi Normale	Loi Gamma
VaR	70%	11 426,57	11 451,46	11 435,53
	80%	11 705,80	11 717,70	11 710,77

Pour un niveau de quantile à 80%, il est possible de calculer l'ajustement pour risque :

Mesure de risque	Niveau de quantile	Loi Lognormale	Loi Normale	Loi Gamma
VaR	80%	694,46	706,36	699,42

Nous retenons un ajustement pour risque à 699,42.

En utilisant le *scaling*, il aurait été à 855,94 car au lieu d'avoir une RMSEP à 839 €, elle aurait été évaluée à :

$$RMSEP_{BFUK} = \frac{11\,011 \times 1\,034}{11\,063} = 1\,030$$

L'ajustement pour risque aurait donc été grandement surestimé en utilisant la méthode du *scaling*.

Conclusion sur l'ajustement pour risque

En pratique, la mesure de risque la plus utilisée est la VaR car elle permet de facilement comparer le niveau d'appétit au risque des assureurs.

Au niveau du choix du quantile, celui-ci reflète le niveau d'aversion au risque des assureurs. Plus le niveau de quantile est élevé et plus les assureurs sont averses au risque.

Concernant la loi choisie, nous remarquons qu'il n'y a pas de différence significative pour nos données réelles.

Nous remarquons que l'ajustement pour risque est plus important pour la classe « Exceptionnelle » que pour la classe « Large » ou « Attritional ». Cela est en cohérence avec les principes généraux évoqués puisque l'ajustement pour risque est bien plus élevé pour les risques peu fréquents mais graves qui ont une incertitude forte et une durée élevée (classe « Exceptionnelle ») que pour les risques fréquents mais peu graves avec une incertitude faible et une durée plus courte (classe « Attritional »).

Concernant la méthode du *scaling*, si celle-ci est couramment utilisée par les assureurs non-vie, elle est très approximative (hypothèse forte que toutes les méthodes déterministes sont équivalentes à la méthode Chain Ladder en termes de volatilité) et les écarts observés sont assez significatifs. Cette méthode n'est donc pas à privilégier.

Conclusion

La norme IFRS 17 qui entrera en vigueur au 1^{er} janvier 2023 va avoir un impact sur les procédures de calculs des provisions des assureurs. Pour les entités d'assurance qui sont éligibles au modèle simplifié, la nouveauté réside dans l'évaluation des passifs au titre des sinistres survenus. En effet, en plus de l'évaluation du « Best Estimate », il est nécessaire de calculer un indicateur supplémentaire : l'ajustement pour risque.

S'il est possible de s'appuyer sur le large panel existant de méthodes déterministes pour le calcul du « Best Estimate », il est néanmoins important de mettre en œuvre de nouveaux processus permettant d'évaluer rigoureusement l'incertitude des provisions aux fins du calcul de l'ajustement pour risque.

Ce mémoire a permis de présenter des extensions stochastiques aux méthodes les plus couramment utilisées (Chain Ladder, Bornhuetter-Ferguson, Coût Moyen, Verbeek-Taylor) et ainsi d'estimer les erreurs de prédiction produites par ces méthodes.

Dans un premier temps, nous avons évalué le montant « Best Estimate » des provisions pour des jeux de données réelles. Dans le cas où toutes les hypothèses des modèles n'étaient pas vérifiées, l'application de pondérations a permis de se ramener dans la situation où les méthodes étaient applicables.

Une fois les méthodes déterministes calibrées, nous avons utilisé les extensions stochastiques pour calculer les erreurs de prédiction. Les résultats obtenus sont assez différents des résultats qui découlent de la méthode du *scaling* couramment appliquée dans l'industrie et montrent que cette approche approximative peut conduire à une estimation erronée (surestimation ou sous-estimation) de l'erreur de prédiction.

A l'aide de données simulées, nous avons également pu évaluer les méthodes et identifier celles qui donnent les meilleurs résultats au sens de la minimisation de l'erreur de prédiction en fonction des caractéristiques des données. Nous avons remarqué que certaines méthodes (Bornhuetter-Ferguson et Verbeek-Taylor notamment) génèrent des erreurs de prédiction plus faibles dans une grande majorité des cas.

Enfin, nous avons calculé l'ajustement pour risque sur nos jeux de données réelles en estimant la distribution des provisions. Nous avons regardé les lois lognormale, normale et gamma obtenues à partir des paramètres d'espérance (« Best Estimate ») et de variance (MSEP). Si les résultats obtenus sont peu sensibles au choix de la distribution, le calibrage par l'actuaire des moments d'ordres 1 et 2 revêt une importance cruciale, mettant ainsi l'accent sur le choix de méthodes appropriées.

Nous avons majoritairement utilisé des formules fermées pour le calcul des erreurs de prédiction et nous avons donc été contraints de formuler une hypothèse sur la distribution. Il pourrait être intéressant d'utiliser des méthodes de rééchantillonnage (comme le Bootstrap par exemple) qui permettent de s'exonérer de ce type d'hypothèse puisque la distribution est directement évaluée par la méthode. Néanmoins, la correcte prise en compte des pondérations (exclusions de coefficients de développement, sélection du nombre de diagonales) est actuellement une limite des approches par rééchantillonnage et pourrait constituer une piste pour des travaux ultérieurs. Enfin, dans d'autres contextes que la norme IFRS 17, des travaux analogues à ceux présentés dans ce mémoire pourraient être menés dans une vision à 1 an du risque de provisionnement.

Bibliographie

ARTZNER, P, DELBAEN, F, EBER, J-M, HEATH D (1999), "Coherent measures of risk". *Mathematical Finance*, Vol. 9, No. 3, pages 203–228.

AVANZI, B, TAYLOR G, WANG, M, WONG, B (2021), "*SynthETIC: An individual insurance claim simulator with feature control*". *Insurance: Mathematics and Economics* 100, pages 296-308.

