Institut des ACTUAIRES

Dauphine | PSL
UNIVERSITÉ PARIS

Mémoire présenté devant l'Université de Paris-Dauphine pour l'obtention du Certificat d'Actuaire de Paris-Dauphine et l'admission à l'Institut des Actuaires le 26 juin 2025

Par : Ghislain ZELLER
Titre : Évaluation du paramètre spécifique (USP) des réserves d'une captive par apprentissage profond

Confidentialité : ☑ Non ☐ Oui (Durée: ☐ 1 an ☐ 2 ans)

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

Membres présents du jury de l'Institut des Actuaires:

Entreprise :
Nom: Marsh France
Signature :

MARSH S.A.S
Au capital de 5 917 915 euros
Tour Ariane - La Défense 9
92088 Paris la Défense Cedex
RCS Nanterre 572 174 415
ORIAS n° 07 001 037

Directeur de Mémoire en entreprise :
Nom: Lucas MIETTON
Signature:

Membres présents du Jury du Certificat d'Actuaire de Paris-Dauphine:

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

Secrétariat :

Signature du responsable entreprise

Bibliothèque :

Signature du candidat

Résumé

Dans la formule standard, le coefficient d'écart type (ou de volatilité) calculé pour mesurer le risque de réserve (module risque de souscription) est fixé par ligne d'activité et unique pour tous les organismes d'assurance. Cela peut avoir pour incidence un coût du capital inadapté aux profils de risque spécifiques. Une démarche auprès du superviseur pour appliquer un paramètre de volatilité spécifique, l'USP (*Undertaking Specific Parameter*), peut être engagée par les organismes ne disposant pas de moyens nécessaires pour se doter d'un modèle interne. C'est le cas des captives d'assurance et de réassurance, véhicules de transfert de risque internes aux entreprises, qui peuvent faire une demande pour appliquer les méthodes de calcul du paramètre. Celles-ci sont précisées dans les actes délégués (EIOPA, 2015).

Alors que de plus en plus de techniques d'apprentissage statistique sont utilisées en provisionnement non-vie, les méthodes réglementaires de calcul des paramètres USP restent quant à elles toujours basées sur des approches historique ou paramétrique. Le récent modèle Mack-Net (Ramos-Pérez et al., 2022), construit à partir de réseaux de neurones récurrents (RNN) et de cellules LSTM (*Long Short-Term Memory*), est une approche fournissant de nouveaux facteurs de développement et coefficients de volatilité pouvant sur-performer les modèles traditionnels de marché comme Mack Chain-Ladder. Ce modèle d'apprentissage profond permet donc de proposer une nouvelle manière de calculer les paramètres USP.

L'objet de ce mémoire est d'étudier la force des méthodes de calcul du paramètre USP pour le risque de réserve sur les données d'une captive, et de proposer une nouvelle approche plus précise basée sur l'apprentissage profond. A partir de plusieurs métriques de performances, notamment le CDR (*Claims Development Result*), le mémoire démontre que les méthodes d'apprentissage profond sous-jacentes au modèle Mack-Net permettent en effet de mieux évaluer le besoin en capital de la captive étudiée.

Mots-clés : *USP*, *risque de réserve*, *captive*, *Deep Learning*, RNN, *LSTM*.

Abstract

In the standard formula, the standard deviation (or volatility) coefficient for reserve risk, as part of the underwriting risk module, is set by line of business. This can result in a cost of capital that is unsuited to the risk profile of insurance organizations. Organizations that do not have the necessary resources to set up an internal model can apply to the supervisor for the application of an USP parameter (*Undertaking Specific Parameter*) to quantify their reserving volatility. This is the case for insurance and reinsurance captives, which can apply to use these methods that are specified in the delegated acts (EIOPA, 2015).

While more and more statistical and deep learning techniques are being used in non-life reserving to estimate future cashflow, regulatory methods for calculating USPs are still based on historical or parametric approaches. The Mack-Net model, built from recurrent neural networks (RNN) and Long Short-Term Memory (LSTM) cells, is an approach that provides new development factors and volatility coefficients that can outperform traditional models on the market, thereby making it possible to propose a new way of calculating USP parameters.

The aim of this thesis is to study the strength of USP methods for reserve risk on a captive's data, and to propose a new, more accurate approach (including the use of deep learning) to assess reserving volatility differently. Using several performance metrics, in particular the CDR (Claims Development Result), this dissertation shows that deep learning methods underlying the Mack-Net model can best assess the capital requirement of the captive studied.

Keywords : *USP*, *reserving*, *captives* , *Deep Learning*, RNN, *LSTM*.

Note de Synthèse

Contexte

Naval Groupe, la Ligue de Football Professionnelle (LFP), les groupes la Poste, Orange, ou encore Safran ... De plus en plus d'entreprises françaises choisissent, en 2023 et 2024, de créer leur propre organisme d'assurance interne à leur organisation : les captives. En tant que courtier d'expertise internationale, *Marsh McLennan* (et son entité *Risk Analytics*) accompagnent ces entreprises dans la mise en place de tels véhicules, notamment dans l'externalisation de la fonction actuarielle et des livrables requis par la Directive Solvabilité II. Les captives répondent au principe de proportionnalité, et dans ce contexte, certaines captives peuvent contourner la formule standard et se différencier dans le calcul du capital réglementaire via l'application de paramètres spécifiques, les USP (*Undertaking Specific Parameters*). Alors que le contexte actuel est propice à une refonte globale de la Directive Solvabilité II (L'Argus de l'assurance, 2025), le mémoire confronte en particulier les méthodes de calcul du paramètre USP associé à la volatilité des réserves pour une captive souscrivant du risque responsabilité automobile (RC). Au-delà d'un état de l'art des méthodes, le présent travail contribue à proposer une nouvelle approche de calcul, à l'aide des méthodes d'apprentissage profond. Mathématiquement, le paramètre spécifique à calculer dans le cadre du capital réglementaire SCR en non-vie est $\sigma_{res,USP}$, et est défini via

$$\mathrm{SCR}_{\text{nl prem res}} = 3 \cdot \sigma_{\text{nl}} \cdot V_{\text{nl}}, \tag{1}$$

avec

$$\begin{cases} \sigma_{\text{nl}} = \frac{1}{V_{\text{nl}}} \sqrt{\sum_{s,t} \mathrm{Corr}(s,t) \cdot \sigma_s \cdot V_s \cdot \sigma_t \cdot V_t} \\ \sigma_s = \sqrt{\frac{\sigma^2_{\text{prem},s} \cdot V^2_{\text{prem},s} + \sigma_{\text{prem},s} \cdot V_{\text{prem},s} \cdot \sigma_{res,USP} \cdot V_{\text{res},s} + \sigma^2_{res,USP} \cdot V^2_{\text{res},s}}{V_{\text{prem},s} + V_{\text{res},s}}} \end{cases},$$

où $V_i, i \in \{nl, prem, res\}$, $V_{j,s}$, $\sigma_{j,s}, j \in \{prem, res\}$ et $\mathrm{Corr}(s,t)$ représentent respectivement les volumes pour le risque de primes et/ou de réserves en non-vie, les volumes ou écarts-types pour le risque de primes et/ou de réserves et enfin le coefficient de corrélation pour le risque de primes et de réserves en non-vie entre les branches d'assurance s et t. Les actes délégués définissent deux méthodes réglementaires et déterministes basées sur les triangles de liquidation. La première méthode réglementaire (ou méthode 1 ou méthode log-normale) consiste à calculer

$$\sigma_{res,USP} = c \cdot \hat{\sigma}(\delta, \gamma) \cdot \sqrt{\frac{T+1}{T-1}} + (1-c) \cdot \sigma_{res,s}, \tag{2}$$

où $\hat{\sigma}$ est le paramètre spécifique à estimer, $\sigma_{res,s}$ le paramètre de marché (9% en RC automobile), c un facteur de crédibilité par ligne d'activité variant selon la longueur de l'historique de sinistralité, δ, γ respectivement les paramètres de mélange et de variation logarithmique du modèle et T le nombre d'années comptables pour lesquelles les données sont disponibles.

La seconde méthode réglementaire (ou méthode 2 ou méthode de Merz-Wüthrich ou méthode des

triangles) des actes délégués consiste à calculer, à partir de la méthode de Mack Chain-Ladder

$$\sigma_{res,USP} = c \cdot \frac{\sqrt{\text{MSEP(CDR)}}}{\left(\sum_{i=0}^{I}(\hat{C}_{i,J} - C_{i,J-i})\right)} + (1-c)\cdot \sigma_{\text{res,s}}, \tag{3}$$

avec

$$\text{MSEP(CDR)} = \sum_{i=1}^{I} \hat{C}_{i,J}^2 \cdot \frac{\hat{Q}_{I-i}}{C_{i,I-i}} + \sum_{i=1}^{I}\sum_{k=1}^{I} \hat{C}_{i,J}\cdot\hat{C}_{k,J}\cdot\left(\frac{\hat{Q}_{I-i}}{S_{I-i}} + \sum_{j=I-i+1}^{J-1}\frac{C_{I-j,j}}{S'_j}\cdot\frac{\hat{Q}_j}{S_j}\right), \tag{4}$$

et où $\hat{Q}_j = \frac{\hat{\sigma}_j^2}{\hat{f}_j^2}$. Le CDR correspond à l'écart de ré-estimation entre deux provisions successives d'une année comptable à l'autre et sa MSEP (*Mean Square Error Prediction*) à une marge de solvabilité en cas de déviation du CDR. Les notations du modèle de Mack Chain-Ladder sont rappelées, notamment

$$\hat{C}_{i,J} = C_{i,I-i}\hat{f}_{I-i}\cdots\hat{f}_{J-1},$$

$$\begin{cases} \hat{f}_j = \dfrac{\sum_{i=0}^{I-j} C_{i,j+1}}{\sum_{i=0}^{I-j} C_{i,j}} \\ \hat{\sigma}_j^2 = \dfrac{1}{J-j-1}\sum_{i=0}^{J-j-1} C_{i,j}\left(\dfrac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j\right)^2 \end{cases},$$

et pour $0 \leq j \leq J-1$

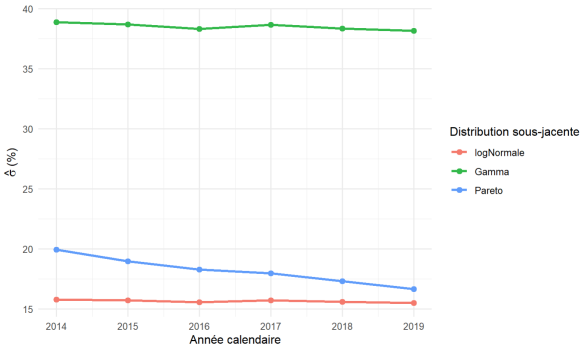
$$\begin{cases} S_j = \sum_{i=0}^{I-j-1} C_{i,j} \\ S'_j = \sum_{i=0}^{I-j} C_{i,j} \end{cases}.$$

Le mémoire souhaite confronter les deux méthodes réglementaires (2) et (3) avec d'autres méthodes candidates utilisées par la captive, qu'elles soient déterministes, stochastiques (état de l'art) ou reposant sur l'apprentissage profond (nouvelle approche). Les données utilisées sont les sinistres RC automobile du portefeuille de la captive.

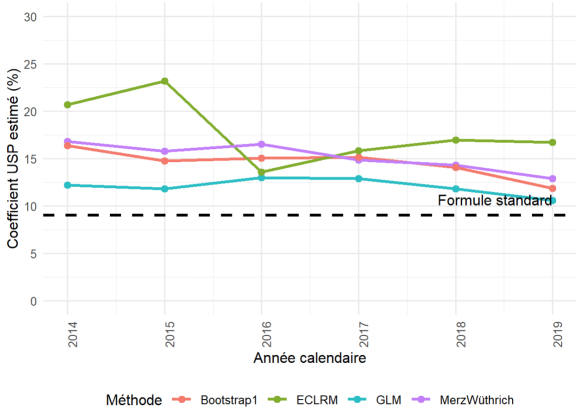
Critiques des méthodes des actes délégués et état de l'art des autres méthodes existantes

La première partie du travail de modélisation consiste à dresser un état de l'art des méthodes concurrentes aux deux méthodes réglementaires exposées en formules (2) et (3). Il existe deux méthodes concurrentes pouvant être comparées à la méthode 1, en modifiant l'hypothèse de loi log-normale sous-jacente : l'approche Gamma et l'approche Pareto. La figure 1a indique les résultats obtenus pour le calcul du paramètre spécifique, réalisé sur des triangles à différents exercices. Les résultats indiquent qu'un changement d'hypothèses peut provoquer des résultats très hétérogènes sur le coefficient de volatilité des réserves.

Concernant la méthode 2 des actes délégués, il existe trois méthodes concurrentes qui permettent de quantifier la MSEP décrites dans l'équation (4) : l'approche par modèle linéaire généralisé (GLM) quasi-Poisson, le bootstrap à un an, et la méthode des charges ECLRM (*Extended-Complementary-Loss-Ratio Method*). Les résultats des paramètres USP en figure 1b indiquent des résultats plutôt similaires pour les méthodes Chain-Ladder, bootstrap à un an et le GLM. La méthode ECLRM, prenant en compte à la fois le triangle des charges et des paiements fournit un paramètre spécifique moins stable pour la captive.



(a) Première méthode des actes délégués (log-normale) et alternatives



(b) Seconde méthode des actes délégués (Merz-Wüthrich) et alternatives

FIGURE 1 : Comparaison des deux méthodes des actes délégués et alternatives

Apport de l'apprentissage profond dans l'évaluation des réserves de la captive

En explorant le modèle Mack-Net (RAMOS-PÉREZ et al., 2022) basé sur les méthodes d'apprentissage profond (*Deep Learning*), le mémoire propose une nouvelle manière de calculer le paramètre USP de volatilité des réserves de la captive. A partir de la formule (4), le but est de recalibrer les paramètres du modèle traditionnel de Mack Chain-Ladder (facteurs de développement et de variance) à partir de réseaux de neurones récurrents (RNN) et de cellules LSTM. L'architecture du modèle est schématisée en figure 2. Pour mettre en place le modèle, on distingue trois ensembles de données, décrits dans la figure 3 : entraînement, validation et test. Les performances des prédictions du modèle Mack-Net sont ensuite comparées à des modèles de marché.

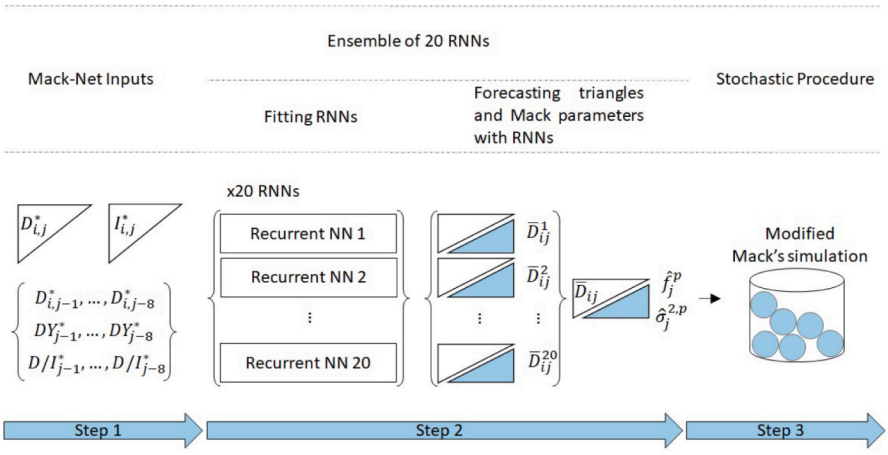


FIGURE 2 : Déploiement du modèle Mack-Net (RAMOS-PÉREZ et al., 2022)

Dans ce modèle, les notations sont les suivantes :

- $C_{i,j}^{Pa}$ représentent les paiements cumulés,
- $C_{i,j}^{In}$ représentent les montants de charges cumulés,

- P_i représente un vecteur d'exposition,
- $D^*_{i,j}$ vaut $\frac{C^{Pa}_{i,j}}{P_i}$,
- DY_j représente les années de développement.

Les données sont entraînées dans K RNN qui fournissent chacun un triangle de montants cumulés prédit $\overline{D}^k$ où $k \in \{1, ..., K\}$. Les triangles prédits permettent d'obtenir un triangle moyenné $\overline{D}$, comme base pour le calcul de nouveaux facteurs de développement et de coefficients de volatilité Mack-Net définis par

$$\hat{f}^p_j = \frac{\sum_{i=I-j+2}^{I} \overline{D}_{i,j}}{\sum_{i=I-j+2}^{I} \overline{D}_{i,j-1}},$$

$$\hat{\sigma}^{2,p}_j = \frac{1}{I-j-1} \sum_{i=0}^{I} \overline{D}_{i,j} \left(\frac{\overline{D}_{i,j}}{\overline{D}_{i,j-1}} - \overline{f}_j \right)^2,$$

avec $\overline{f}_j = \frac{\sum_{i=0}^{I} \overline{D}_{ij}}{\sum_{i=0}^{I} \overline{D}_{ij-1}}$ et $\overline{D}_{ij} = \frac{\sum_{k=1}^{K} \overline{D}^k_{ij}}{K}$.