BJÖRKWALL, S, HÖSSJER, O, OHLSSON, E (2010), "*Bootstrapping the Separation Method in Claims Reserving*". *ASTIN Bulletin*, 40, pages 845-869.

DE VYLDER, F (1978), "*Estimation of IBNR claims by least squares*". *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, volume 78, pages 249-254.

DOSSOU, C (2022), "*Prise en compte de l'inflation dans le processus d'inventaire en assurance non-vie*", Mémoire EURIA.

ENGLAND, P.D. , VERRALL, R.J. (2002), "*Stochastic claims reserving in general insurance*". *British Actuarial Journal*, volume 8, pages 443-518.

International Accounting Standards Board (2017), "*IFRS 17 Insurance contracts*".

International Accounting Standards Board (2020), "*Amendments to IFRS 17 Insurance contracts*".

MACK, T (1993), "*Distribution-free calculation of the standard error of the Chain Ladder reserve estimates*". *ASTIN BULLETIN*, volume 23, pages 213-225.

MACK, T (1994), "*Measuring the Variability of Chain Ladder Reserve Estimates*". *Casualty Actuarial Society Forum*, volume 1, pages 101-182.

MACK, T (2006), "*Parameter Estimation for Bornhuetter Ferguson*". *Casualty Actuarial Society Forum*, pages 141-157.

MACK, T (2008), "*The Prediction Error of Bornhuetter Ferguson*". *ASTIN Bulletin*, volume 38, pages 87-103.

PARTRAT, C, LECOEUR, E, NESSI, J-M, NISIPASU, E et REIZ, O (2007), "*Provisionnement technique en assurance non-vie : Perspectives actuarielles modernes*". *Economica*.

PwC (2017), "*IFRS 17 marks a new epoch for insurance contract accounting*", *In depth*, No. INT2017-04.

PwC (2021), "*Formation IFRS 17 : principes généraux*", document interne.

TAYLOR, G.C (1977), "*Separation on inflation and other effects from the distribution of non-life insurance claim delays*". ASTIN Bulletin, volume 9, pages 217-230.

VERBEEK, H.G (1972), "*An approach to the analysis of claims experience in motor liability excess of loss reinsurance*". ASTIN Bulletin, volume 6, pages 195-202.

ANNEXES

Annexe 1 : Présentation des données

Les données issues des portefeuilles RC Automobile français et anglais sont présentées en base 100.

- Données RC Automobile France – classe « Attritional »

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25		
1996	100,0	99,6	92,9	89,7	87,0	86,2	84,7	84,2	83,8	83,8	84,0	83,9	83,9	83,7	84,0	84,2	84,1	84,2	84,2	82,5	84,3	84,3	84,4	84,3	84,4	84,3	84,4	
1997	89,0	92,8	86,7	82,8	80,3	78,5	76,7	75,0	74,2	73,7	73,5	72,7	72,3	71,3	70,5	70,2	70,3	70,1	69,2	69,3	69,3	69,4	70,0	70,0	69,5			
1998	95,1	98,0	90,7	87,4	84,9	83,4	83,6	82,2	82,3	81,5	81,5	81,2	80,4	80,7	80,8	81,0	80,9	80,7	80,6	80,6	79,9	79,6	79,8	79,8				
1999	97,7	104,1	97,4	93,0	91,9	88,5	87,7	86,2	86,1	86,0	85,6	85,3	84,7	84,6	84,5	84,5	84,3	84,1	84,3	84,1	84,2	84,2	84,2					
2000	95,8	101,5	95,3	91,6	88,9	87,2	86,2	85,8	84,9	83,9	83,4	83,2	83,6	83,8	84,0	84,3	83,5	83,6	83,7	81,5	81,1	81,3						
2001	100,8	104,5	96,2	92,9	90,6	89,2	88,2	86,5	87,0	86,6	86,4	85,7	85,4	85,5	85,7	86,3	86,2	85,9	86,0	86,1	85,9							
2002	97,0	100,8	95,7	92,2	88,5	88,2	87,3	87,1	87,2	86,9	85,3	84,7	84,5	84,2	86,0	86,1	85,6	85,8	85,6	85,4								
2003	94,4	94,5	87,4	83,3	82,2	80,7	79,5	79,3	79,5	79,1	78,7	78,7	78,7	78,7	78,3	78,2	78,2	78,4	78,1									
2004	79,8	80,8	75,2	71,5	69,4	68,1	67,9	67,5	67,3	67,3	67,0	66,5	66,9	66,9	66,5	66,8	66,7	66,6										
2005	79,4	78,5	73,5	70,1	67,5	66,3	65,6	65,1	64,8	64,4	64,5	64,4	64,6	64,5	64,4	64,3	64,4											
2006	73,8	76,9	71,1	67,3	65,4	64,3	64,6	63,5	63,1	63,1	62,8	62,9	62,6	62,5	62,5	62,5												
2007	66,8	76,1	70,5	67,2	65,7	64,8	64,1	63,4	63,6	63,5	63,3	63,2	63,2	63,0	63,0													
2008	63,7	75,3	70,6	67,7	65,9	64,9	63,9	63,7	63,6	63,5	63,3	63,3	63,1	63,1														
2009	67,7	78,0	72,2	69,5	68,4	66,8	66,6	66,3	66,0	65,7	65,7	65,6	65,6															
2010	66,4	82,2	76,6	74,4	72,1	71,5	71,2	70,6	70,4	70,3	70,1	70,1																
2011	67,5	82,0	77,0	73,2	72,0	71,2	70,4	70,4	69,9	69,8	70,0																	
2012	68,6	82,4	76,7	75,1	72,6	71,4	71,2	70,9	70,6	70,6																		
2013	62,8	75,0	71,1	71,0	69,9	69,1	68,9	68,4	67,4																			
2014	59,1	73,3	69,5	67,2	65,5	64,4	64,3	64,3																				
2015	62,8	72,3	69,6	67,3	66,1	65,9	65,3																					
2016	61,7	75,8	72,9	70,0	68,8	68,6																						
2017	63,9	77,3	74,5	73,1	71,9																							
2018	65,4	76,8	74,7	72,5																								
2019	65,6	80,2	77,5																									
2020	54,0	65,7																										
2021	66,2																											

Triangle cumulé des charges

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25		
1996	35,4	66,6	74,8	78,6	81,1	82,0	82,3	83,0	83,1	83,2	83,4	83,7	83,8	83,9	83,9	83,8	83,9	83,9	83,9	84,0	83,8	84,2	84,2	84,3	84,3	84,3	84,3	
1997	36,0	67,3	74,5	77,6	79,6	79,9	80,4	80,4	79,0	78,1	77,8	77,4	76,3	76,3	75,4	75,3	74,9	74,6	74,1	74,1	73,6	72,5	71,9	70,3	70,1			
1998	37,2	68,6	76,3	80,0	82,4	82,5	83,1	82,8	82,9	82,6	82,8	83,0	82,9	83,0	82,7	82,7	82,8	82,9	82,9	82,9	82,9	82,9	82,8	81,5	81,5			
1999	36,3	67,5	76,2	80,2	82,5	84,1	85,4	84,7	84,9	84,8	84,9	84,5	84,7	84,8	84,8	84,5	84,6	84,4	84,4	84,3	84,1	84,1	84,2					
2000	35,9	69,1	78,2	82,8	84,7	85,6	87,1	87,3	87,0	86,6	86,3	86,0	85,9	85,9	85,8	85,9	85,0	84,9	84,9	84,9	84,7	84,2						
2001	36,4	69,7	79,2	84,0	86,1	87,9	88,1	88,4	88,1	88,2	88,5	88,5	88,3	88,4	88,1	86,3	85,8	86,0	86,1	86,1	86,0							
2002	36,7	68,5	78,4	82,7	85,4	85,1	86,0	86,3	86,3	86,3	86,3	86,4	86,6	86,3	86,3	86,3	86,2	86,0	85,7	85,7								
2003	33,0	61,6	69,0	72,5	73,9	75,7	76,5	77,4	77,8	77,7	77,7	77,9	78,0	78,0	78,0	78,0	78,0	78,0	78,0	78,1								
2004	28,7	52,2	59,1	61,5	63,2	64,4	65,3	65,1	65,7	66,2	66,2	66,3	66,5	66,5	66,0	66,6	66,7	66,7										
2005	27,4	50,5	56,4	59,1	60,6	62,0	63,1	63,1	63,6	63,8	63,8	63,9	64,2	64,2	64,3	64,4	64,4											
2006	26,0	48,8	55,1	58,1	59,7	60,5	61,2	61,4	61,7	62,0	62,2	62,3	62,3	62,3	62,3	62,3												
2007	24,1	48,3	55,0	58,2	59,6	60,6	61,4	61,7	61,9	62,3	62,5	62,7	62,8	62,8	62,9													
2008	23,6	47,9	54,7	57,8	59,6	60,6	61,2	61,7	61,9	62,1	62,1	62,3	62,3	62,4														
2009	24,3	50,7	57,5	60,8	62,2	63,0	63,7	64,2	64,6	64,7	64,9	65,0	65,1															
2010	24,0	52,5	60,4	63,6	66,0	67,0	67,6	68,2	68,7	69,1	69,1	69,3																
2011	23,9	53,0	60,4	64,1	65,7	66,7	67,6	68,1	68,4	68,6	68,9																	
2012	24,4	52,3	59,9	63,3	65,3	66,6	67,5	68,0	68,4	68,7																		
2013	22,7	49,2	57,1	61,2	64,0	65,4	66,0	66,5	66,7																			
2014	22,1	47,3	54,3	57,7	59,2	60,4	60,7	60,8																				
2015	23,6	48,4	55,4	58,4	59,8	61,0	61,5																					
2016	24,6	50,1	57,6	60,4	62,3	63,7																						
2017	24,5	50,6	58,1	61,8	63,9																							
2018	24,2	50,2	57,4	61,1																								
2019	24,1	50,2	58,7																									
2020	18,8	41,7																										
2021	23,3																											

Triangle cumulé des paiements

• Données RC Automobile France – classe « Large »