L'idée finale du mémoire est de choisir les paramètres optimaux des réseaux de neurones qui minimisent une métrique de provisionnement adéquate. D'après la formule (4), la métrique finale choisie est le CDR, puisque le but est de minimiser la variance (MSEP) et ainsi obtenir un paramètre USP plus précis. Cette métrique CDR est définie par

$$\begin{aligned} CDR^{I+1}_i &= R^{I+1}_{i,j*} - R^{I}_{i,j*} + \left(Y_{i,j*} - \hat{Y}^I_{i,j*} \right) \\ &= R^{I+1}_{i,j*} - R^{I}_{i,j*} + AvE^{I+1}_{i,j*}, \end{aligned}$$

où $j^* = I - i$ avec I représentant la dernière année de développement du triangle, i une année d'accident, les paiements incrémentaux $Y_{i,I-i+1} = C_{i,I-i+1} - C_{i,I-i}$, les réserves $R^I_i = C_{i,I} - C_{i,I-i}$ et AvE l'écart entre la première diagonale estimée et observée (*Actual versus Expected*).

En d'autres termes,

$$CDR = AvE + \Delta IBNR.$$

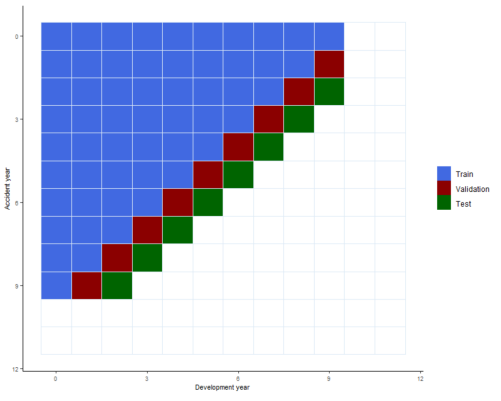


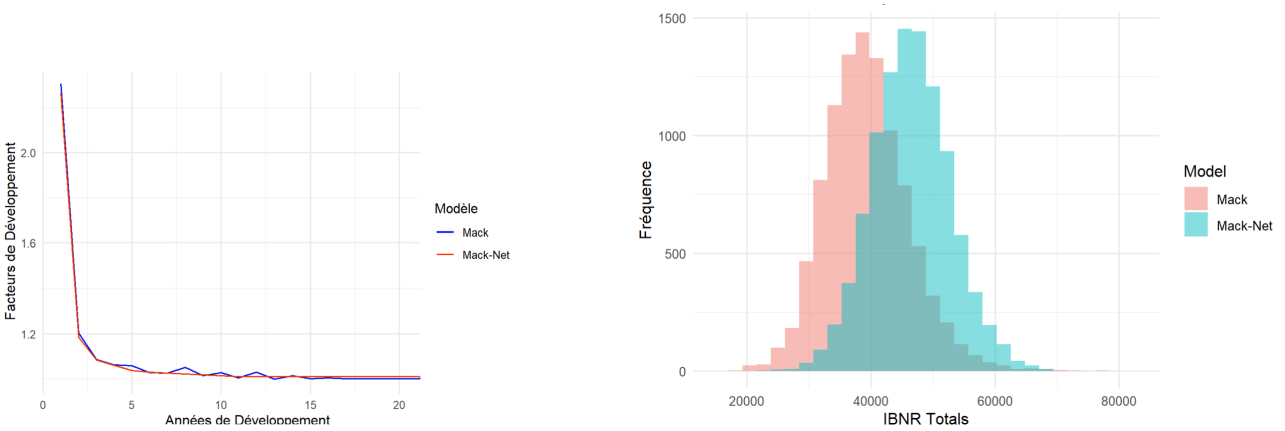
FIGURE 3 : Entraînement, validation et test (PITTARELLO, 2023) sur triangle de provisionnement

En particulier, les métriques usuelles d'optimisation du provisionnement qui en découlent sont

$$CDR_{\text{score}} = \sqrt{\frac{\sum_{i=1}^{I} |Y_{i,j*}| \cdot (CDR^{I+1}_i)^2}{\sum_{i=1}^{I} |Y_{i,j*}|}} \text{ et } \quad AvE_{\text{score}} = \sqrt{\frac{\sum_{i=1}^{I} |Y_{i,j^*}| \cdot (AvE^{I+1}_i)^2}{\sum_{i=1}^{I} |Y_{i,j^*}|}}.$$

Nouvelle approche de calcul du paramètre USP avec l'apprentissage profond

Entre les deux modèles Mack et Mack-Net, les facteurs de développement estimés et comparés en figure 4a convergent vers 1. Le modèle Mack-Net, qui considère en entrée à la fois les triangles des charges et des paiements, lisse l'aléa statistique lié aux facteurs de développement et atténue les valeurs trop volatiles dans le modèle de Mack. Par la suite, les facteurs recalibrés avec l'approche Mack-Net sont incorporés dans la formule (4). La prise en compte des triangles des charges en plus des paiements dans les prédictions des réserves est une information pertinente, notamment en RC automobile, puisque certains sinistres peuvent rester ouverts longtemps, notamment pour les dommages corporels. Enfin, des simulations de bootstrap peuvent être effectuées comme en figure 4b avec le modèle Mack-Net afin d'obtenir une distribution des réserves.



(a) Facteurs de lien de Mack et Mack-Net

(b) Procédure bootstrap de Mack et Mack-Net

FIGURE 4 : Comparaison des paramètres du modèle de Mack Chain-Ladder et Mack-Net

Au-delà d'une simple comparaison entre la méthode de Mack Chain-Ladder et sa version *Deep Learning* via le modèle Mack-Net, il est intéressant, dans le cadre d'un *benchmark* en table 1, de confronter les scores de performances avec d'autres modèles de marché, certains basés sur l'apprentissage profond, comme le modèle bCCNN (GABRIELLI et al., 2019) ou *DeepTriangle* (KUO, 2018), et d'autres plus traditionnels utilisant notamment, comme dans le modèle Mack-Net, le triangle des charges : la méthode ECLRM et celle de Munich Chain-Ladder. Dans le modèle Mack-Net, la prise en compte de ces deux informations détériore moins les performances que pour un modèle de type Munich ou ECLRM, grâce à la flexibilité des structures en réseaux de neurones. La combinaison du *Deep Learning* et la prise en compte des charges conduit finalement à de meilleures performances.

Métriques	**Mack-Net**	**Mack Chain-Ladder**	**ECLRM**	**Munich Chain-Ladder**	**bCCNN**	***DeepTriangle***
AvE_{score}	504,7	501,4	753,7	490,0	722,3	1631
CDR_{score}	729,8	770,9	1381,1	1075,8	1296,2	1776
MAE^{1yr}	141,6	163,5	304,9	129,4	342,0	257
$MAPE^{\text{1yr}}$	1,6%	2,1%	3,9%	1,6%	4,7%	2,7%
MAE^{ult}	470,2	175,0	1070,9	181,0	1249,3	354
$MAPE^{\text{ult}}$	5,7%	2,3%	11,6%	1,5%	12,7%	1,7%

TABLE 1 : Comparaison des performances moyennes du modèle Mack-Net (en milliers €)

Le paramètre USP, estimé par exercice à partir des $\hat{Q}_j$ de l'équation (4) nouvellement calculés dans le modèle Mack-Net, peut ainsi être comparé en figure 5 avec les autres approches standards. Dans le modèle Mack-Net, les hyperparamètres du modèle minimisent le CDR_{score} ce qui fiabilise l'approche.

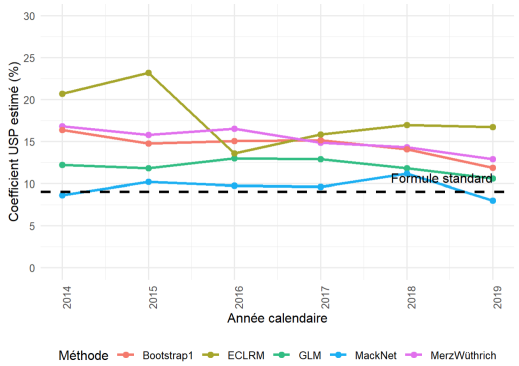


FIGURE 5 : Paramètre USP obtenu avec le nouveau modèle Mack-Net

Conclusion

Dans le cadre du calcul du SCR pour le risque de réserve d'une captive, le présent travail met en lumière les méthodes qui peuvent se substituer à l'utilisation du paramètre fixé par les actes délégués. Dans un contexte où de plus en plus de captives d'assurance ou de réassurance voient le jour sur le territoire national et où la Directive Solvabilité II devient moins contraignante pour ces dernières, le but du mémoire est d'explorer de nouvelles méthodes pour estimer le paramètre USP des réserves de la captive, notamment via l'apprentissage profond. Il s'avère que le paramètre obtenu avec le modèle Mack-Net est plus précis que l'approche réglementaire, car le modèle permet de réduire l'erreur liée au calcul de la métrique CDR, fondamentale dans le calcul du paramètre USP. La difficulté en matière de qualité des données et de vérification d'hypothèses rend l'approbation par le superviseur du paramètre USP dans la pratique plus ardue. Il existe aujourd'hui sur le marché une variété importante de méthodes pour quantifier la volatilité des réserves à un an, mais le mémoire a pour objectif d'étudier la contribution des méthodes de *Deep Learning* comme celles avec les réseaux de neurones récurrents sur un sujet réglementaire.

Synthesis note

Context

Naval Groupe, Ligue de Football Professionnelle (LFP), La Poste, Orange, Safran, ... In 2023 and 2024, more and more French companies are choosing to set up their own in-house insurance vehicles, known as captives. As a broker with international expertise, *Marsh* and its entity *Risk Analytics* support these companies in setting up such vehicles, particularly in the actuarial function and the outsourcing of Solvency II Directive deliverables. These insurance organizations are subject to the principle of proportionality, and in this context, certain captives can bypass the standard formula and differentiate themselves in the calculation of regulatory capital through the application of specific parameters, the USPs (*Undertaking Specific Parameters*). Taking these various elements into account, and adding to this a context in which the Solvency II Directive is being reviewed (L'Argus de l'assurance, 2025), the present thesis questions in particular the methods for using the USP parameter of reserve volatility for a captive underwriting motor liability risk, and goes further by proposing a new calculation method using deep learning. Mathematically, the specific parameter to be calculated within the framework of the regulatory SCR capital in non-life is $\sigma_{res,USP}$ and is defined by

$$\mathrm{SCR}_{\text{nl prem res}} = 3 \cdot \sigma_{\text{nl}} \cdot V_{\text{nl}}, \tag{5}$$

avec

$$\begin{cases} \sigma_{\text{nl}} = \frac{1}{V_{\text{nl}}} \sqrt{\sum_{s,t} \mathrm{Corr}(s,t) \cdot \sigma_s \cdot V_s \cdot \sigma_t \cdot V_t} \\ \sigma_s = \sqrt{\frac{\sigma^2_{\text{prem},s} \cdot V^2_{\text{prem},s} + \sigma_{\text{prem},s} \cdot V_{\text{prem},s} \cdot \sigma_{res,USP} \cdot V_{\text{res},s} + \sigma^2_{res,USP} \cdot V^2_{\text{res},s}}{V_{\text{prem},s} + V_{\text{res},s}}} \end{cases},$$

where $V_i,\ i \in \{nl,\ prem,\ res\}$, $V_{j,s}$, $\sigma_{j,s},\ j \in \{prem,\ res\}$, and $\mathrm{Corr}(s,t)$ respectively represent the volumes for the premium and/or reserve risk in non-life insurance, the volumes or standard deviations for the premium and/or reserve risk and finally the correlation coefficient for the premium and reserve risk in non-life insurance between lines s and t. The delegated acts define two deterministic methods based on liquidation triangles. The first official method consists of calculating

$$\sigma_{res,USP} = c \cdot \hat{\sigma}(\delta, \gamma) \cdot \sqrt{\frac{T+1}{T-1}} + (1-c) \cdot \sigma_{res,s}. \tag{6}$$

where $\hat{\sigma}$ is the specific parameter to be estimated, $\sigma_{res,s}$ is the market parameter (9% in automobile liability insurance), δ, γ respectively represent the mixing and log-variation parameters of the model, c is a credibility factor per line of business that varies according to the length of the claims history, and T is the number of accounting years for which data is available.

The second method of delegated acts consists in calculating, based on the Mack Chain-Ladder method

$$\sigma_{res,USP} = c \cdot \frac{\sqrt{\mathrm{MSEP(CDR)}}}{\left(\sum_{i=0}^{I} (\hat{C}_{i,J} - C_{i,J-i})\right)} + (1-c) \cdot \sigma_{\text{res,s}},$$

where

$$\text{MSEP(CDR)} = \sum_{i=1}^{I} \hat{C}_{i,J}^2 \cdot \frac{\hat{Q}_{I-i}}{C_{i,I-i}} + \sum_{i=1}^{I} \sum_{k=1}^{I} \hat{C}_{i,J} \cdot \hat{C}_{k,J} \cdot \left(\frac{\hat{Q}_{I-i}}{S_{I-i}} + \sum_{j=I-i+1}^{J-1} \frac{C_{I-j,j}}{S'_j} \cdot \frac{\hat{Q}_j}{S_j} \right), \tag{7}$$

where $\hat{Q}_j = \frac{\hat{\sigma}_j^2}{\hat{f}_j^2}$. The CDR corresponds to the re-estimation deviation between two successive provisions from one accounting year to the next, and the MSEP (Mean Square Error Prediction) represents a solvency margin in case of a deviation in the CDR. The notations of the Mack Chain-Ladder model are recalled, namely

$$\hat{C}_{i,J} = C_{i,I-i} \hat{f}_{I-i} \dots \hat{f}_{J-1},$$

$$\begin{cases} \hat{f}_j = \dfrac{\sum_{i=0}^{I-j} C_{i,j+1}}{\sum_{i=0}^{I-j} C_{i,j}} \\ \hat{\sigma}_j^2 = \dfrac{1}{J-j-1} \sum_{i=0}^{J-j-1} C_{i,j} \left(\dfrac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j \right)^2 \end{cases},$$

and for $0 \leq j \leq J-1$

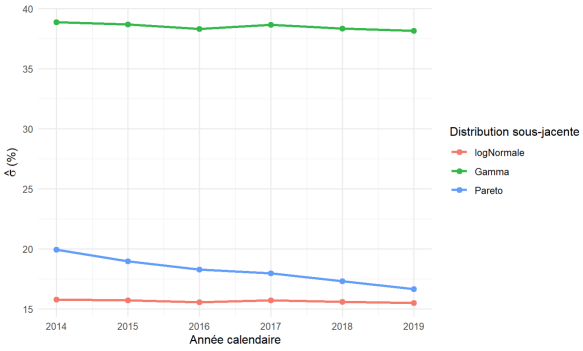
$$\begin{cases} S_j = \sum_{i=0}^{I-j-1} C_{i,j} \\ S'_j = \sum_{i=0}^{I-j} C_{i,j} \end{cases}.$$

The thesis aims to compare the two regulatory methods (2) and (3) with other candidate approaches used by the captive, whether they are deterministic, stochastic (state-of-the-art), or based on deep learning (new approach). The data used consist of motor third-party liability (MTPL) claims from the captive's portfolio.

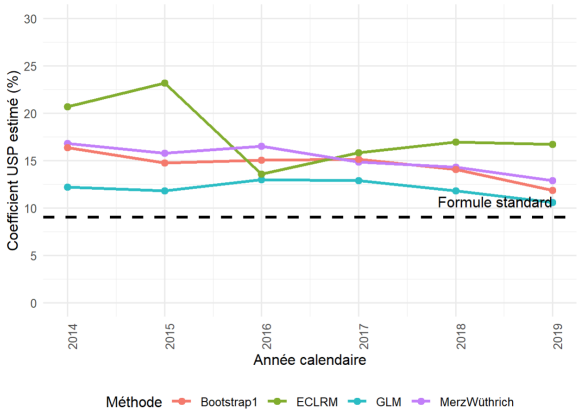
Criticism of the delegated acts methods and state-of-the-art of other existing methods

The first part of the modeling consists of a state-of-the-art review of various alternative methods to the two delegated acts methods. There are two alternative methods that can be compared to Method 1 by modifying the underlying assumption of the parametric model: the Gamma approach and the Pareto approach. Figure 6a below shows the results obtained for the calculation of the specific parameter, performed on triangles at different evaluation exercises. The results indicate that a change in assumptions can lead to highly heterogeneous outcomes regarding the reserve volatility coefficient.

Regarding the second delegated acts method, there are three alternative methods to quantify the MSEP: the quasi-Poisson GLM, the one-year bootstrap, and the ECLRM method (*Extended-Complementary-Loss-Ratio Method*). The results in figure 6b indicate fairly similar outcomes for the Chain-Ladder, one-year bootstrap, and GLM methods. The implemented ECLRM method provides a less stable specific parameter.