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1996	100,0	133,5	131,6	135,5	134,6	135,1	133,0	131,2	128,8	128,4	125,6	116,6	116,3	116,7	117,4	117,1	117,0	117,2	116,9	116,9	117,0	118,2	119,0	119,0	119,6	120,0
1997	130,9	155,8	149,6	154,2	143,0	137,6	132,3	127,7	129,3	127,4	125,6	131,4	130,7	130,3	130,6	129,4	129,9	135,0	137,9	139,1	140,5	141,2	142,2	142,1	147,7	
1998	222,2	257,8	254,4	245,1	220,9	209,4	202,1	202,7	205,1	203,6	194,6	197,2	199,2	196,8	199,4	199,5	203,4	202,1	203,0	204,2	200,7	209,7	211,2	214,7		
1999	180,7	222,9	216,5	219,5	203,0	190,9	210,4	209,2	206,5	201,8	199,8	201,7	201,1	200,6	202,3	204,4	205,4	207,3	209,8	209,2	211,7	210,0	213,1			
2000	150,2	206,0	213,9	219,6	222,6	213,0	206,6	197,9	202,2	199,2	197,6	194,0	193,0	192,9	191,6	192,5	192,0	195,6	198,3	203,2	201,2	201,6				
2001	153,4	179,9	185,9	186,0	188,9	178,5	178,3	178,6	171,4	170,0	172,6	177,6	173,2	174,1	179,3	176,3	178,1	183,6	184,3	181,9	179,1					
2002	111,9	139,8	144,8	154,1	156,8	160,1	158,2	156,9	151,2	159,6	159,9	165,0	167,5	169,1	168,5	177,3	182,1	181,7	179,7	177,8						
2003	118,4	166,5	178,8	179,2	186,6	182,0	181,6	164,0	161,8	174,1	174,5	172,0	170,8	176,8	178,5	179,8	182,0	186,9	187,2							
2004	98,1	109,2	107,7	113,6	114,8	114,1	108,3	105,8	105,3	105,2	106,1	103,0	102,5	100,9	99,8	98,3	104,1	107,7								
2005	141,8	196,4	211,5	203,9	191,2	196,0	199,0	201,3	196,5	204,7	196,8	198,4	198,0	199,8	198,9	200,5	196,4									
2006	118,7	166,8	168,8	179,9	179,6	178,3	171,9	169,5	167,6	169,7	166,1	163,2	171,6	167,2	162,9	166,2										
2007	125,9	172,4	183,7	171,2	179,7	177,3	174,6	160,0	146,7	143,4	138,9	137,9	131,7	130,4	128,9											
2008	185,3	221,7	232,4	242,7	259,2	268,2	267,2	248,9	240,2	228,0	226,2	225,7	224,9	230,2												
2009	122,6	186,5	205,0	204,7	206,9	200,6	185,4	173,8	174,0	178,0	176,6	176,0	169,6													
2010	130,2	156,5	180,9	197,5	209,2	213,8	226,5	224,6	228,3	212,8	209,5	218,3														
2011	175,5	213,5	229,8	242,7	239,8	215,3	213,5	217,1	215,8	212,2	212,0															
2012	164,8	256,9	286,6	278,8	279,3	290,6	274,5	265,9	274,2	276,9																
2013	139,7	171,7	202,3	205,3	220,6	213,8	220,4	213,0	222,1																	
2014	186,4	233,4	243,2	240,7	246,0	250,8	242,0	235,8																		
2015	186,4	238,9	254,3	263,0	276,1	272,7	262,6																			
2016	222,2	274,9	280,0	278,7	273,0	255,3																				
2017	210,5	268,3	264,7	279,0	288,6																					
2018	182,9	245,6	269,5	296,0																						
2019	156,8	264,2	306,8																							
2020	210,3	263,6																								
2021	247,5																									

Triangle cumulé des charges

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1996	7,8	20,9	32,5	49,3	53,8	76,0	89,0	94,9	100,7	104,7	106,6	108,8	109,6	109,7	113,0	113,6	113,6	113,6	114,1	114,6	114,6	115,1	115,1	116,3	116,5	118,0
1997	8,1	21,8	34,2	54,1	69,7	85,7	92,8	97,4	104,2	108,5	112,1	119,8	120,2	122,0	122,3	123,9	123,9	125,0	131,3	132,0	134,1	134,4	134,4	136,0	138,1	
1998	7,5	39,5	61,2	86,7	116,2	130,1	142,1	150,0	160,1	169,0	174,0	174,4	181,1	184,4	185,9	189,0	191,4	192,0	192,1	192,0	192,2	193,5	198,1	198,8		
1999	6,3	34,4	61,3	87,3	110,4	119,6	145,0	154,3	164,3	174,3	178,8	184,7	193,1	194,1	196,8	200,0	201,0	204,2	204,9	204,9	207,4	207,7	208,9			
2000	6,0	37,8	54,4	78,3	100,9	120,7	135,3	148,9	162,1	170,7	170,0	176,6	179,1	183,8	188,9	190,0	188,8	189,6	192,9	196,1	200,7	200,8				
2001	11,1	39,9	65,9	85,6	113,4	129,7	136,5	148,6	161,2	151,4	153,3	155,4	156,7	158,7	162,2	162,2	164,3	170,2	172,0	172,1	173,0					
2002	13,1	28,4	42,3	59,5	70,6	93,7	114,0	128,2	134,0	139,5	143,5	146,7	155,7	161,8	163,4	165,5	165,8	169,0	173,6	175,3						
2003	3,2	18,8	42,2	68,4	91,2	97,3	120,2	135,1	141,6	151,7	156,0	157,3	159,1	161,9	166,0	166,9	169,2	173,0	177,1							
2004	4,0	13,3	29,1	38,8	45,2	60,2	73,0	78,6	79,1	82,9	87,6	88,0	88,3	91,9	93,8	93,8	94,0	96,1								
2005	5,6	23,5	49,5	77,1	105,4	115,7	129,0	138,9	146,7	163,4	172,8	175,8	176,6	177,6	179,0	182,6	184,4									
2006	7,7	26,7	39,7	55,7	68,4	78,6	96,9	105,3	124,5	132,2	140,5	143,2	150,1	152,7	152,7	153,5										
2007	6,1	23,7	48,3	58,8	75,8	89,1	99,8	107,3	116,0	122,9	125,3	127,2	128,2	128,2	128,2											
2008	6,9	48,7	66,2	90,2	105,5	131,4	160,8	172,3	184,0	193,5	204,3	209,3	211,2	212,3												
2009	5,5	36,9	52,1	75,6	98,3	122,3	131,9	140,0	144,8	151,9	155,0	157,8	157,9													
2010	5,1	25,6	52,0	67,8	92,7	114,2	133,4	146,4	155,8	170,2	175,3	179,3														
2011	7,6	31,1	55,5	78,0	99,6	112,6	129,9	134,1	146,1	153,5	164,7															
2012	9,2	52,5	79,7	112,1	147,3	172,7	189,9	200,3	208,0	219,9																
2013	5,6	26,0	46,8	72,3	89,9	99,1	124,0	144,8	173,0																	
2014	8,6	28,7	45,5	67,6	90,9	107,5	124,4	133,9																		
2015	2,5	42,4	76,9	102,0	133,5	149,4	165,0																			
2016	5,0	31,5	55,5	86,3	107,0	127,1																				
2017	4,7	33,6	61,7	96,8	122,6																					
2018	2,7	28,6	52,8	77,6																						
2019	2,3	25,3	63,8																							
2020	4,5	23,3																								
2021	0,4																									

Triangle cumulé des paiements

- Données RC Automobile UK

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
2002	100,0	212,7	363,8	530,9	458,8	403,6	433,8	437,7	465,1	465,4	464,4	464,9	464,8	465,0	465,0	471,1	471,7	467,3	468,1	466,9
2003	119,6	464,3	696,2	832,8	783,1	828,3	913,6	917,5	1 068,1	1 159,1	1 105,9	1 062,7	1 052,2	1 047,8	1 201,1	1 170,6	1 135,3	1 140,8	1 124,5	
2004	40,4	436,5	491,9	553,1	659,4	909,7	875,7	1 044,2	1 050,1	988,7	1 046,7	1 078,7	1 076,3	1 223,1	1 166,4	1 119,0	1 099,0	1 089,7		
2005	182,3	595,2	601,1	695,8	499,8	485,7	539,2	565,5	593,6	603,5	594,7	608,7	602,8	621,5	621,1	623,1	650,7			
2006	227,1	409,7	410,9	445,5	457,8	372,0	385,9	413,3	436,1	435,7	431,4	495,3	497,2	450,9	456,6	457,4				
2007	172,1	391,6	468,2	571,6	592,7	697,9	709,8	733,8	631,4	608,6	800,5	1 010,2	915,4	907,9	945,9					
2008	200,1	370,1	466,7	484,5	597,8	683,2	674,5	741,1	775,2	909,8	938,4	918,3	917,6	857,1						
2009	87,5	342,9	548,0	581,0	528,5	557,3	520,6	511,6	782,2	749,4	693,1	676,6	686,2							
2010	184,7	482,3	539,0	547,5	598,2	679,6	644,4	976,9	1 003,0	890,0	881,5	930,9								
2011	353,5	804,2	793,2	841,9	813,8	755,2	1 231,9	1 155,7	990,6	1 030,7	1 036,9									
2012	410,6	792,7	1 023,8	1 027,0	1 027,9	1 720,6	1 664,1	1 618,4	1 535,5	1 453,0										
2013	420,3	851,7	1 041,1	962,1	2 099,5	2 195,0	1 927,8	1 742,7	1 703,0											
2014	408,1	922,8	1 402,7	2 754,0	2 859,4	2 475,6	2 224,5	2 185,5												
2015	349,7	975,2	2 791,6	2 512,9	2 273,6	2 100,7	2 168,7													
2016	203,9	1 837,2	2 130,7	1 835,3	1 748,4	1 746,7														
2017	123,7	749,8	671,1	574,1	556,6															
2018	2,1	12,8	17,1	59,1																
2019	12,4	12,3	28,3																	
2020	7,5	25,1																		
2021	213,5																			

Triangle cumulé des charges

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
2002	-	0,2	6,9	35,7	152,1	186,8	212,9	287,3	387,5	418,7	435,9	438,1	439,9	440,9	441,8	448,8	450,2	450,9	452,0	452,6
2003	-	-	-	80,3	133,0	222,9	275,0	366,9	454,7	473,8	444,6	480,4	490,3	501,3	526,2	550,6	571,8	588,9	597,5	
2004	-	-	10,4	18,9	66,7	160,3	221,1	249,3	270,9	317,9	342,0	355,7	372,1	410,0	439,8	481,8	624,3	640,7		
2005	-	-	30,4	53,0	75,1	152,2	208,3	244,5	269,2	276,0	285,2	291,4	303,6	313,3	321,2	326,3	330,9			
2006	-	-	11,0	31,7	67,1	94,9	118,3	128,0	195,2	213,8	221,8	223,6	234,2	238,3	256,9	262,9				
2007	8,2	10,3	10,8	51,2	112,5	162,5	216,3	276,8	285,7	290,7	308,8	314,3	319,5	322,5	323,9					
2008	-	-	0,6	36,3	66,4	139,1	208,9	230,6	253,5	316,8	344,1	394,7	407,8	440,7						
2009	-	-	9,6	16,6	70,8	180,6	221,0	298,7	302,0	336,8	352,7	388,4	395,5							
2010	-	0,0	28,7	40,8	65,6	133,8	201,1	247,3	254,6	267,2	316,2	318,8								
2011	-	-	11,9	82,2	171,2	260,5	320,0	354,5	381,7	397,3	444,1									
2012	-	19,1	28,9	76,3	128,4	202,3	364,1	470,4	548,7	565,7										
2013	-	-	21,5	113,7	167,3	604,6	694,5	899,2	1 078,2											
2014	-	-	2,2	53,3	326,1	503,9	900,0	1 001,6												
2015	-	-	13,3	128,7	245,7	463,5	641,3													
2016	-	11,5	13,8	103,2	145,2	167,0														
2017	-	-	-	7,6	45,1															
2018	-	-	4,3	4,3																
2019	-	-	-																	
2020	-	-	-																	
2021	-	-	-																	

Triangle cumulé des paiements

Annexe 2 : Démonstration de la décomposition de la variance pour la méthode du coût moyen stochastique

$$\begin{aligned}
 \text{Var}(C|D) &= \text{Var}(N.\bar{X}|D) \\
 &= \text{Var}[E(N.\bar{X}|N, D)|D] + E[\text{Var}(N.\bar{X}|N, D)|D] \\
 &= \text{Var}[N.E(\bar{X}|D)|D] + E[N^2\text{Var}(\bar{X}|N, D)|D] \\
 &= \text{Var}[N.E(\bar{X}|D)|D] + E\left[N^2 \times \frac{1}{N}\text{Var}(X|D)\right|D] \\
 &= E(X|D)^2\text{Var}[N|D] + E[N|D]\text{Var}(X|D)
 \end{aligned}$$