(a) First delegated acts method (log-normal) and alternatives



(b) Second delegated acts method (Merz-Wüthrich) and alternatives

Figure 6: Comparison of the two methods of delegated acts and alternatives

Contribution of deep learning methods to reserve evaluation

By studying a *Deep Learning* model, the Mack-Net model (Ramos-Pérez et al., 2022), this study proposes a new way to calculate the USP volatility coefficient of the captive's reserves. Based on formula (7), the goal is to recalibrate the parameters of the traditional Mack Chain-Ladder model (development and variance factors) using recurrent neural networks (RNNs) and LSTM cells. The model architecture is illustrated in figure 7. To implement the model, three datasets are distinguished, as described in figure 8: training, validation, and test. The performance of the Mack-Net model is then compared to market models.

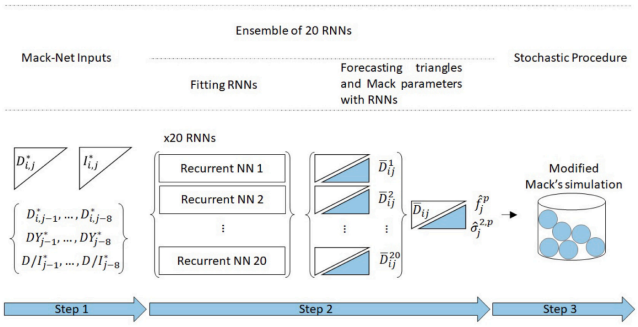


Figure 7: Deployment of the Mack-Net model (Ramos-Pérez et al., 2022)

In this model, the notations are as follows:

- $C_{i,j}^{Pa}$ represent the cumulative payments,
- $C_{i,j}^{In}$ represent the cumulative incurred amounts,
- P_i represents an exposure vector,
- $D_{i,j}^{*}$ is equal to $\frac{C_{i,j}^{Pa}}{P_i}$,
- DY_j represents the development years.

The data are trained through K RNN models, each providing a predicted triangle of cumulative amounts $\overline{D}^k$ where $k \in \{1, ..., K\}$. The predicted triangles allow for computing an averaged triangle $\overline{D}$, used as a basis for calculating new development factors and Mack-Net volatility coefficients defined by

$$\hat{f}_j^p = \frac{\sum_{i=I-j+2}^{I} \overline{D}_{i,j}}{\sum_{i=I-j+2}^{I} \overline{D}_{i,j-1}},$$

$$\hat{\sigma}_j^{2,p} = \frac{1}{I-j-1} \sum_{i=0}^{I} \overline{D}_{i,j} \left(\frac{\overline{D}_{i,j}}{\overline{D}_{i,j-1}} - \overline{f}_j \right)^2,$$

with $\overline{f}_j = \frac{\sum_{i=0}^{I} \overline{D}_{ij}}{\sum_{i=0}^{I} \overline{D}_{ij-1}}$ and $\overline{D}_{ij} = \frac{\sum_{k=1}^{K} \overline{D}_{ij}^k}{K}$.

According to formula (7), the final idea of this study is to select the optimal neural network parameters that minimize the CDR provisioning metric in order to reduce its variance and thus obtain more precise USP coefficients. As a result, both ultimate and one-year horizon estimated reserves are more accurate. This CDR metric is defined as

$$\begin{aligned} CDR_i^{I+1} &= R_{i,j*}^{I+1} - R_{i,j*}^{I} + \left(Y_{i,j*} - \hat{Y}_{i,j*}^{I} \right) \\ &= R_{i,j*}^{I+1} - R_{i,j*}^{I} + AvE_{i,j*}^{I+1}, \end{aligned}$$

where $j^* = I - i$, with I representing the last development year of the triangle, i an accident year, the incremental payments $Y_{i,I-i+1} = C_{i,I-i+1} - C_{i,I-i}$, the reserves $R_i^I = C_{i,I} - C_{i,I-i}$ and AvE the comparison between the estimated and observed diagonal (*Actual versus Expected*).

In other words,

$$CDR = AvE + \Delta IBNR.$$

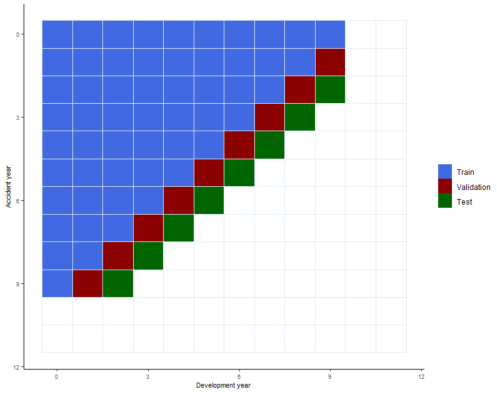


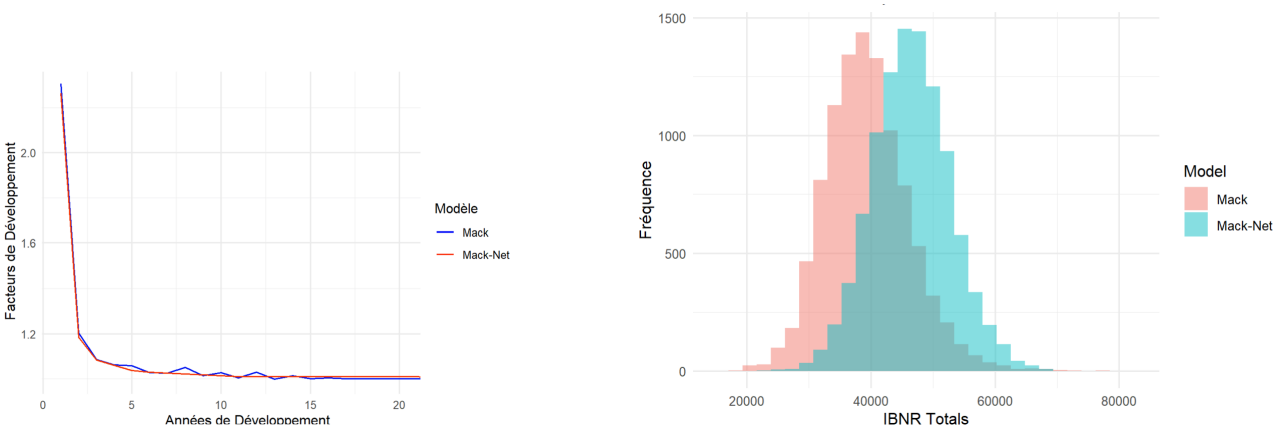
Figure 8: Training, validation, and testing (Pittarello, 2023) on the runoff triangle

In particular, the usual reserving optimization metrics that result are

$$CDR_{\text{score}} = \sqrt{\frac{\sum_{i=1}^{I} |Y_{i,j*}| \cdot (CDR_i^{I+1})^2}{\sum_{i=1}^{I} |Y_{i,j*}|}} \quad \text{and} \quad AvE_{\text{score}} = \sqrt{\frac{\sum_{i=1}^{I} |Y_{i,j^*}| \cdot (AvE_i^{I+1})^2}{\sum_{i=1}^{I} |Y_{i,j^*}|}}.$$

New approach for calculating the USP coefficient with deep learning

Between the two models, Mack and Mack-Net, the estimated and compared in figure 9a development factors converge towards 1. The Mack-Net model, which takes both the incurred and paid triangles as inputs, smooths out the statistical randomness associated with development factors and mitigates excessively volatile values in the Mack model. Thus, the recalibrated factors using the Mack-Net approach are incorporated into equation (7). Considering the incurred triangles in addition to the payments in reserve predictions is relevant information, particularly in motor liability insurance, as some claims can remain open for a long time, especially those involving bodily injuries. Finally, bootstrap simulations can be performed like in figure 9b with the Mack-Net model to obtain a distribution of reserves.



(a) Link factors of Mack and Mack-Net

(b) Bootstrap procedure of Mack and Mack-Net

Figure 9: Comparison of the Mack Chain-Ladder and Mack-Net model features

In addition to comparing the Mack Chain-Ladder method with the Mack-Net model, it is also interesting, as part of a *benchmark* available in table 2, to confront these scores with other market models—some based on deep learning, such as the bCCNN model (Gabrielli et al., 2019) or *DeepTriangle* (Kuo, 2018), and others more traditional, such as the ECLRM method and the Munich Chain-Ladder model, which also utilize the incurred claims triangle like the Mack-Net model. In the Mack-Net model, incorporating this additional information degrades performance less than in Munich or ECLRM models, thanks to the flexibility of neural network structures. Ultimately, the combination of *Deep Learning* and claim cost incorporation leads to better performance.

Metrics	**Mack-Net**	**Mack Chain-Ladder**	**ECLRM**	**Munich Chain-Ladder**	**bCCNN**	***DeepTriangle***
AvE_{score}	504,7	501,4	753,7	490,0	722,3	1631
CDR_{score}	729,8	770,9	1381,1	1075,8	1296,2	1776
MAE^{1yr}	141,6	163,5	304,9	129,4	342,0	257
$MAPE^{\text{1yr}}$	1,6%	2,1%	3,9%	1,6%	4,7%	2,7%
MAE^{ult}	470,2	175,0	1070,9	181,0	1249,3	354
$MAPE^{\text{ult}}$	5,7%	2,3%	11,6%	1,5%	12,7%	1,7%

Table 2: Comparison of the average performance of the Mack-Net model (in thousands €)

The USP coefficients, estimated from the newly calculated $\hat{Q}_j$ in the Mack-Net model, can thus be compared in figure 10 with the other standard approaches. In the Mack-Net model, the model's hyperparameters minimize CDR_{score}, which enhances the reliability of the approach.

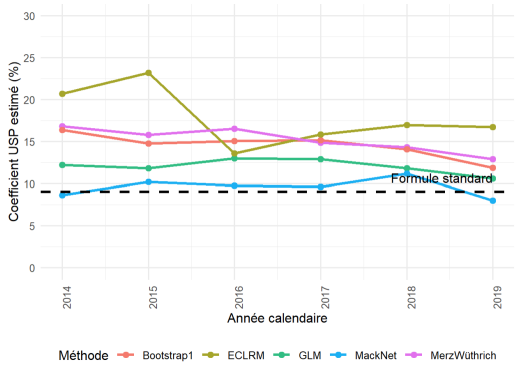


Figure 10: USP coefficients obtained with the new Mack-Net model

Conclusion

In the context of SCR calculation for the reserve risk of a captive, the present work highlights methods that can substitute the use of the parameter set by the delegated acts. In a context where more and more insurance or reinsurance captives are emerging on national territory, and where the Solvency II Directive is becoming less restrictive for them, the aim of this thesis is to explore new methods for estimating the USP parameter of the captive's reserves, notably through deep learning. It turns out that the parameter obtained with the Mack-Net model is more accurate than the regulatory approach, as the model reduces the error related to the calculation of the CDR metric, which is fundamental in the USP parameter computation. The challenge regarding data quality and hypothesis verification makes the supervisor's approval of the USP parameter more difficult in practice. Today, there exists a wide variety of methods on the market to quantify one-year reserve volatility, but the aim of this thesis is to study the contribution of *Deep Learning* methods, such as those using recurrent neural networks, to a regulatory topic.

Remerciements

Je tiens à remercier chaleureusement toutes les personnes qui m'ont permis de réaliser ce mémoire. En premier lieu, je tiens à exprimer ma gratitude envers mon encadrant de mémoire, Lucas Mietton, ainsi que Delphine Do'Huu et Laurent Bonnet, directeurs de l'équipe *Risk Analytics* de Marsh France, pour leur soutien tout au long du stage.

Je tiens à remercier également mon tuteur académique, Quentin Guibert, pour ses conseils avisés jusqu'à l'aboutissement final du projet, et plus globalement pour son dévouement intégral auprès des étudiants.

Merci à ma famille, mes parents, frères et soeurs, Quitterie et Patrice, mes amis, notamment Pierre et Benjamin, ainsi que mes camarades de promotion du master 2 Actuariat de l'université Paris-Dauphine pour les encouragements tout au long de l'année 2024.

Merci enfin à Eduardo Ramos-Pérez et Andrea Gabrielli d'avoir accepté d'échanger par mail sur les modèles de provisionnement explorés dans ce mémoire. Merci à Asaph Tapsoba. Enfin, merci à Arnaud Lacoume et Thomas Boyet du groupe Marsh McLennan pour l'expertise apportée lors du stage.

Table des matières

Résumé **3**

Abstract **4**

Note de Synthèse **5**

Synthesis note **11**

Remerciements **17**

Table des matières **19**

Introduction **21**

1 Les USP pour le risque de réserve d'une captive **22**
1.1 Qu'est-ce qu'une captive ? . . . 22
1.2 Les USP pour le risque de réserve . . . 29
1.3 Profil de risque de la captive . . . 40

2 Panorama et critiques des méthodes USP **58**
2.1 Critiques théoriques de la méthode réglementaire 1 . . . 58
2.2 Les alternatives à la méthode réglementaire 2 . . . 60
2.3 Application des méthodes . . . 69

3 Version *Deep Learning* de Mack pour le calcul de l'USP **77**
3.1 Le choix du modèle de provisionnement . . . 77
3.2 Utilisation des réseaux de neurones et réseaux de neurones récurrents (RNN) . . . 78
3.3 Le modèle Mack-Net . . . 91
3.4 Présentation des résultats . . . 97
3.5 *Benchmark* du modèle Mack-Net . . . 102
3.6 Limites des travaux . . . 105

Conclusion **107**

Bibliographie **108**

A Annexes **113**
A.1 Utilisation de l'IA générative . . . 113
A.2 Aboutissement à la formule de la méthode 1 des actes délégués . . . 113
A.3 Aboutissement à la formule de la méthode 2 des actes délégués . . . 115
A.4 Preuves du modèle linéaire généralisé (GLM) Poisson sur-dispersé . . . 116
A.5 Preuves de la méthode ECLRM . . . 121

A.6 Code associé à la méthode ECLRM 125
A.7 Algorithmes BFGS, L-BFGS et L-BFGS-B 127
A.8 Réserves estimées du chapitre 2 129
A.9 Modèle interne partiel avec approche bayésienne 129
A.10 Le modèle *DeepTriangle* de Kévin Kuo 132

Introduction

Depuis sa récente publication (Autorité de Contrôle Prudentiel et de résolution, 2024), l'ACPR effectue une revue de la Directive Solvabilité II, dressant les contours d'une refonte à l'horizon 2026. En particulier, l'un des points phares de cette révision porte sur le principe dit de proportionnalité, prônant « une simplification des exigences pour la partie la « moins complexe » du marché ». Dans le même temps, de plus en plus de groupes industriels décident de se lancer dans la création de leur propre entité d'assurance, les captives, notamment en France, à la suite d'un décret du 7 juin 2023 favorable à leur domiciliation sur le territoire national. Ces captives d'assurance ou de réassurance répondent au principe clé de proportionnalité. Avec l'instauration de la Directive Solvabilité II en 2016, les organismes d'assurance et de réassurance européens font face à des exigences renforcées en termes de capital réglementaire. Selon leur volume d'activité, ces organismes suivent, la plupart du temps, l'approche générale de la formule standard ou décident d'établir un modèle interne (partiel ou total, généralement coûteux financièrement et en temps de calcul). Néanmoins, pour certains modules de risques, certaines entreprises ont la possibilité d'utiliser des paramètres spécifiques, les USP (*Undertaking Specific Parameters*) qui permettent de mieux refléter leur profil de risque. L'objet de ce mémoire consiste à étudier la pertinence des méthodes USP imposées par les règlements délégués (EIOPA, 2015) pour le paramètre de volatilité des réserves.

De surcroît, ces dernières années, l'apprentissage automatique (*Machine Learning*) contribue pour beaucoup à l'élaboration de modèles actuariels robustes. En particulier, l'apport de l'apprentissage profond (*Deep Learning*) s'accélère en provisionnement non-vie, notamment avec Mulquiney (2006), Kuo (2018), Poon (2019), Gabrielli et al. (2019), Gabrielli (2019), Al-Mudafer et al. (2022). Les réseaux de neurones, très flexibles dans leur conception et leurs structures versatiles, ont déjà fait leurs preuves dans la production de résultats convaincants (Rossouw et Richman, 2019). Dans la mesure où de plus en plus de captives adhèrent au principe de proportionnalité, les règles imposées par la Directive Solvabilité II pourraient devenir moins contraignantes et ainsi les captives pourraient évaluer de manière plus indépendante leurs propres risques. Le cœur du mémoire s'intéresse à la précision d'un modèle de *Deep Learning*, le modèle Mack-Net (Ramos-Pérez et al., 2022), dans l'estimation du paramètre spécifique USP, par rapport aux méthodes concurrentes existantes. En basant l'étude sur des données responsabilité civile (RC) automobile d'une captive, le but du présent travail est d'explorer et de se demander si une méthode plus habile et précise permet d'apprécier la volatilité du provisionnement. En d'autres termes il s'agit de répondre à la question suivante : **peut-on adapter la mesure du risque de réserve, dans le cadre du calcul du paramètre USP de la captive ?**

Il convient dans un chapitre 1 d'introduire le contexte lié aux captives et à la justification d'emploi des paramètres USP. Dans le chapitre 2, il s'agit de dresser l'état de l'art des critiques actuelles autour des méthodes USP des actes délégués et d'étudier les méthodes concurrentes existantes. Enfin, le chapitre 3 s'articule autour du potentiel d'une récente méthode de provisionnement avec l'apport du *Deep Learning*, qui se révèle être une alternative crédible aux approches usuelles des calculs du paramètre USP pour le risque de réserve.