Or :

$$\begin{aligned}
 \text{Var}(\bar{X}|D) &= \text{Var}[E(\bar{X}|N, D)|D] + E[\text{Var}(\bar{X}|N, D)|D] \\
 &= \text{Var}[E(X|D)|D] + E\left[\frac{1}{N}\text{Var}(X|D)\right|D] \\
 &= \text{Var}(X|D)E\left(\frac{1}{N}\right|D)
 \end{aligned}$$

D'où :

$$\begin{aligned}
 \text{Var}(C|D) &= E(X|D)^2\text{Var}[N|D] + E[N|D]\text{Var}(X|D) \\
 &= E(X|D)^2\text{Var}[N|D] + E[N|D]\frac{\text{Var}(\bar{X}|D)}{E\left(\frac{1}{N}\right|D)}
 \end{aligned}$$

□

Annexe 3 : Calcul de l'estimateur de F^2

$$F^2 = \left(\sum_{k=n+1-i}^{n-1} S_k \right)^2 = \sum_{k=n+1-i}^{n-1} S_k^2 + 2 \sum_{j < k} S_j S_k \approx \sum_{k=n+1-i}^{n-1} E(S_k^2|B_k) + 2 \sum_{j < k} E(S_j S_k|B_k)$$

Notations :

- $\hat{f}_k^{(N)} = \frac{\sum_{i=1}^{n-k} f_{i,k}^{(N)} N_{i,k}^\alpha w_{i,k}^N}{\sum_{i=1}^{n-k} N_{i,k}^\alpha w_{i,k}^N}$ avec $f_{i,k}^{(N)} = \frac{N_{i,k+1}}{N_{i,k}}$ et $\hat{f}_k^{(X)} = \frac{\sum_{i=1}^{n-k} f_{i,k}^{(X)} \bar{X}_{i,k}^\alpha w_{i,k}^{(X)}}{\sum_{i=1}^{n-k} \bar{X}_{i,k}^\alpha w_{i,k}^{(X)}}$ avec $f_{i,k}^{(X)} = \frac{\bar{X}_{i,k+1}}{\bar{X}_{i,k}}$
- $f_k^{(X)} f_k^{(N)} = f_k$ et $\hat{f}_k^{(X)} \hat{f}_k^{(N)} = \hat{f}_k$
- $Y_i = f_{i,k}^{(N)} N_{i,k}^\alpha w_{i,k}^N$ et $Z_j = f_{j,k}^{(X)} \bar{X}_{j,k}^\alpha w_{j,k}^{(X)}$

Calcul de $E(S_j S_k | B_k)$

Pour $j < k$:

$$E(S_j S_k | B_k) = S_j E(S_k | B_k) = S_j \hat{f}_{n+1-i} \dots \hat{f}_{k-1} E(f_k - \hat{f}_k | B_k) f_{k+1} \dots f_{n-1}$$

Or :

$$E(\hat{f}_k | B_k) = f_k$$

D'où :

$$E(S_j S_k | B_k) = 0$$

Calcul de $E(S_k^2 | B_k)$

$$E(S_k^2 | B_k) = \hat{f}_{n+1-i}^2 \dots \hat{f}_{k-1}^2 E((f_k - \hat{f}_k)^2 | B_k) f_{k+1}^2 \dots f_{n-1}^2$$

Comme $E(\hat{f}_k | B_k) = f_k$:

$$E((f_k - \hat{f}_k)^2 | B_k) = \text{Var}(\hat{f}_k | B_k) = \frac{\text{Var}(\sum_{i=1}^{n-k} \sum_{j=1}^{n-k} Y_i Z_j | B_k)}{(\sum_{i=1}^{n-k} N_{i,k}^\alpha W_{i,k}^N)^2 (\sum_{i=1}^{n-k} \bar{X}_{i,k}^\alpha W_{i,k}^{(X)})^2}$$

Or :

$$\begin{aligned} \text{Var}\left(\sum_{i=1}^{n-k} \sum_{j=1}^{n-k} Y_i Z_j | B_k\right) &= \sum_i \sum_j \text{Var}(Y_i Z_j | B_k) + \sum_i \sum_j \sum_{s \neq j} \text{Cov}(Y_i Z_j, Y_i Z_s | B_k) \\ &\quad + \sum_i \sum_j \sum_{r \neq i} \text{Cov}(Y_i Z_j, Y_r Z_j | B_k) + \sum_i \sum_j \sum_{r \neq i} \sum_{s \neq j} \text{Cov}(Y_i Z_j, Y_r Z_s | B_k) \\ &= \sum_i \text{Var}(Y_i Z_i | B_k) + \sum_i \sum_{j \neq i} \text{Var}(Y_i | B_k) \text{Var}(Z_j | B_k) + \sum_i \sum_{s \neq i} \text{Cov}(Y_i Z_i, Y_i Z_s | B_k) \\ &\quad + \sum_i \sum_{j \neq i} \sum_{s \neq j} \text{Cov}(Y_i Z_j, Y_i Z_s | B_k) + \sum_j \sum_{i \neq j} \sum_{r \neq i} \text{Cov}(Y_i Z_j, Y_r Z_j | B_k) \\ &\quad + \sum_j \sum_{r \neq j} \text{Cov}(Y_j Z_j, Y_r Z_j | B_k) + \sum_i \sum_j \sum_{r \neq i} \sum_{s \neq j} \text{Cov}(Y_i Z_j, Y_r Z_s | B_k) \end{aligned}$$

$$\begin{aligned}
&= \sum_i \text{Var}(Y_i Z_i | B_k) + \sum_i \sum_{j \neq i} \text{Var}(Y_i | B_k) \text{Var}(Z_j | B_k) + \sum_i \sum_{s \neq i} \text{Cov}(Y_i Z_i, Y_i | B_k) E(Z_s | B_k) \\
&+ \sum_i \sum_{j \neq i} \sum_{s \neq j} E(Z_j | B_k) E(Z_s | B_k) \text{Var}(Y_i | B_k) + \sum_j \sum_{i \neq j} \sum_{r \neq i} E(Y_i | B_k) E(Y_r | B_k) \text{Var}(Z_j | B_k) \\
&+ \sum_i \sum_{r \neq i} \sum_{s \neq i} \text{Cov}(Y_i Z_i, Y_r Z_s | B_k) + \sum_i \sum_{i \neq j, j \neq r} \sum_{r \neq i} \text{Cov}(Y_i Z_j, Y_r Z_r | B_k) \\
&+ \sum_j \sum_{r \neq j} \text{Cov}(Y_j Z_j, Z_j | B_k) E(Y_r | B_k) + \sum_i \sum_j \sum_{r \neq i} \sum_{s \neq j} \text{Cov}(Y_i Z_j, Y_r Z_s | B_k)
\end{aligned}$$

Les termes en orange valent 0.

Il reste à calculer :

- $\text{Var}(Y_i Z_i | B_k)$
- $\text{Var}(Y_i | B_k)$
- $\text{Var}(Z_j | B_k)$
- $E(Y_i | B_k)$
- $E(Z_j | B_k)$
- $\text{Cov}(Y_i Z_i, Y_i | B_k)$
- $\text{Cov}(Y_j Z_j, Z_j | B_k)$

Calcul de $\text{Var}(Y_i Z_i | B_k)$:

$$\text{Var}(Y_i Z_i | B_k) = N_{i,k}^{2\alpha-2} w_{i,k}^2 (N) \bar{X}_{i,k}^{2\alpha-2} w_{i,k}^2 (X) \text{Var}(N_{i,k+1} \bar{X}_{i,k+1} | B_k) \quad \text{avec}$$

$$\begin{aligned}
\text{Var}(N_{i,k+1} \bar{X}_{i,k+1} | B_k) &= E(N_{i,k+1}^2 \bar{X}_{i,k+1}^2 | B_k) - [E(N_{i,k+1} \bar{X}_{i,k+1} | B_k)]^2 \\
&= E[E(N_{i,k+1}^2 \bar{X}_{i,k+1}^2 | N_{i,k+1}) | B_k] - [E(E(N_{i,k+1} \bar{X}_{i,k+1} | N_{i,k+1}) | B_k)]^2 \\
&= E(N_{i,k+1}^2 | B_k) E(\bar{X}_{i,k+1}^2 | B_k) - [E(N_{i,k+1} | B_k)]^2 [E(\bar{X}_{i,k+1} | B_k)]^2 \\
&= \left(\text{Var}(N_{i,k+1} | B_k) + [E(N_{i,k+1} | B_k)]^2 \right) \times \left(\text{Var}(\bar{X}_{i,k+1} | B_k) + [E(\bar{X}_{i,k+1} | B_k)]^2 \right) \\
&\quad - [E(N_{i,k+1} | B_k)]^2 [E(\bar{X}_{i,k+1} | B_k)]^2 \\
&= \left(\frac{\sigma_k^{2(N)} \cdot N_{i,k}^2}{w_{i,k}^{(N)} \cdot N_{i,k}^\alpha} + f_k^{2(N)} N_{i,k}^2 \right) \left(\frac{\sigma_k^{2(X)} \cdot \bar{X}_{i,k}^2}{w_{i,k}^{(X)} \cdot \bar{X}_{i,k}^\alpha} + f_k^{2(X)} \bar{X}_{i,k}^2 \right) - [E(N_{i,k+1} | B_k)]^2 [E(\bar{X}_{i,k+1} | B_k)]^2
\end{aligned}$$

D'où :

$$\begin{aligned}
\text{Var}(Y_i Z_i | B_k) &= N_{i,k}^\alpha w_{i,k}^{(N)} \bar{X}_{i,k}^\alpha w_{i,k}^{(X)} \sigma_k^{2(X)} \sigma_k^{2(N)} + N_{i,k}^\alpha \bar{X}_{i,k}^{2\alpha} w_{i,k}^2 (X) w_{i,k}^{(N)} \sigma_k^{2(N)} f_k^{2(X)} \\
&\quad + N_{i,k}^{2\alpha} \bar{X}_{i,k}^\alpha w_{i,k}^2 (N) w_{i,k}^{(X)} \sigma_k^{2(X)} f_k^{2(N)}
\end{aligned}$$

Calcul de $Var(Y_i|B_k)$:

$$Var(Z_j|B_k) = N_{i,k}^{2-\alpha} w_{i,k}^{2(N)} Var\left(\frac{N_{i,k+1}}{N_{i,k}} \middle| B_k\right) = N_{i,k}^\alpha w_{i,k}^{(N)} \sigma_k^{2(N)}$$

Calcul de $Var(Z_j|B_k)$:

$$Var(Z_j|B_k) = \bar{X}_{i,k}^{2-\alpha} w_{i,k}^{2(X)} Var\left(\frac{\bar{X}_{i,k+1}}{\bar{X}_{i,k}} \middle| B_k\right) = \bar{X}_{i,k}^\alpha w_{i,k}^{(X)} \sigma_k^{2(X)}$$

Calcul de $E(Y_i|B_k)$:

$$E(Y_i|B_k) = N_{i,k}^\alpha w_{i,k}^{(N)} E\left(\frac{N_{i,k+1}}{N_{i,k}} \middle| B_k\right) = N_{i,k}^\alpha w_{i,k}^{(N)} f_k^{(N)}$$