Chapitre 1

Les USP pour le risque de réserve, un choix logique pour une captive ?

1.1 Qu'est-ce qu'une captive ?

Il convient tout d'abord de rappeler le cadre qui permet aux captives d'assurance et de réassurance d'exister, l'environnement et les variétés qui évoluent de nos jours sur le marché. De plus amples informations sont détaillées dans la note de l'Institut des actuaires (Institut des Actuaires, 2021). D'après ce rapport, le terme « captive » émerge en 1955, désignant à l'époque l'entité d'assurance d'une entreprise minière. En Europe, la première captive s'établit au Royaume-Uni, dans l'industrie chimique. En France, le terme devient populaire lorsque les groupes Peugeot et Citroën créent leur propre organisme d'assurance interne.

D'après l'article 13 de la Directive Solvabilité II (Parlement Européen et Conseil de l'Union Européenne, 2009), une captive d'assurance ou de réassurance est une « entreprise d'assurance ou de réassurance détenue soit par une entreprise financière autre qu'une entreprise d'assurance ou de réassurance [...] soit par une entreprise non financière, dont l'objet est de fournir une couverture d'assurance exclusivement pour les risques de l'entreprise ou des entreprises auxquelles elle appartient ou d'une entreprise ou des entreprises du groupe dont elle fait partie ». Une captive est donc considérée comme un organisme d'assurance ou de réassurance détenu par un groupe industriel, commercial ou financier dans le but d'assurer ou réassurer exclusivement tous ou une partie des risques du groupe auquel elle appartient.

1.1.1 Les variétés de captives

Plusieurs classes de captives, dont la terminologie est résumée dans la table 1.1, se distinguent. Les captives pures incluent les captives monoparentales et les captives groupes. Les captives sponsorisées incluent les compartiments captifs.

Type de captive	Description
Captives pures	Détenues entièrement par les assurés
Captives sponsorisées	Non-obligatoirement détenues par les assurés et dont le panel de mutualisation ne se restreint pas au risque des assurés
Captives monoparentales	Concernent les captives exclusivement à usage d'une entreprise
Captives groupes	Détenues par plusieurs acteurs d'un même secteur industriel
Captives d'association	Peuvent être détenues par plusieurs acteurs de secteurs industriels distincts
Captives de location	Concernent les entreprises qui n'ont pas l'opportunité d'établir leur propre captive mais qui, en payant un ticket d'entrée, bénéficient d'une couverture
Compartiments captifs	Détenus par un porteur de risque agréé, elles sont composées de plusieurs cellules indépendantes. Ces cellules séparées ont parfois vocation à devenir elles-mêmes des captives pures.

TABLE 1.1 : Les différentes structures de captives (INSTITUT DES ACTUAIRES, 2021)

Pour ce qui est de leur fonctionnement interne, les captives font généralement appel à des sociétés de gestion, généralement des courtiers. La figure 1.1 présente le fonctionnement générique d'une captive d'assurance (directe) ou de réassurance.

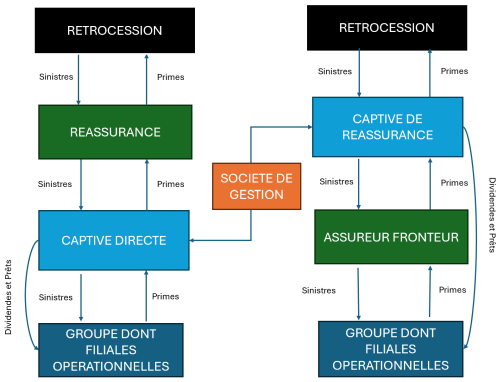


FIGURE 1.1 : Fonctionnement d'une captive d'assurance et de réassurance

La dernière catégorie de la table 1.1 peut parfois se révéler être un choix judicieux pour les sociétés ne détenant pas suffisamment de fonds propres et ne pouvant assumer des frais aussi élevés qu'une captive pure à part entière. Dans le présent travail, les données de sinistralité proviennent d'un groupe de location de véhicules transférant une partie de sa sinistralité à un compartiment captif de ce type. Comme sur la figure 1.2 et en relation avec la table 1.1, un ensemble de compartiments captifs s'apparente à un *pool* d'assurance ou de réassurance, où plusieurs cellules gravitent autour d'un seul propriétaire, nommé *Protected Cell Company* (PCC). Le PCC est l'entité à l'origine du montage financier, fournissant le capital et les frais initiaux nécessaires à sa constitution. Son autorité repose sur son pouvoir de distribution à chaque cellule de la licence d'assurance pour qu'elle puisse s'auto-assurer, et plus largement se réassurer. Aujourd'hui, de tels montages se révèlent être un choix de plus en plus prisé par les groupes qui souhaitent détenir leur captive. D'après le rapport annuel des captives de Marsh (MARSH CAPTIVE SOLUTIONS, 2023), un quart des nouvelles captives gérées par le courtier sont des cellules captives.

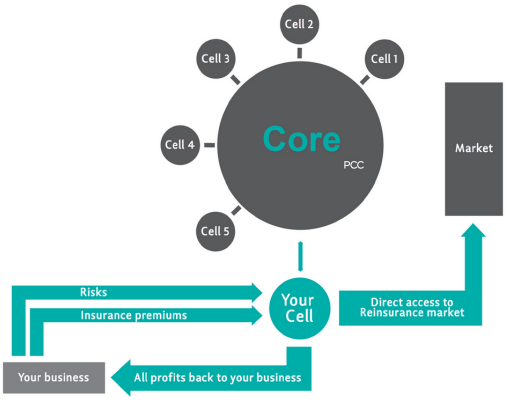


FIGURE 1.2 : Montage d'une cellule captive (COMPAGNIE NATIONALE DES SERVICES DE CONSEIL EN RISQUES & ASSURANCES, 2019)

1.1.2 Les enjeux pour une captive

Comme expliqué en section 1.1.1 et dans la figure 1.1, une captive repose sur le principe d'auto-assurance. La captive d'assurance peut émettre des polices et recevoir des primes de la part de la maison-mère du groupe auquel elle appartient. Elle peut aussi faire appel aux réassureurs pour transférer une partie de son risque. La captive de réassurance, quant à elle, ne souscrit pas sur le marché de l'assurance directe mais engage un assureur « fronteur » dont la responsabilité est l'émission des contrats et l'encaissement de la prime au nom de la captive, en échange de commissions de *fronting* (souvent à hauteur de 5-10% de la prime brute captive). Un mécanisme de rétrocession peut par la suite être mis en place si la captive de réassurance décide de céder du risque à son tour. La table 1.2 présente les risques et bénéfices potentiels à tirer à la suite d'une création de captive. On définit souvent l'appétit au risque d'une captive par un seuil de tolérance équivalent au montant maximum de sinistres acceptés. Ce seuil est un indicateur important pour la captive afin qu'elle puisse assumer les divers coûts de frottement (taxes, commissions etc.).

Opportunités	**Défis**
Maîtrise du financement des risques et conservation des primes encaissées (en cas de faible sinistralité) et des réserves détenues par le groupe au détriment de l'assureur, notamment les sinistres de fréquence	Expertise actuarielle requise pour justifier du niveau de primes encaissées par la captive ainsi que du niveau de sinistralité maximum en cas d'année défavorable selon l'appétit au risque
Couverture des risques plus adaptée car certains risques parfois non pris en charge par le marché assurantiel	Capacité à assumer divers frais (gestion, *fronting*, etc.)
Mutualisation des risques à l'échelle du groupe afin de peser sur les négociations commerciales	Détention suffisante de fonds propres réglementaires
Optimisation du pilotage de ses propres risques et meilleure gestion des programmes d'assurance	
Réduction de taxes	
Accès direct au marché de la réassurance	

TABLE 1.2 : Opportunités et défis suite à la création d'une captive

En France et au-delà, les captives s'établissent à peu près sur tous les continents comme indiqué sur la figure 1.3. Par exemple, en Europe, les captives se domicilient principalement à Dublin, Luxembourg, Malte, en Suède et en Suisse. Aux États-Unis, plus de la moitié des captives mondiales sont localisées, principalement dans le Vermont. Les îles sont aussi le lieu d'établissement de captives (environ un tiers), notamment aux Bermudes. Enfin, au Moyen-Orient et en Asie, on en trouve dans les Emirats et à Singapour (Souter, 2023).

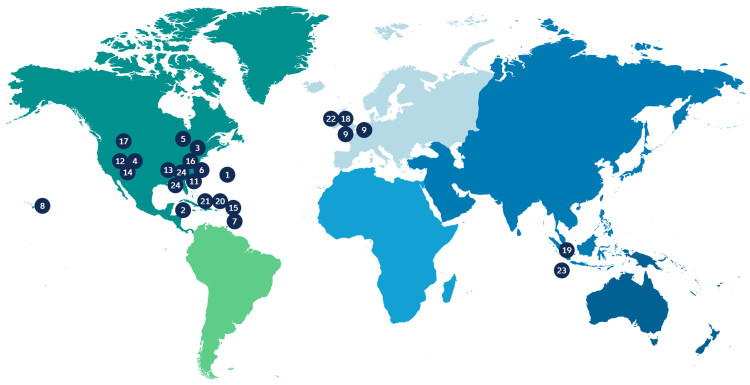


Figure 1.3 : Principales localisations des captives dans le monde en 2022 (Souter, 2023)

Au total, il existerait aujourd'hui plus de 6 000 captives dans le monde (dont plus de 200 en France), alors qu'elles étaient 1 000 en 1980. Au cours des dix dernières années, le nombre de captives a augmenté de 20% (Allianz, 2023). A titre d'exemple, pour un échantillon représentatif du marché et d'après le rapport Marsh (Marsh Captive Solutions, 2023), le courtier compte 370 nouvelles captives dans son portefeuille à l'échelle mondiale, dont 100 en 2020, 132 en 2021, et 138 en 2022, ce qui porte à 1 900 le nombre d'établissements captifs gérés par le courtier américain. Le montant de primes des 1 900 captives s'élève en 2023 à 70 milliards de dollars, et le surplus (correspondant à un excédent d'actifs par rapport aux passifs) s'élève à 118.5 milliards de dollars. Comme le souligne la figure 1.4, les entreprises décident de se lancer dans le projet d'une captive lorsque les tarifs commerciaux de l'assurance traditionnelle deviennent élevés.

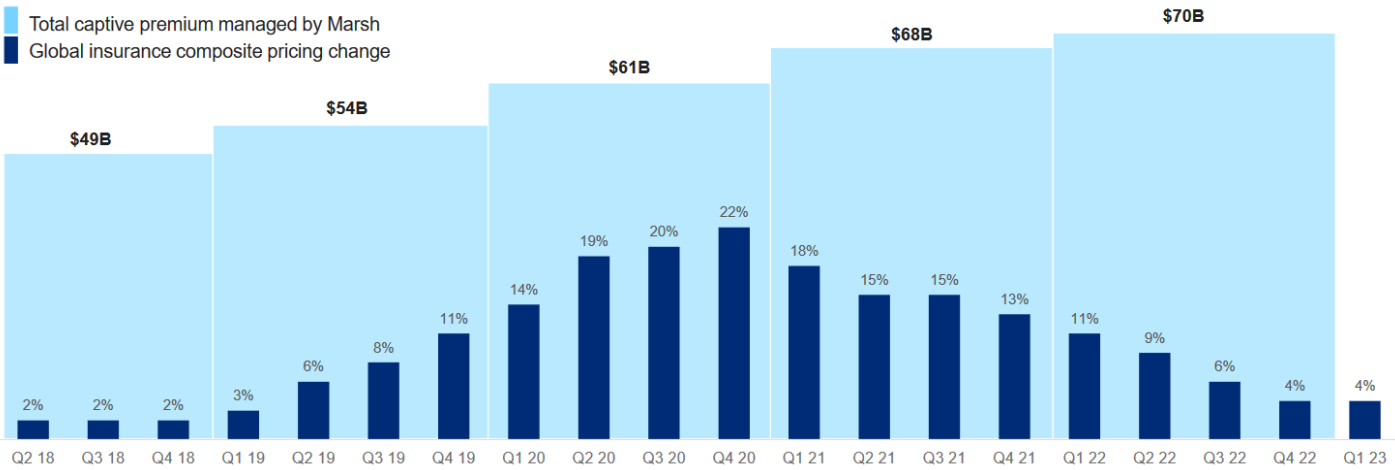


Figure 1.4 : Évolution de la prime des captives selon la variation des primes du marché mondial de l'assurance entre 2018 et 2023 des clients de Marsh (Marsh Captive Solutions, 2023)

Enfin, en termes de souscription, le podium des principales lignes d'activité des polices des captives

(gérées par le courtier) en figure 1.5 est composé de l'assurance non-vie (P&C, *Property and Casualty*, ou IARD, 36% de la prime), suivie par l'assurance vie (30%) et les contrats d'avantages sociaux (santé, retraite, 19%). Les lignes financières sont composées notamment de couvertures cyber, qui prennent de plus en plus d'ampleur ces dernières années sur le marché (le nombre de captives souscrivant ces contrats croit de 75% en 2022).

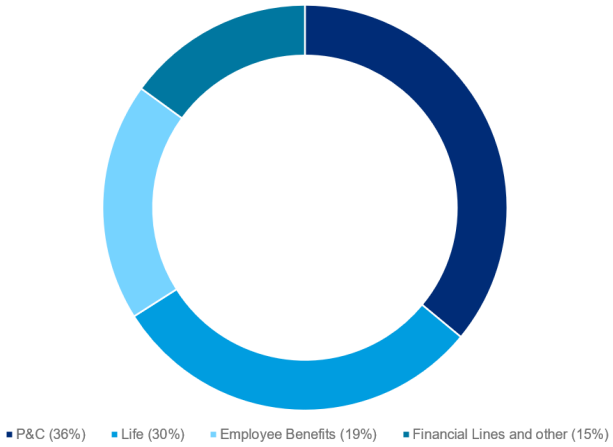


FIGURE 1.5 : Répartition de la prime captive selon les polices souscrites des clients de Marsh (MARSH CAPTIVE SOLUTIONS, 2023)

En termes de réglementation, les captives sont soumises à deux types de normes : la réglementation fiscale et celle des assurances. Longtemps perçues comme des véhicules d'optimisation fiscale, les captives sont soumises à différents contrôles. Au niveau européen, le projet BEPS (*Base Erosion and Profit Sharing*) lancé en 2014 par l'OCDE a pour objectif de repérer les organismes exploitant les failles entre les règles nationales et internationales. Les captives peuvent être soumises à ce contrôle afin de démontrer d'une pleine transparence. Par ailleurs, en tant qu'organisme d'assurance à part entière, les captives européennes sont soumises à la Directive Solvabilité II. Par exemple, pour le provisionnement des sinistres, les captives sont soumises aux mêmes règles de provisionnement, même si leur portefeuille est généralement plus restreint. En outre, pour lutter contre la volatilité de leurs réserves, des marges de sécurité sont souvent requises. C'est le cas par exemple de la Provision pour Fluctuation de Sinistralité (PFS) pour certaines captives de réassurance (basées au Luxembourg). Cette provision correspond à une provision pour égalisation globale spécifique.

En France, l'analogue de la PFS est la provision pour résilience (issue du décret du 7 juin 2023). Elle s'applique uniquement aux captives de réassurance issues d'entreprises non-financières pour certains risques IARD. Sa dotation annuelle est exemptée d'impôts pendant quinze ans (pas de limite au Luxembourg) et est plafonnée au maximum à 90% (100% au Luxembourg) des bénéfices techniques, par branche de risque, et à dix fois la moyenne des trois derniers MCR (*Minimum Capital Requirement*). Jusqu'à la fin 2022, la différence entre une domiciliation au Luxembourg et en France pour les captives de réassurance se base principalement sur la capacité à se doter de cette provision. Avant le décret du 7 juin 2023, les captives implantées en France ne bénéficient pas d'avantages fiscaux par rapport à une société commerciale, c'est-à-dire que leurs résultats sont taxés via l'Impôt sur les Sociétés (IS). Avec le nouveau décret, le Gouvernement, en faisant évoluer le Code Général des Impôts (Article 39 quinquies G), permet aux captives de réassurance en France de pouvoir lisser la charge des risques dans le temps. Ce processus permet aux captives de conserver les bons résultats d'une année sans application de l'IS, pour pouvoir les réutiliser en cas d'une année où la sinistralité excéderait les primes encaissées. Un autre avantage de la France pour l'établissement d'une captive par rapport au Luxembourg est la capacité à être exonéré d'impôt sur la fortune (établi sur l'actif net de la société). La table 1.3 expose

d'autres différences techniques entre les principaux domiciles de captives en Europe.

	France	**Luxembourg**	**Dublin**	**Malte**
Régulateur	ACPR	CAA	CBI	MFSA
Taux d'imposition	27.5%	24.94%	12.5%	35%
Financement	Prêts intragroupe	Prêts intragroupe, lettres de crédit (garanties bancaires)	Prêts intragroupe limités	Prêts intragroupe après accord
Provision pour égalisation	Exclu (avant juin 2023) sauf risques spécifiques (exemple : risque nucléaire)	Pour les captives de réassurance	Exclu	Exclu

TABLE 1.3 : Comparaison des principaux domiciles des captives en Europe

1.1.3 L'environnement Solvabilité II d'une captive

Tout comme les organismes classiques d'assurance européens, les captives sont soumises aux trois piliers de Solvabilité II. Elles doivent détenir une licence délivrée par l'autorité de contrôle nationale pour exercer leur activité. Leur structure de gouvernance doit être transparente et suffisamment développée.