Calcul de $E(Z_j|B_k)$:

$$E(Z_j|B_k) = \bar{X}_{i,k}^\alpha w_{i,k}^{(X)} E\left(\frac{\bar{X}_{i,k+1}}{\bar{X}_{i,k}} \middle| B_k\right) = \bar{X}_{i,k}^\alpha w_{i,k}^{(X)} f_k^{(X)}$$

Calcul de $Cov(Y_i Z_i, Y_i|B_k)$:

$$\begin{aligned} Cov(Y_i Z_i, Y_i|B_k) &= cov(f_{i,k} N_{i,k}^\alpha w_{i,k}^{(N)} \bar{X}_{i,k}^\alpha w_{i,k}^{(X)}, N_{i,k}^\alpha w_{i,k}^{(N)} f_k^{(N)} | B_k) \\ &= N_{i,k}^{2\alpha} w_{i,k}^{2(N)} \bar{X}_{i,k}^\alpha w_{i,k}^{(X)} cov\left(\frac{N_{i,k+1}}{N_{i,k}} \cdot \frac{\bar{X}_{i,k+1}}{\bar{X}_{i,k}}, \frac{N_{i,k+1}}{N_{i,k}} \middle| B_k\right) \\ &= N_{i,k}^{2\alpha-2} w_{i,k}^{2(N)} \bar{X}_{i,k}^{\alpha-1} w_{i,k}^{(X)} cov(N_{i,k+1} \bar{X}_{i,k+1}, N_{i,k+1} | B_k) \quad \text{avec} \end{aligned}$$

$$\begin{aligned} cov(N_{i,k+1} \bar{X}_{i,k+1}, N_{i,k+1} | B_k) &= E(N_{i,k+1}^2 \bar{X}_{i,k+1} | B_k) - E(N_{i,k+1} \bar{X}_{i,k+1} | B_k) E(N_{i,k+1} | B_k) \\ &= E(N_{i,k+1}^2 | B_k) E(\bar{X}_{i,k+1} | B_k) - E(N_{i,k+1} | B_k)^2 E(\bar{X}_{i,k+1} | B_k) \\ &= Var(N_{i,k+1} | B_k) E(\bar{X}_{i,k+1} | B_k) \\ &= \frac{\sigma_k^{2(N)} \cdot N_{i,k}^2}{w_{i,k}^{(N)} \cdot N_{i,k}^\alpha} \times \bar{X}_{i,k} f_k^{(X)} \end{aligned}$$

D'où :

$$Cov(Y_i Z_i, Y_i|B_k) = N_{i,k}^\alpha w_{i,k}^{(N)} \bar{X}_{i,k}^\alpha w_{i,k}^{(X)} \sigma_k^{2(N)} f_k^{(X)}$$

Calcul de $Cov(Y_i Z_i, Z_i | B_k)$:

$$\begin{aligned}
 Cov(Y_i Z_i, Z_i | B_k) &= cov(f_{i,k} N_{i,k}^\alpha w_{i,k}^{(N)} \bar{X}_{i,k}^\alpha w_{i,k}^{(X)}, \bar{X}_{i,k}^\alpha w_{i,k}^{(X)} f_k^{(X)} | B_k) \\
 &= N_{i,k}^\alpha w_{i,k}^{2(X)} \bar{X}_{i,k}^{2\alpha} w_{i,k}^{(N)} cov\left(\frac{N_{i,k+1}}{N_{i,k}} \cdot \frac{\bar{X}_{i,k+1}}{\bar{X}_{i,k}}, \frac{N_{i,k+1}}{N_{i,k}} \mid B_k\right) \\
 &= \bar{X}_{i,k}^{2\alpha-2} w_{i,k}^{2(X)} N_{i,k}^{\alpha-1} w_{i,k}^{(N)} cov(N_{i,k+1} \bar{X}_{i,k+1}, \bar{X}_{i,k+1} | B_k) \quad \text{avec}
 \end{aligned}$$

$$cov(N_{i,k+1} \bar{X}_{i,k+1}, N_{i,k+1} | B_k) = \frac{\sigma_k^{2(X)} \cdot \bar{X}_{i,k}^2}{w_{i,k}^{(X)} \cdot \bar{X}_{i,k}^\alpha} \times N_{i,k} f_k^{(N)}$$

D'où :

$$Cov(Y_i Z_i, Z_i | B_k) = N_{i,k}^\alpha w_{i,k}^{(N)} \bar{X}_{i,k}^\alpha w_{i,k}^{(X)} \sigma_k^{2(X)} f_k^{(N)}$$

Finalemnt, en rassemblant tous les termes :

$$Var(\hat{f}_k | B_k) = \frac{\sigma_k^{2(N)} \sigma_k^{2(X)}}{\left(\sum_{i=1}^{n-k} N_{i,k}^\alpha w_{i,k}^{(N)}\right) \left(\sum_{i=1}^{n-k} \bar{X}_{i,k}^\alpha w_{i,k}^{(X)}\right)} + \frac{\sigma_k^{2(N)} f_k^{2(X)}}{\left(\sum_{i=1}^{n-k} N_{i,k}^\alpha w_{i,k}^{(N)}\right)} + \frac{\sigma_k^{2(X)} f_k^{2(N)}}{\left(\sum_{i=1}^{n-k} \bar{X}_{i,k}^\alpha w_{i,k}^{(X)}\right)}$$

L'estimateur de F^2 est donc donné par :

$$\hat{F}^2 = \sum_{k=n+1-i}^{n-1} \hat{f}_{n+1-i}^2 \cdots \hat{f}_{k-1}^2 Var(\hat{f}_k | B_k) f_{k+1}^2 \cdots f_{n-1}^2$$

□