Le principe de proportionnalité. L'un des principes général clé de droit dans la Directive est celui de proportionnalité (articles 56 et 88 des actes délégués (EIOPA, 2015), correspondant à un allègement des restrictions si l'activité et la taille de l'organisme d'assurance le permettent et s'appliquant à l'environnement des captives. Il prévoit que la Directive Solvabilité II s'applique de manière proportionnée selon la nature, la complexité, et l'échelle des risques de chaque organisme d'assurance car autrement le respect strict des règles pourrait leur être beaucoup trop défavorable. En 2020, une clause de revoyure Solvabilité II insistant sur ce principe est mise en place, aboutissant en 2024 à la publication de l'ACPR (AUTORITÉ DE CONTRÔLE PRUDENTIEL ET DE RÉSOLUTION, 2024) afin de le renforcer. Sans mettre en péril les engagements envers les assurés, le but est de réduire davantage voire de s'affranchir totalement du poids de la Directive sur les organismes les plus petits. À l'horizon 2026, la revue de la Directive (ACTUELIA, 2024) pourrait par conséquent exclure les plus « petits » organismes d'assurance de toutes exigences réglementaires si ces derniers détiennent moins de 15 millions € de primes brutes émises annuelles (contre 5 millions € aujourd'hui), et moins de 50 millions € de provisions techniques (contre 25 millions € aujourd'hui). Pour les organismes autorisés à appliquer ce principe (sur une durée renouvelable de deux ans), les critères d'éligibilité sont décrits dans la table 1.4. La table correspond aux critères d'un organisme d'assurance non-vie, ce qui est le cas pour la captive étudiée dans le présent travail.

Critère	Niveau
Ratio combiné moyen	inférieur à 100% sur les trois dernières années
Primes brutes acquises hors UE	inférieures à 20 M€ ou 10% des primes brutes acquises totales
Primes brutes acquises	inférieures à 100 M€
SCR marché et contrepartie	somme inférieure à 20% des investissements
Acceptation en réassurance	inférieure à 50% de l'encaissement annuel brut de primes émises

TABLE 1.4 : Critères d'éligibilité du principe de proportionnalité pour un organisme d'assurance non-vie (ACTUELIA, 2024)

Les conséquences en termes d'applications concrètes du principe de proportionnalité peuvent alors être :

- la réalisation de l'ORSA tous les deux ans à la place d'un rendu annuel,
- l'audit du bilan facultatif,
- la revue des politiques écrites, réalisée tous les cinq ans, et limitée à la gestion du risque, au contrôle interne, l'audit interne et aux rémunérations,
- la publication du rapport narratif RSR (*Reporting Supervisory Report* au sein du pilier 3) tous les cinq ans, au lieu de tous les trois ans.

Piliers Solvabilité II. Concernant le pilier 1 et les exigences quantitatives au niveau du bilan en vision économique, les captives n'ont souvent pas besoin de détenir d'actions ni d'obligations pour adosser leur passif, mais du capital prêté par le groupe aux filiales et à disposition de la captive suffit et est souvent transféré (*cash-pooling*). Cet accord entre la maison-mère et la filiale doit être pris en compte dans le calcul du SCR marché s'il peut être considéré comme un prêt, sinon dans le SCR de contrepartie s'il est évalué comme un dépôt à la banque, d'après un récent opinion (EIOPA, 2024). Concernant le capital minimum requis (MCR), il doit être au minimum de 2.7 millions € pour les captives d'assurance non-vie (sauf en cas de souscription de risque RC automobile, aviation, marine, crédit et caution auquel cas le niveau requis est de 4 millions €), de 4 millions € pour les captives d'assurance vie, et de 1.3 million € pour les captives de réassurance (article 129 de la Directive (PARLEMENT EUROPÉEN et CONSEIL DE L'UNION EUROPÉENNE, 2009)). Par ailleurs, la plupart des captives utilisent une version simplifiée de la formule standard, comme dans la figure 1.6. Enfin, en France, l'ACPR suggère (officieusement) un ratio de solvabilité minimum de 120% (VADJOUX, 2021).

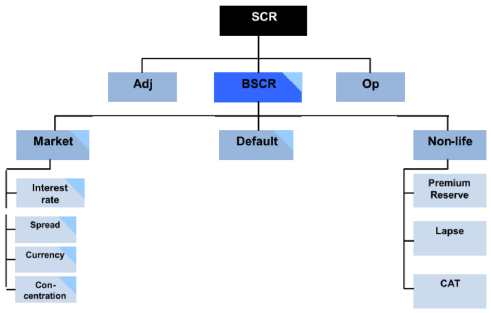


FIGURE 1.6 : Exemple de formule standard pour une captive d'assurance IARD

Concernant le pilier 2, la plupart des captives, en parallèle de leur structure de gouvernance, pratiquent l'externalisation que ce soit de leur système informatique, des calculs et rapports Solvabilité II, de la gestion des sinistres, des réserves, ou encore des fonctions clés (fonction actuarielle, gestion des risques, audit interne, conformité). Concernant le pilier 3, les captives mandatent aussi des sociétés de gestion (comme des sociétés de courtage) pour les travaux de *reporting*, que ce soit le SFCR (*Solvency and Financial Condition Report*), les QRT (*Quantitative Reporting Templates*) ou RSR. Au sein des trois piliers, les actes délégués (EIOPA, 2015) visent et exemptent explicitement les captives de certains aspects de la réglementation. La table 1.5 renseigne sur les articles concernés.

N° article	**Sujet**	**Adaptation aux captives**
88	Proportionnalité	« Évaluation de la nature, de l'ampleur et de la complexité des risques (modules ou sous-modules pertinents). » « Évaluation qualitative ou quantitative de l'erreur introduite dans les résultats du calcul simplifié. »
89	Dispositions générales pour les simplifications pour les entreprises captives	Simplifications des articles 90, 103, 105 et 106 autorisées si les engagements d'assurance ou de réassurance concernent uniquement les entités du groupe.
90	**Calcul du SCR primes et réserve en non-vie**	Formule simplifiée
103	Calcul simplifié de l'exigence de capital pour risque de taux d'intérêt	Formule des chocs simplifiée
105	Calcul simplifié de l'exigence de capital pour risque de *spread*	« Les captives peuvent fonder le calcul de l'exigence de capital pour risque de *spread* sur l'hypothèse selon laquelle tous les actifs sont affectés au troisième échelon de qualité de crédit. »
106	Calcul simplifié de l'exigence de capital pour risque de concentration du risque de marché	Les captives peuvent exclure les accords intragroupe de regroupement d'actifs de l'assiette de calcul si des « clauses contractuelles juridiquement contraignantes » existent. Le seuil d'exposition en excès est égal à 15% pour les expositions sur signature unique.

TABLE 1.5 : Exemptions pour les captives dans la réglementation Solvabilité II (EIOPA, 2015)

Comme l'indique l'article 90, le risque de **réserve en assurance non-vie** est un élément de différenciation pour les captives. Dans la section 1.2.3, il convient de définir ce risque et ses implications dans le cadre réglementaire.

1.2 Les USP pour le risque de réserve

Les USP (*Undertaking Specific Parameters*) sont des paramètres spécifiques propres à chaque organisme d'assurance pouvant remplacer (sur demande) certains coefficients utilisés dans la formule standard. En France, cette demande se matérialise sous la forme d'un dossier déposé auprès de l'ACPR composé d'une lettre formelle de demande, d'une note et des fichiers de calculs sur justification d'une telle utilisation par branche, d'un gage de qualité, de connaissances et d'un historique suffisant par rapport aux données (jugées « complètes, précises et appropriées »).

1.2.1 Les USP : des exigences importantes en termes de qualité des données

En vue de l'application des paramètres USP, la Directive impose comme pré-requis de détenir des données de qualité suffisante, notamment à travers ses articles 48, 82 et 86. Elle exige en effet que les organismes d'assurance suivent des mesures de gouvernance internes permettant de « garantir le caractère approprié, l'exhaustivité et l'exactitude des données utilisées dans le calcul de leurs provisions techniques ». Plus précisément, les critères d'évaluation à vérifier avant la modélisation et les calculs sont :

- la pertinence : les données doivent être adéquates pour les calculs et pertinentes par rapport aux risques du portefeuille,
- l'exhaustivité : les données sont suffisamment granulaires et volumineuses en termes de profondeur d'historique disponible pour effectuer les calculs,
- l'exactitude : les données sont propres, non-biaisées et constantes dans la durée.

Dans l'hypothèse où l'une des conditions serait non respectée sur les données, la société doit documenter (via la piste d'audit et le rapport de traçabilité des données) les limites et les ajustements effectués. La traçabilité correspond à un document relatif au traitement et l'utilisation des données, composée d'une cartographie des risques permettant de vérifier l'environnement des systèmes informatiques, des flux de données (de la source jusqu'aux comptes prudentiels) internes et externes permettant de transiter du répertoire de données aux contrôles de qualité. En supplément de la cartographie « macro », une cartographie « micro » (ou lignage) représentant les points de passage intermédiaires de la donnée jusqu'à son application finale peut être mise en place.

La qualité des données est un enjeu essentiel pour le bon fonctionnement d'une société d'assurance dans un objectif d'organisation, de maîtrise et de suivi sur les données qui sont transférées des systèmes de gestion vers les outils de modélisation. La Directive Solvabilité II impose la mise en place d'une gouvernance des données permettant d'évaluer et de mieux piloter la qualité des données (Actuelia, 2015). Comme décrit dans la figure 1.7 , la gouvernance des données s'inscrit dans un cycle de contrôle interne et de gestion des risques. Elle implique l'obligation d'instaurer des politiques écrites, dont une dédiée à la qualité des données. Des responsables ou propriétaires des données sont par la suite nommés comme maillons clés du processus, parmi eux :

- un responsable des systèmes d'information (SI), en charge de la cohérence des contrôles techniques,
- un coordinateur qualité, qui travaille et documente la qualité des données,
- un responsable pilier 1 en charge de la définition des données, des contrôles métiers, et de l'évaluation sur la qualité,
- la fonction actuarielle qui émet des avis et recommandations sur la qualité et la suffisance des données utilisées pour le calcul des provisions techniques,
- un responsable QDD (Qualité des Données) qui dispose d'une vision transversale des données Solvabilité II. Il est responsable du répertoire des données qui décrit les caractéristiques et les sources, des données. Il rassemble toutes les sources, internes ou externes à la société, ainsi que les données intermédiaires qui aboutissent aux données finales. Le répertoire des données contient les exigences de l'ACPR : description, localisation, source, usage, « criticité » (tests de sensibilité), propriétaire, modalités, et fréquences de mise à jour.

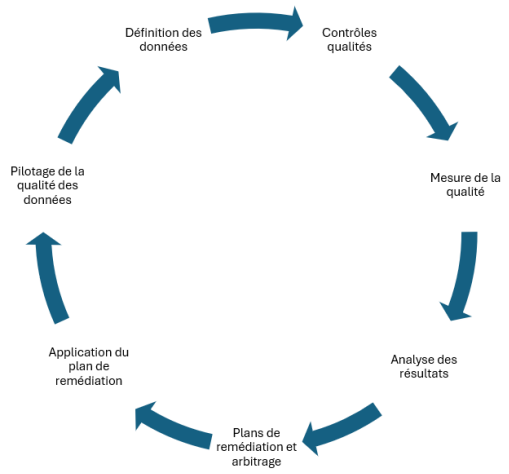


FIGURE 1.7 : Processus de pilotage de la QDD d'un organisme d'assurance

Des personnes relais peuvent aussi être nommées à chaque échelon des tâches opérationnelles (gestion des contrats, gestion des sinistres, gestion des actifs, comptabilité, informatique, etc.). Dans le cadre des contrôles internes opérationnels de premier niveau et second niveau (dont le but est l'audit de la qualité des contrôles de premiers niveau), une cartographie et une évaluation des risques de non-qualité des données sont mises en place afin d'améliorer la prise de décision et la synthèse des informations pertinentes pour la réalisation des missions.

L'établissement d'un dictionnaire de données constitue un premier moyen de gouvernance de la qualité des données. Ce dictionnaire est la base des exigences en matière de bilan prudentiel, provisions, SCR de marché, et comprend notamment la source ou la profondeur d'historique. Un dictionnaire permet de faire l'inventaire des données utilisées. Enfin, un plan de remédiation et un arbitrage doivent être mis en place définissant notamment la nature et le budget des remédiations, la durée et la date à laquelle l'intervention aboutit. Comme dispositif pour maîtriser le périmètre des données, les sociétés d'assurance se fient à la documentation relatives aux exigences sur la qualité des données (ACPR, 2023), en particulier :

- « l'entreprise conçoit et formalise un ou des documents (processus et procédures) organisant le dispositif de maîtrise de la qualité des données »,
- « le principe de proportionnalité n'exonère pas l'entreprise de se conformer aux exigences en termes de maîtrise de la qualité des données »,
- « les résultats obtenus dans l'ORSA sur la capacité de l'entreprise à faire face à certains événements adverses doivent être robustes, ce qui induit de les fonder sur des données de qualité qui sont, pour une bonne part, les mêmes que les données critiques utilisées pour établir les informations transmises à l'autorité de contrôle ».

L'ACPR recommande enfin d'instaurer un comité de gouvernance et de pilotage des données Solvabilité II dont les objectifs sont :

- le suivi des tableaux de bord des indicateurs de qualité,
- l'allocation du budget dédié à la maîtrise de la qualité des données,
- la mise en place des plans de remédiation de long terme,

- la mise en place de feuilles de route pour les équipes opérationnelles.

Ce comité est constitué *a minima* du directeur général, du responsable QDD, des chefs des fonctions informatique, actuarielle, et de la gestion des risques. Il décide notamment du niveau global de tolérance au risque en cas de non-qualité des données et d'une grille de seuils selon les actions menées pour les maîtriser (corrections, investigations plus approfondies etc.).

L'organisme d'assurance reste seul responsable de la traçabilité des données. Si les données proviennent de l'extérieur, un accord explicite écrit quant à la qualité des données doit toutefois exister entre le prestataire et l'organisme afin de clarifier les méthodes, les métriques, les fréquences d'évaluation du prestataire et la sûreté des données. Cet accord renseigne également sur les attentes (granularité, définitions des données, etc.), les modalités d'envoi et les contrôles de qualité effectués par le prestataire.

Ainsi, le risque de non-qualité des données est un risque connexe au risque de souscription et au risque opérationnel. L'organisme doit démontrer que les données sont représentatives du risque à un an d'horizon. Un rapport quantitatif détaillant les réflexions, les tests statistiques et de sensibilité doit être établit en conséquence.

1.2.2 Les USP : mesure entre formule standard et modèle interne partiel

Dans le cadre du pilier 2 de la Directive Solvabilité II, l'appréciation individuelle du risque est obligatoire. Elle est effectuée dans le cadre de l'ORSA. Une mesure de la volatilité du provisionnement tout à fait libre (autrement que par des méthodes standards comme la formule standard ou les USP) peut ainsi être réalisée. En effet, tout organisme d'assurance ou de réassurance doit intégrer à son activité une auto-évaluation de sa propre solvabilité et de ses risques à court et moyen termes, d'après l'article 45 de la Directive (Parlement Européen et Conseil de l'Union Européenne, 2009). Dans l'ORSA, le pilotage de l'organisme s'étudie non pas sur la prochaine année mais sur un horizon de 3 à 5 ans. L'ORSA est aussi un outil stratégique pour la gestion du capital, car il contient des projections des résultats des prochains exercices ainsi que du développement des produits d'assurance. Son but final est de quantifier les besoins en capital et de s'assurer de la cohérence des scénarii considérés dans le pilier 1. Cet ensemble d'informations fait l'objet d'un rapport soumis au régulateur, et est spécifique à chaque organisme d'assurance, selon sa spécificité et son appétence au risque. En résumé, l'ORSA correspond à une vision critique du pilier 1. Il permet de se demander si le calcul du pilier est suffisant et cohérent. Il fait partie de la stratégie de l'entreprise et le conseil d'administration y apporte une contribution active, en validant les scénarii et les résultats.

L'emploi de paramètres spécifiques USP peut être perçu comme une mesure entre la formule standard et un modèle interne partiel (MIP), c'est-à-dire un modèle interne appliqué localement (au module souscription, dans ce cas précis). Dans le cas d'une critique de la formule des USP dans l'ORSA, une société, si elle en fait la demande et obtient la validation de l'ACPR, peut recourir à un MIP, davantage sophistiqué. Déjà, dans les spécifications techniques du QIS5 (*Quantitative Impact Studies*) de l'EIOPA (Commission Européenne, 2010), le document fait référence à l'utilisation des USP : ≪ *Since none of the methods is considered to be perfect, undertakings should apply a variety of methods to estimate their volatility* ≫.

Le MIP correspond à une solution intermédiaire entre une formule standard et un modèle interne total (MI) où le périmètre d'application est limité, justifié et où les principes de la Directive doivent être respectés. L'article 122 des actes délégués (EIOPA, 2015) mentionne en effet que ≪ les entreprises d'assurance et de réassurance peuvent utiliser une période ou une mesure du risque différente [...] pour calculer le capital de solvabilité requis de manière à garantir aux preneurs et aux bénéficiaires un

niveau de protection équivalent ». L'utilisation d'un MIP, conditionnée à une exigence de qualité des données élevée, met en lumière la rigidité de la formule standard, comme l'illustre la table 1.6.

	Formule standard	**USP**	**MIP**
Intérêts	• Analysée à partir de scénarii • Plus simple à mettre en oeuvre	• Basée sur les données historiques • Plus adapté au profil de risque. Incitation par l'ACPR si le risque est mal pris en compte	• Souplesse sur l'approche méthodologique, sans la lourdeur d'un MI • Charge en capital requise davantage adaptée que les USP
Limites	• Ne prend pas en compte le profil de risque spécifique des entités	• Plus forte exigence sur la QDD • Forte demande pour la documentation	• Processus encore plus détaillé et coûteux

TABLE 1.6 : Comparaison des enjeux entre formule standard, USP et modèle interne partiel (MIP) (CERCHIARA et MAGATTI, 2013)

Pour que l'ACPR octroie l'utilisation d'un MIP, les hypothèses sous-jacentes de chaque méthode doivent être vérifiées. Si la société en fait la demande, un cycle de validation de MIP se met en place, où l'objectif est de confronter le modèle développé à l'expérience via un *backtesting*, d'analyser la stabilité du MIP en effectuant des tests de sensibilité et d'énoncer les hypothèses clés en spécifiant la littérature scientifique à l'appui. Pour qu'un MIP soit validé, une grille de validation est établie avec un calendrier préalablement défini et la désignation de responsables compétents et indépendants (*Fit & Proper*). Cependant, la mise en place d'un MIP peut être coûteuse, que ce soit en termes financiers ou en temps de calcul ; un budget alloué suffisant est donc nécessaire. L'utilisation d'un MIP à part entière reste aujourd'hui peu répandue sur le marché des captives.

1.2.3 Le risque de réserve en vision Solvabilité II

Rappels sur le provisionnement en assurance non-vie. Le risque de réserve correspond au risque d'un provisionnement des sinistres insuffisament estimé de la part d'un organisme d'assurance. Il reflète l'incertitude tout au long de la vie d'un sinistre, de son point de départ (sa date de survenance) à sa clotûre définitive, comme en figure 1.8. L'incertitude provient souvent d'un mauvais traitement des données ou d'une mauvaise évaluation, liée à l'erreur humaine ou à un écart statistique sous-jacent au modèle de provisionnement employé : on parle respectivement d'erreur d'estimation et d'aléa statistique. En général, au sein du provisionnement, on distingue les provisions pour sinistres suivantes (SURU, 2012) :

- une provision dossier/dossier : à la suite de la déclaration, cette provision correspond à une estimation moyenne du coût du sinistre au regard des éléments de contexte disponibles,
- une provision d'IBNER (*Incurred But Not Enough Reported*) : l'IBNER représente un ajout à la provision dossier/dossier à la suite d'une aggravation décalée d'un sinistre,

- une provision d'IBNYR (*Incurred But Not Yet Reported)* : certains sinistres peuvent être déclarés tardivement car, à la clotûre de l'exercice, tous les sinistres survenus peuvent être encore inconnus de l'assureur. L'IBNYR est un complément de la provision dossier/dossier et de l'IBNER.

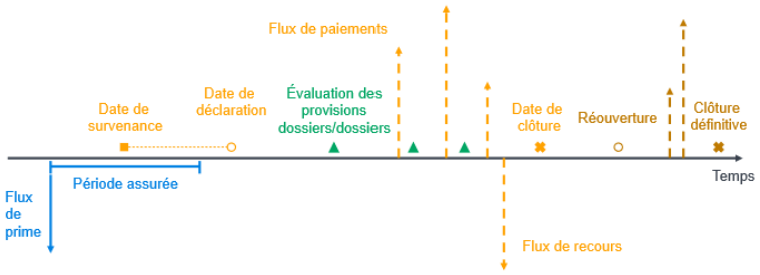


FIGURE 1.8 : Durée de vie d'un sinistre (ORTIZ, 2019)

La provision d'IBNR est alors définie comme IBNR = IBNER + IBNYR. Enfin, la PSAP (Provision pour Sinistres à Payer) s'exprime comme

$$\text{PSAP} = \text{Provision dossier/dossier} + \text{IBNER} + \text{IBNYR}.$$

En provisionnement, la tâche principale de l'actuaire est de prédire les *cashflow* futurs de la manière la plus juste possible. Dans l'optique de répondre aux exigences de solvabilité, il s'intéresse notamment à la vision à l'horizon un an.

La vision à un an du provisionnement. Les travaux de mémoires précédents permettent de mieux appréhender une métrique de la volatilité des réserves à horizon un an (LACOUME, 2009). Cette volatilité repose sur l'estimation du CDR, *Claims Development Result* (MERZ et WÜTHRICH, 2008) . Cette estimation, qui correspond à un élément du compte du résultat de l'organisme d'assurance, et qui traduit directement sa capacité à respecter les engagements envers les assurés, se base notamment sur le calcul d'une erreur quadratique de prédiction (NOUAR, 2015), la MSEP (*Mean Square Error Prediction*).

En considérant t et $t+1$ deux exercices comptables successifs et $C_{i,j}$ les montants cumulés en l'année d'accident $i \in \{0,\dots,I\}$ et l'année de développement $j \in \{0,\dots,J\}$ (la plupart du temps, $I = J$ et l'on a $t \leq I$), les notations usuellement employées sont

$D_I = \{C_{i,j} : i+j \leq I \text{ et } i \leq I\}$,

$D_{I+1} = \{C_{i,j} : i+j \leq I+1 \text{ et } i \leq I\}$,

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

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

Les paiements incrémentaux entre I et $I+1$ sont notés $Y_{i,I-i+1} = C_{i,I-i+1} - C_{i,I-i}$.

Le modèle de Mack Chain-Ladder (MACK, 1993), fondamental pour la suite de l'étude, suppose les hypothèses suivantes.

Hypothèse 1.2.1. *(a)* $\forall i \neq k, C_{i,j} \perp C_{k,j}$,

(b) Il existe $f_0,\dots,f_{I-1} > 0$ *tels que pour tout* i,j

$$\mathbb{E}(C_{i,j} \mid C_{i,0},\dots,C_{i,j-1}) = \mathbb{E}(C_{i,j} \mid C_{i,j-1}) = f_{j-1}C_{i,j-1},$$

(c) Pour tout i, j, $C_{i,j}$ est une chaîne de Markov telle que

$$\mathbb{V}(C_{i,j} \mid C_{i,j-1}) = \sigma_{j-1}^2 C_{i,j-1}.$$

Le modèle permet d'introduire les estimateurs sans biais

$$\begin{cases} \hat{C}_{i,J}^{I} = \hat{C}_{i,J} = C_{i,I-i}\hat{f}_{I-i}\dots\hat{f}_{J-1} \\ \hat{C}_{i,J}^{I+1} = C_{i,I-i+1}\hat{f}_{I-i+1}^{I+1}\dots\hat{f}_{J-1}^{I+1}, \end{cases}$$

et

$$\begin{cases} \hat{R}_i^I = \hat{C}_{i,J} - C_{i,I-i} \\ \hat{R}_i^{I+1} = \hat{C}_{i,J}^{I+1} - C_{i,I-i+1}, \end{cases}$$

où

$$\begin{cases} \hat{f}_j^I = \hat{f}_j = \dfrac{\sum_{i=0}^{I-j} C_{i,j+1}}{\sum_{i=0}^{I-j} C_{i,j}} \\ \hat{f}_j^{I+1} = \dfrac{\sum_{i=0}^{I-j+1} C_{i,j+1}}{\sum_{i=0}^{I-j+1} C_{i,j}}, \end{cases}$$

et pour $0 \leq j \leq J-1$

$$\begin{cases} \hat{\sigma}_j^2 = \dfrac{1}{J-j-1}\sum_{i=0}^{J-j-1} C_{i,j}\left(\dfrac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j\right)^2 \\ \hat{\sigma}_{J-1}^2 = \min\left(\hat{\sigma}_{J-2}^2, \hat{\sigma}_{J-3}^2, \dfrac{\hat{\sigma}_{J-2}^4}{\hat{\sigma}_{J-3}^2}\right). \end{cases}$$

Les notations suivantes sont également introduites

$$\begin{cases} S_j = \sum_{i=0}^{I-j-1} C_{i,j} \\ S'_j = \sum_{i=0}^{I-j} C_{i,j} \end{cases}.$$

La littérature définit usuellement la MSEP comme mesure d'incertitude du provisionnement. A l'ultime,

$$\mathrm{MSEP}(\hat{C}_{i,J}) = \mathbb{E}\left[(C_{i,J} - \hat{C}_{i,J})^2 \middle| D_I\right].$$

La volatilité du provisionnement se scinde en deux composantes : une volatilité de l'évaluation des flux futurs (l'erreur de processus) et une volatilité de l'estimation des paramètres (l'erreur d'estimation) estimées dans le modèle de provisionnement, d'où

$$\mathrm{MSEP}(\hat{C}_{i,J}) = \underbrace{\mathbb{V}(C_{i,J} \mid D_I)}_{\text{erreur de processus}} + \underbrace{\left(\mathbb{E}(C_{i,J} \mid D_I) - \hat{C}_{i,J}\right)^2}_{\text{erreur d'estimation}}. \tag{1.1}$$

On définit ensuite le CDR comme l'écart de ré-estimation entre deux provisions successives d'une année comptable à l'autre, illustré en figure 1.9. En effet, la Directive Solvabilité II, qui se base sur une projection à un an, requiert des organismes d'assurance de détenir suffisamment de réserves pour honorer leurs engagements sur cette période. Formellement, on définit alors le CDR avec les notations précédentes comme

$$\mathrm{CDR}_i(I+1) = R_i^I - R_i^{I+1} - Y_{i,I-i+1}. \tag{1.2}$$

La littérature distingue également souvent deux visions du CDR :

- une vision rétrospective, expliquée par le CDR réel défini par $\text{CDR}_i(I+1) = \mathbb{E}[C_{i,J} \mid D_I] - \mathbb{E}[C_{i,J} \mid D_{I+1}]$, correspondant au CDR en se plaçant « dans le futur » (cependant, cette quantité n'est empiriquement pas observable),
- une vision prospective expliquée par le CDR observable désigné par $\text{CDR}_i(I+1) = \hat{C}_{i,J}^{I} - \hat{C}_{i,J}^{I+1}$, correspondant à l'erreur d'estimation dans les provisions d'une année à l'autre par rapport à une déviation d'ultimes supposée théoriquement nulle entre les deux exercices.

Ainsi, la MSEP du CDR réel s'écrit

$$\text{MSEP}_{CDR_i(I+1)|D_I}(\widehat{CDR}_i(I+1)) = \mathbb{E}\left[\left(CDR_i(I+1) - \widehat{CDR}_i(I+1)\right)^2 \mid D_I\right],$$

et la MSEP du CDR observable s'écrit

$$\text{MSEP}_{\text{CDR}_i(I+1)}(0) = \mathbb{E}\left[\left(\widehat{\text{CDR}}_i(I+1) - 0\right)^2 \mid D_I\right], \tag{1.3}$$

pour une année d'accident i. Cette dernière quantité mesure l'aléa dans la solvabilité par rapport à l'absence de déviation (d'où le 0) dans les estimations à la fin de chaque année comptable. Elle s'apparente à une marge de solvabilité en cas de déviation du CDR.

La MSEP agrégée qui quantifie l'erreur de provisionnement à un an est alors

$$\text{MSEP}_{\sum_i \widehat{\text{CDR}}_i(I+1)|D_I}(0) = \sum_{i=1}^{I} \widehat{\text{MSEP}}_{CDR_i(I+1)}(0) + 2 \sum_{k>i>0} \hat{C}_{i,J}\hat{C}_{k,J} \left[\frac{\hat{\sigma}_{I-i}^2}{S_{I-i}\hat{f}_{I-i}^2} + \sum_{j=I-i+1}^{J-1} \frac{C_{I-j,j}}{S'_j} \frac{\hat{\sigma}_j^2}{S_j \hat{f}_j^2}\right]. \tag{1.4}$$

Pour passer des équations (1.3) à (1.4), les auteurs (Merz et Wüthrich, 2008) utilisent une série de notations et d'approximations intermédiaires qui sont rappelées dans l'annexe A.3 et qui permettent d'aboutir à la formule des USP des actes délégués.

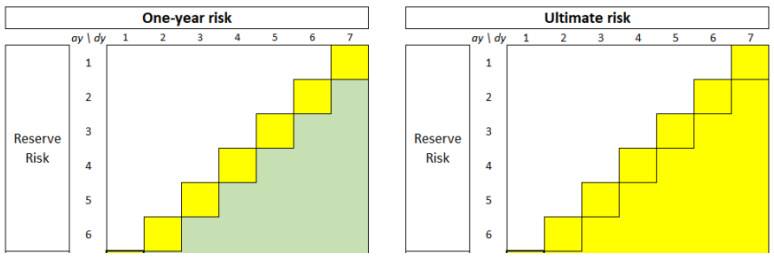


Figure 1.9 : Visions à 1 an et à l'ultime du provisionnement (Delong et Szatkowski, 2021)

1.2.4 Les deux méthodes USP des actes délégués associées au risque de réserve

L'application des USP repose sur l'analyse des données historiques. Les paramètres USP de volatilité calculés peuvent, pour les organismes qui en font la demande, aboutir sur un résultat plus faible ou plus élevé que ceux fixés par la formule standard, ce qui a comme conséquence une exigence en capital requise différente. Le périmètre global d'application des USP étant précisé dans la figure 1.10 (les modules ciblés sont colorés en orange), le présent travail se focalise sur celui du risque de réserve (sous-module prime et réserve du module souscription) en non-vie. Dans ce module, les paramètres qui peuvent être modifiés pour chaque ligne d'activité sont :

- l'écart-type du risque de primes en non-vie à un an mentionné à l'article 117 des actes délégués,
- l'écart-type du risque de primes à un an,
- le facteur d'ajustement pour la réassurance non-proportionnelle,
- l'écart-type du risque de réserve à un an en non-vie mentionné à l'article 117. En particulier, au sein du module, le SCR primes et réserves en non-vie s'exprime comme

$$\mathrm{SCR}_{\mathrm{nl\ prem\ res}} = 3 \cdot \sigma_{\mathrm{nl}} \cdot V_{\mathrm{nl}}, \tag{1.5}$$

avec

$$\begin{cases} \sigma_{\mathrm{nl}} = \frac{1}{V_{\mathrm{nl}}} \sqrt{\sum_{s,t} \mathrm{Corr}(s,t) \cdot \sigma_s \cdot V_s \cdot \sigma_t \cdot V_t} \\ \sigma_s = \sqrt{\frac{\sigma^2_{\mathrm{prem},s} \cdot V^2_{\mathrm{prem},s} + \sigma_{\mathrm{prem},s} \cdot V_{\mathrm{prem},s} \cdot \sigma_{\mathrm{res},s} \cdot V_{\mathrm{res},s} + \sigma^2_{\mathrm{res},s} \cdot V^2_{\mathrm{res},s}}{V_{\mathrm{prem},s} + V_{\mathrm{res},s}}} \end{cases},$$

où $V_i, i \in \{nl, prem, res\}$, $V_{j,s}$, $\sigma_{j,s}, j \in \{prem, res\}$ et $\mathrm{Corr}(s,t)$ représentent respectivement les volumes pour le risque de primes et/ou de réserves en non-vie, les volumes ou écarts-types pour le risque de primes et/ou de réserves et enfin le coefficient de corrélation pour le risque de primes et de réserves en non-vie entre les branches d'assurance s et t.

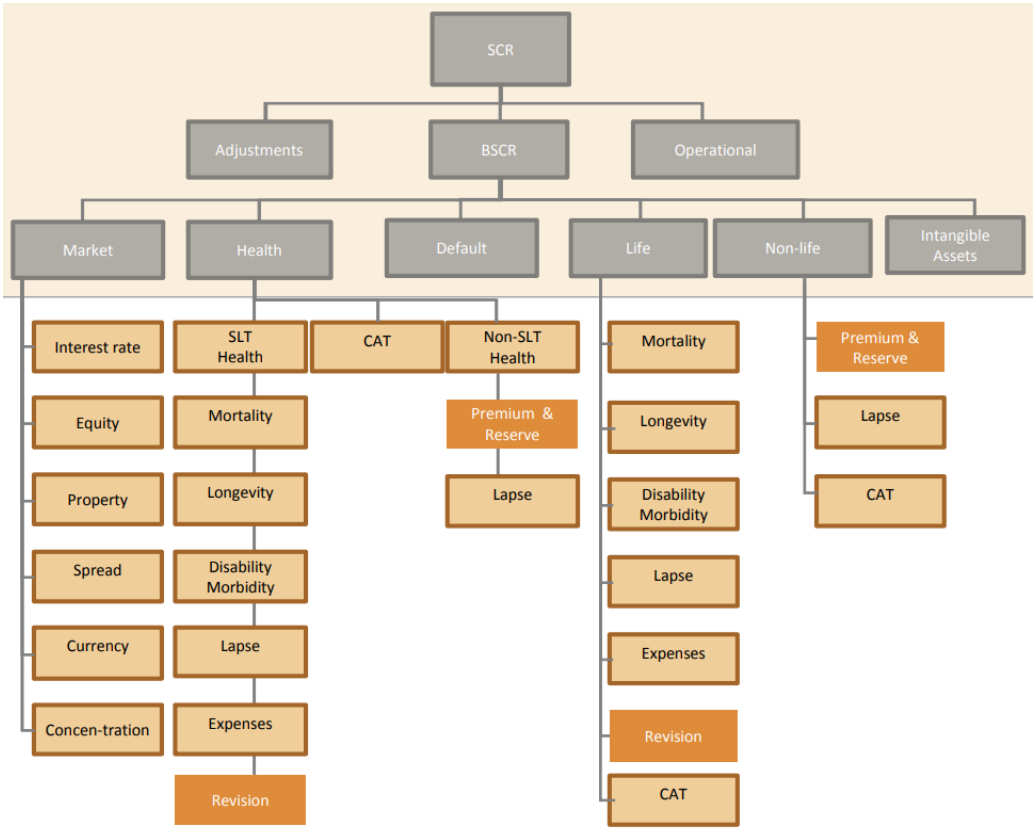


FIGURE 1.10 : Périmètre d'application des paramètres USP (LLOSA et al., 2015)

Ce dernier paramètre contient celui sur lequel le présent travail se base ; **plus précisément il s'agit du paramètre $\sigma_{\mathbf{res},s}$**. Ce paramètre doit refléter la volatilité implicite en cas de sous-provisionnement si un changement soudain dans l'évolution des règlements passés survient. Dans la formule standard, en RC automobile, l'écart-type est fixé à **9%** et précisé en annexe II des actes

délégués (EIOPA, 2015). Ce pourcentage diffère selon les LoB (*Lines of Business*). Il a eu l'occasion d'être remis en question à plusieurs reprises (EIOPA, 2011) notamment dans le cadre des QIS, par exemple QIS5 (Commission Européenne, 2010).

Il existe deux approches (Siegenthaler et al., 2017) par formules fermées pour calibrer la volatilité des réserves par LoB avec l'approche USP : la méthode paramétrique log-normale (« méthode 1 ») et la méthode des triangles (Merz et Wüthrich, 2008), dite « méthode 2 ». L'EIOPA impose d'appliquer les deux méthodes puis de sélectionner le coefficient le plus élevé. Cependant, d'après l'article 101 de la Directive, s'il est prouvé que l'une des méthodes produit des résultats trop erratiques, une seule méthode peut alors être appliquée.

Méthode 1 : la méthode log-normale. La méthodologie de cette approche est la même pour le calcul du paramètre USP du risque des primes. Soit y_t la « somme de la meilleure estimation de la provision établie à la fin de l'exercice pour les sinistres à payer en début d'exercice et des paiements effectués durant l'exercice » et la variable x_t la « meilleure estimation de la PSAP durant l'exercice comptable t ». Le modèle suppose les hypothèses suivantes.

Hypothèse 1.2.2. *(a) Il existe une relation linéaire entre l'espérance de y_t et x_t*

$$\mathbb{E}(y_t) = \beta x_t.$$

(b) Il existe une relation quadratique entre la variance de y_t et x_t, où $\delta \in [0;1]$ correspond au paramètre de mélange du modèle

$$\mathbb{V}(y_t) = \sigma^2((1-\delta)\bar{x}x_t + \delta x_t^2), \quad \bar{x} = \frac{1}{T}\sum_{t=1}^{T} x_t.$$

(c) La variable y_t suit une distribution log-normale, $\log(y_t) \sim \mathcal{N}(\mu, \omega)$ avec

$$\omega = \ln\{1 + \sigma^2[(1-\delta)\bar{x} + \delta x_t^2]\} \quad et \quad \mu = \ln(\beta x_t) - \frac{\omega}{2}.$$

(d) La méthode d'estimation par maximum de vraisemblance est appropriée.

Pour vérifier les hypothèses 1.2.2 *(a)* et *(b)*, des régressions peuvent être effectuées. L'hypothèse 1.2.2 *(c)* peut être vérifiée avec un test d'Anderson-Darling ou de Shapiro-Wilk. Enfin, l'hypothèse 1.2.2 *(d)* peut être vérifiée en observant sur une représentation 3D de la régularité de la fonction de volatilité. Dans le cadre des USP, l'écart-type du risque de réserve s'évalue comme

$$\sigma_{\text{res},s,\text{USP}} = c \cdot \hat{\sigma}(\delta,\gamma) \cdot \sqrt{\frac{T+1}{T-1}} + (1-c) \cdot \sigma_{\text{res},s}, \tag{1.6}$$

où $\hat{\sigma}$ est le paramètre spécifique estimé par l'entité, $\sigma_{res,s}$ le paramètre de marché (9% en RC automobile), c un facteur de crédibilité par LoB, précisé dans la table 1.7 variant selon la longueur de l'historique des sinistres, et T le nombre d'années comptables pour lesquelles les données sont disponibles. Ainsi, plus l'entité spécifique possède un historique profond, plus le poids de l'USP est élevé. Le paramètre à estimer $\hat{\sigma}(\delta,\gamma)$ en équation (1.6) correspond à

$$\hat{\sigma}(\delta,\gamma) = \exp\left(\gamma + \frac{\frac{1}{2}T + \sum_{t=1}^{T} \pi_t(\delta,\gamma) \cdot \ln\left(\frac{y_t}{x_t}\right)}{\sum_{t=1}^{T} \pi_t(\delta,\gamma)}\right), \tag{1.7}$$

avec

$$\pi_t(\delta,\gamma) = \frac{1}{\ln\left(1+(1-\delta)\cdot\frac{\bar{x}}{x_t}+\delta\right)\cdot e^{2\gamma}},$$

où $\gamma < 0$ correspond au paramètre de variation logarithmique du modèle. Les paramètres δ et γ sont obtenus par minimisation de la fonction suivante (Klugman et al., 2012)

$$l(\delta,\gamma) = \sum_{t=1}^{T}\pi_t(\delta,\gamma)\left(\ln\left(\frac{y_t}{x_t}\right)+\frac{1}{2\cdot\pi_t(\delta,\gamma)}+\gamma-\ln\left(\hat{\sigma}(\delta,\gamma)\right)\right)^2-\sum_{t=1}^{T}\ln\left(\pi_t(\delta,\gamma)\right). \tag{1.8}$$

La minimisation peut s'effectuer via des algorithmes de la famille Quasi-Newton.

Années d'historique	5	6	7	8	9	10	11	12	13	14	15 et plus
c	34%	43%	51%	59%	67%	74%	81%	87%	92%	96%	100%

Table 1.7 : Facteurs de crédibilité c de l'EIOPA en RC automobile en fonction du nombre d'année d'historique disponibles

Méthode 2 : la méthode des triangles de Merz & Wüthrich. La méthode est construite sur les triangles des paiements cumulés d'un historique de minimum cinq années. Elle repose sur le modèle de Merz & Wüthrich (Merz et Wüthrich, 2008), dont les hypothèses sont les hypothèses 1.2.1 de Mack ainsi qu'une quatrième hypothèse 1.2.3.

Hypothèse 1.2.3. *(a)*

$$\frac{\hat{\sigma}^2_{J-i}/(\hat{f}_{J-i})^2}{C_{i,J-i}} \ll 1.$$

L'écart-type USP du risque de réserve dans cette méthode est calculé comme

$$\sigma_{\text{res,s,USP}} = c\cdot\frac{\sqrt{\text{MSEP(CDR)}}}{\left(\sum_{i=0}^{I}(\hat{C}_{i,J}-C_{i,J-i})\right)}+(1-c)\cdot\sigma_{\text{res,s}},$$

avec

$$\text{MSEP(CDR)} = \sum_{i=1}^{I}\hat{C}^2_{i,J}\cdot\frac{\hat{Q}_{I-i}}{C_{i,I-i}}+\sum_{i=1}^{I}\sum_{k=1}^{I}\hat{C}_{i,J}\cdot\hat{C}_{k,J}\cdot\left(\frac{\hat{Q}_{I-i}}{S_{I-i}}+\sum_{j=I-i+1}^{J-1}\frac{C_{I-j,j}}{S'_j}\cdot\frac{\hat{Q}_j}{S_j}\right), \tag{1.9}$$

où $\hat{Q}_j = \frac{\hat{\sigma}^2_j}{\hat{f}^2_j}$.

Dans cette méthode, les provisions sont calculées à partir du modèle de Mack Chain-Ladder. Le paramètre $\sigma_{res,s}$ correspond toujours au paramètre de marché, et c au même facteur de crédibilité de la table 1.7 défini dans la méthode 1. Ainsi les méthodes USP 1 et 2 présentent l'avantage d'être des formules fermées, mais elles reposent sur un grand nombre d'hypothèses non toujours vérifiées empiriquement. Des alternatives à ces méthodes existent aujourd'hui et sont présentées dans le chapitre 2. Pour rappel, le but du présent travail est d'explorer les possibilités pour calculer les paramètres USP. Par exemple, le chapitre 3 revisite exclusivement l'approche de la méthode 2.

Cependant, pour l'heure, il convient de présenter en amont les données sous-jacentes utilisées pour les calculs et le contexte environnant. Comprendre les mécanismes implicites des données permet de

juger leur homogénéité et leur qualité qui est cruciale dans le cadre de la problématique du travail (comme expliqué en section 1.2.1). Il convient donc, dans la section 1.3 de décrire le contexte sous-jacent de la RC automobile en France dont sont issues les données, puis de commenter les données ligne à ligne (ou individuelles) des sinistres, utilisées exclusivement comme échantillon de contrôle de l'homogénéité de sinistralité avant de présenter les triangles de liquidation (agrégés), seules bases de calculs et de comparaisons des paramètres USP finaux obtenus dans le chapitre 2.

1.3 Profil de risque de la captive

Les données de sinistralité de la captive utilisées pour répondre à la problématique s'étendent de 1998 à 2019. Il convient d'abord en section 1.3.1 de dresser l'état des lieux de la situation du marché en France jusqu'en 2019 puis d'étudier les évolutions de marché de 2019 à aujourd'hui.

1.3.1 La RC automobile en France

Les données chiffrées du marché jusqu'en 2019. En 2019, le chiffre d'affaires en assurance automobile (RC et dommages aux véhicules) est de 22.8 milliards € pour les organismes d'assurance en France (Guy Carpenter, 2021), représentant 39% du total des primes d'assurance IARD. Les figures 1.12a et 1.12b indiquent que, depuis le début des années 2010, le volume de primes émises ainsi que le parc automobile assuré augmentent. De plus, le paysage assurantiel compte plus de cent organismes d'assurance. La figure 1.11 montre que les dix premiers groupes représentent plus de 85% des parts de marché de l'assurance automobile.

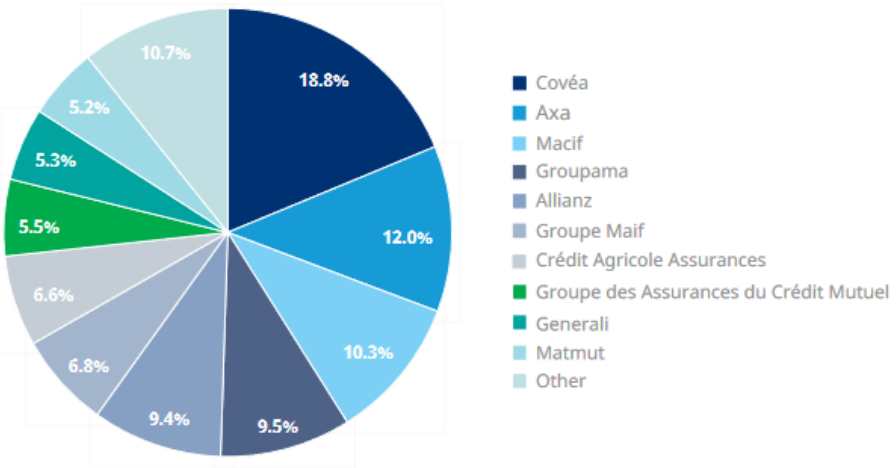
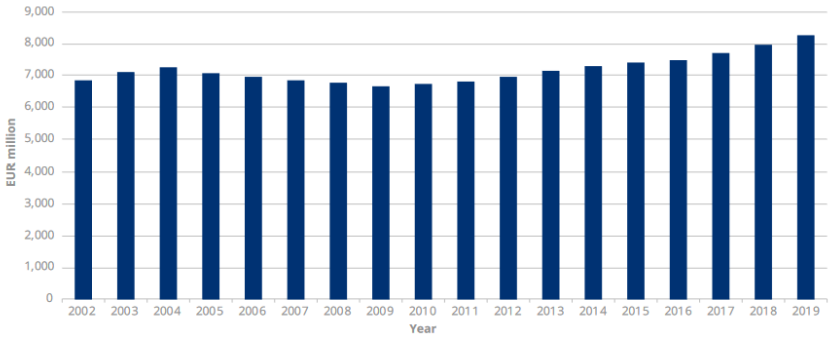
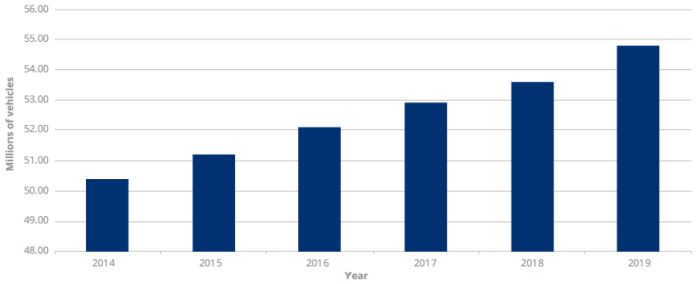


Figure 1.11 : Parts de marché de l'assurance automobile française en 2020 (Guy Carpenter, 2021)



(a) Évolution des primes émises en RC automobile



(b) Parc automobile assuré français

FIGURE 1.12 : L'assurance RC automobile en France (GUY CARPENTER, 2021)

En termes d'indicateurs du marché, les figures 1.13 et 1.14 permettent d'apprécier les ratios Sinistres/-Primes (S/P) et ratios combinés, indiquant l'évolution temporelle de rentabilité de la ligne d'activité.

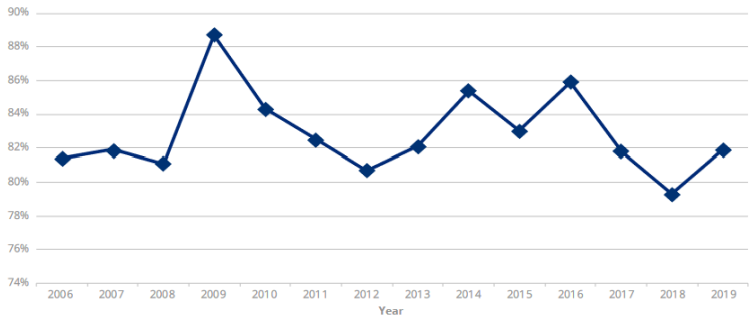


FIGURE 1.13 : Ratios S/P en RC et dommages automobiles (GUY CARPENTER, 2021)

Pour rappel, le ratio combiné correspond (en termes comptables) au ratio S/P additionné aux ratio d'acquisition (charges d'acquisition rapportées aux primes acquises) et d'administration (frais d'administration rapportés aux primes acquises). Hormis l'année 2018, les ratios combinés sont supérieurs à 100% alors que les ratios S/P fluctuent entre 80 et 90%.

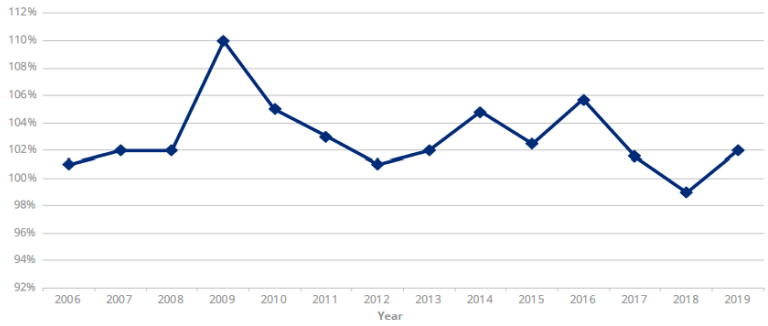
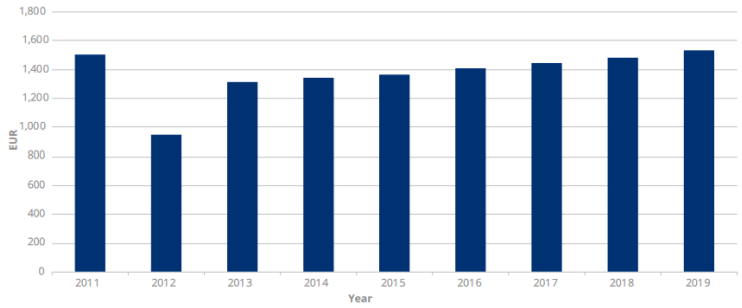
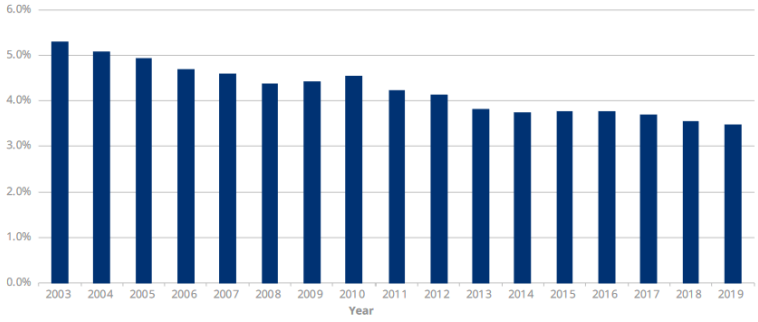


FIGURE 1.14 : Ratios combinés en en RC et dommages automobiles (GUY CARPENTER, 2021)

En termes de fréquence et de sévérité, sur la même période, les coûts moyens des sinistres augmentent alors que la fréquence diminue (d'après les figures 1.15a et 1.15b). Par exemple en 2019, le nombre d'accidents corporels diminue de 1.2% par rapport à l'année précédente et le coût moyen d'un sinistre (RC) augmente pour atteindre 1 530€. Ces évolutions peuvent s'expliquer par divers facteurs comme l'inflation qui augmente le coût des matériaux et pièces détachées ou par les campagnes de sensibilisation/prévention mises en place par l'état. Par exemple, le conseil national de la sécurité routière (CNSR), propose depuis 2001 des campagnes de prévention sur la sécurité routière, notamment auprès des jeunes où les taux de mortalité sont deux fois supérieurs à la moyenne.



(a) Coût moyen des sinistres



(b) Fréquence des sinistres

FIGURE 1.15 : Sinistralité en RC automobile (GUY CARPENTER, 2021)

En termes d'évolution des sinistres dans le temps, la RC automobile est réputée pour être une branche longue. A l'échelle du marché, les règlements de sinistres évoluent en moyenne sur plus d'une quinzaine d'années. La figure 1.16 indique l'évolution du marché des paiements jusqu'à leur ultime.

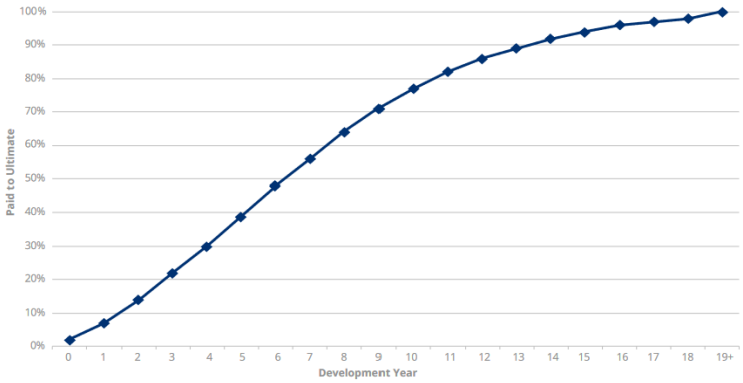
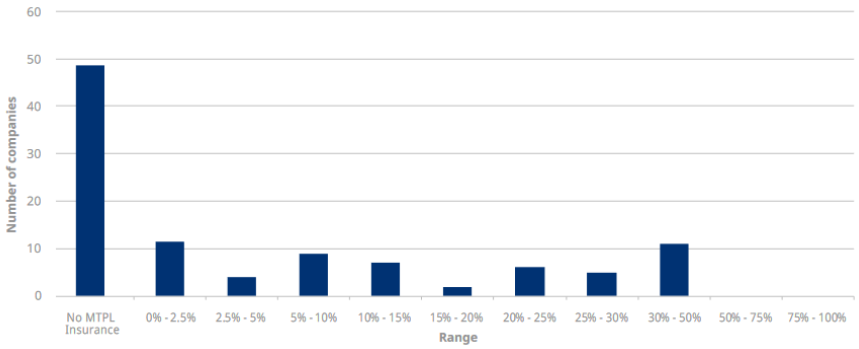
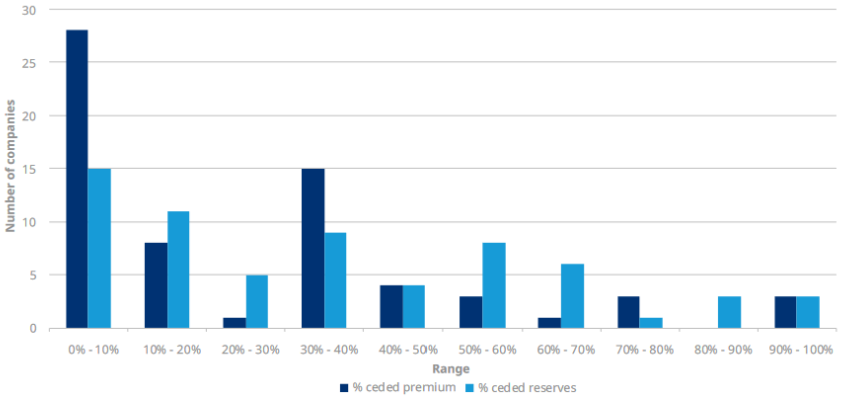


FIGURE 1.16 : Cadence de règlements des paiements en RC automobile (GUY CARPENTER, 2021)

En termes de solvabilité, la part du SCR primes et réserves en RC automobile représente plus de 25% du SCR total pour environ 25% des sociétés d'un échantillon représentatif du marché. De plus, la plupart des organismes d'assurance cèdent moins de 30% de leurs réserves en RC automobile en réassurance, comme l'indiquent les figures 1.17a et 1.17b.



(a) Part du SCR Primes et Réserves sur le SCR Total



(b) Cessions de primes et de réserves à la réassurance

FIGURE 1.17 : Vision Solvabilité II des primes et réserves RC automobile (GUY CARPENTER, 2021)

Les mutations du marché de 2019 à aujourd'hui. Jusqu'en 2023, la situation du marché confirme les observations faites jusqu'en 2019 (France Assureurs, 2023). Le parc assuré ne cesse d'augmenter (+1.2% en moyenne entre 2019 et 2023) tout comme les primes RC (+3% en moyenne entre 2019 et 2023), même si la pandémie de Covid-19 ralentit quelque peu cette tendance (+1.8% en 2020 en termes de primes) . Le confinement imposé marque en outre une diminution très marquée du nombre d'accidents pour les véhicules de première catégorie, c'est-à-dire de moins de 3.5 tonnes (-25.7% en RC matérielle, -26.8% en RC corporelle) mais l'effet atypique de l'année 2020 s'atténue très vite avec une recrudescence naturelle de sinistralité l'année suivante comme l'indique la figure 1.18. Enfin, en termes de sévérité des sinistres, le coût moyen d'un sinistre en RC matérielle passe de 1 053 € en 2022 à 1 118 € (+6.2%) en 2023 alors qu'en RC corporelle il passe de 3 850€ à 5 334€ (+38.5%). Les ratios combinés restent dans la même fourchette de la figure 1.14 (98%-102%) sauf en 2020 où l'effet Covid est positif sur la rentabilité des organismes d'assurance (94%).

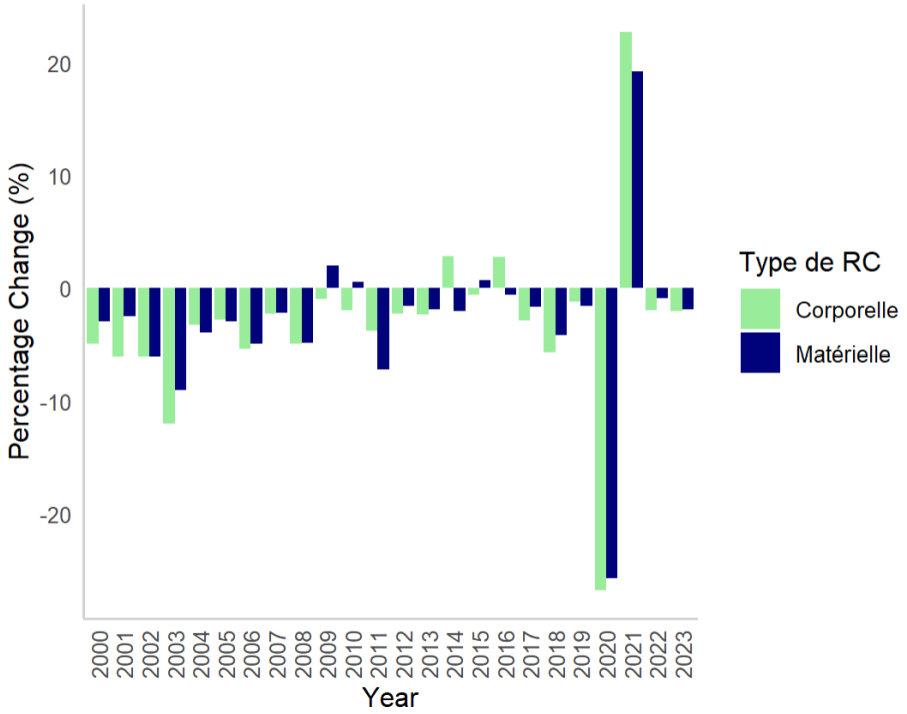


Figure 1.18 : Variation relative du nombre d'accidents des véhicules de première catégorie en RC matérielle et corporelle de 2000 à 2023 (France Assureurs, 2023)

Les spécificités françaises de la RC automobile. Des études de marché menées par Swiss Re (Swiss Re, 2022) et Guy Carpenter (Guy Carpenter, 2021) permettent de comprendre que la RC automobile en France possède de nombreuses particularités. Elle concerne les dommages causés par le conducteur à bord de son véhicule, qu'ils soient matériels ou corporels sur des tiers, et garantit une indemnisation aux victimes d'accident (d'où son appellation - assurance au tiers). En France il existe trois types de couverture RC automobile : l'assurance au tiers, l'assurance au tiers « plus » (bris de glace, vols et incendies inclus) et l'assurance tous risques (dommages tous accidents inclus). Les exclusions de garantie sont la conduite sans permis, la conduite sur circuit automobile, les dommages intentionnels et le transport de matières dangereuses.

En France, la RC automobile est une assurance obligatoire depuis 1945. L'entièreté des véhicules mobiles motorisés doivent être couverts par une police d'assurance automobile, même si le véhicule n'est pas utilisé. Le BCT (Bureau Central de Tarification) est responsable de traiter les cas particuliers

où un organisme d'assurance refuserait de prendre en charge le risque pour un assuré, et d'imposer le cas échéant, un organisme approprié qui sera obligé de lui fournir une couverture. Le code des assurances spécifie que si l'assureur continue de refuser de porter le risque d'un assuré, eu égard à l'obligation du BCT, alors cet assureur est considéré non aligné avec sa fonction de conformité et s'expose donc à un retrait d'agrément de la part du régulateur.

Les conventions IRSA (Indemnisation et Recours entre Sociétés d'Assurance) et IRCA (Indemnisation et Recours Corporel Automobile) sont également des éléments singuliers en RC automobile qui permettent une prise en charge accélérée des sinistres. Le principe est le suivant : à la suite d'un accident ou d'un constat à l'amiable, l'expertise estime un montant des dégâts et un pourcentage de responsabilité et, en fonction de ces informations et à l'aide d'une grille d'évaluation, une indemnisation est proposée aux deux parties. La convention IRSA a pour objectif de protéger les dégâts matériels causés aux véhicules lors d'accidents. L'assuré est d'abord couvert par son assurance, qu'il soit fautif ou non, avant que cette dernière n'engage potentiellement une procédure de recours. L'IRSA présente un barème de treize sinistres automobiles types (comme un stationnement irrégulier, une collision, le non-respect de la signalisation etc.) avec une indemnisation associée. L'assureur engage alors, à hauteur du niveau de responsabilité (responsable/à moitié responsable/non-responsable) contre l'organisme d'assurance de l'autre partie prenante à l'accident, une procédure de recours. Deux types d'indemnisation sont alors envisageables :

- un recours réel, c'est-à-dire le remboursement total des dégâts, s'ils sont estimés supérieurs à 6 500€,
- un recours forfaitaire (forfait fixe de la table 1.8), proportionnel au niveau de responsabilité de l'assuré, s'ils sont estimés à moins de 6 500€. On parle aussi de demi-forfait lorsqu'un recours forfaitaire est divisé par deux, c'est-à-dire lorsque la responsabilité de l'assuré est de moitié.

Années	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**
IRSA	1204	1236	1242	1276	1308	1354	1420	1446	1482	1568€	1678	1706	1776

TABLE 1.8 : Chronique des forfaits de la convention IRSA (en €)

La convention IRCA a quant à elle pour objectif de protéger les victimes d'accidents de circulation entre deux véhicules (ou plus) en indemnisant les dommages corporels légers collatéraux. Elle découle des articles 4 et 5 de la loi Badinter (BADINTER, 1985). Si le taux d'AIPP (Atteinte à l'Intégrité Physique et Psychique) de la victime est inférieur à 5%, la convention s'applique. De même que pour la convention IRSA, il existe deux types de recours :

- un recours réel, si le taux d'AIPP est évalué entre 1% et 5%,
- un recours forfaitaire (forfait fixe de la table 1.9), si le taux d'AIPP est nul. Le forfait peut être partagé entre les deux parties en cas de responsabilité à 50%. Ces types de recours interviennent généralement dans 70% des accidents corporels (INDEX ASSURANCE, 2023).

Années	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**
IRCA	1480	1480	1480	1480	1480	1254	1200

TABLE 1.9 : Chronique des forfaits de la convention IRCA (en €)

La loi Badinter (BADINTER, 1985) établit qu'en cas d'accident de voitures, les victimes à indemniser sont toujours les piétons, les cyclistes et les passagers, sauf en cas d'infraction. La négligence des

conducteurs peut néanmoins amener à un refus d'indemnisation de la part de l'assureur (aléa moral). L'aléa moral consiste en la modification du comportement d'un individu se sachant assuré. En d'autres termes, sachant qu'il est couvert, le conducteur peut se permettre de prendre plus de risques ou de se comporter moins prudemment sur la route. Grâce à cette loi, il est estimé (Swiss Re, 2022) que 10% des accidents en dommages corporels sont susceptibles d'être réglés devant les juridictions, et que 90% sont réglés à l'amiable, ce qui accélère les versements, simplifie les recours et réduit les procédures juridiques. Trois groupes de victimes se distinguent dont :

- les victimes non-conducteurs : les piétons, cyclistes et passagers transportés sont toujours des victimes couvertes pour leurs dommages subis à moins que leur faute ne soit inexcusable. Par exemple, traverser à un autre endroit qu'un passage piéton n'est pas une faute inexcusable. Leur droit à être indemnisé pour cette catégorie de personnes est de 100%,
- les victimes « super-privilégiées » : si l'âge de la victime est inférieur à 16 ans, supérieur à 70 ans, ou si leur taux d'AIPP est supérieur à 80%, ces victimes seront totalement indemnisées, sauf en cas de blessures volontaires (tentative de suicide par exemple),
- le conducteur : sa situation est plus délicate, car la réparation du préjudice qu'il subit peut ne pas être (totalement) prise en charge.

En termes de limites d'indemnisation, il est indiqué que les limites d'assurance en RC automobile doivent être révisées tous les cinq ans par rapport à l'évolution de l'indice européen des prix à la consommation en dommage au véhicule (*property damage*, ou RC matérielle). Par exemple, fin 2019, la dernière modification datait de mai 2017 et la limite minimum légale passe de 1,12 million € à 1,22 million € (Guy Carpenter, 2021). Concernant les dommages corporels (*bodily injury*, ou RC corporelle), les limites d'indemnisation sont illimitées, donc il n'y a pas de mise à jour.

Enfin, le Fonds de Garantie des Assurances Obligatoires de dommages (FGAO) permet aux victimes d'accidents subissant des délits de fuite d'être totalement indemnisées. Les assurés contribuent à 2% à sa pérennité. De plus, avec le nombre de conducteurs ne s'assurant plus aujourd'hui qui augmente, il y aurait, par exemple en 2021, entre 500 000 et 1 million de personnes non assurées en France (Guy Carpenter, 2021). Ce constat est amplifié avec l'arrivée des nouveaux véhicules électriques (NIEV). Toujours d'après le courtier, plus de 10% des accidents en dommages corporels sont causés par ces non-assurés. Le suivi d'un sinistre corporel est le suivant : l'assureur du véhicule causant l'accident met en place, avec la victime un diagnostic médical (supervisé par le médecin et l'avocat de la victime) pour évaluer les dommages corporels. Ce diagnostic, détaillant les handicaps temporaires et permanents de la victime, représente l'assiette de calculs pour le paiement des dommages et intérêts à venir. Huit mois après l'accident (ou trois mois après le décès de la victime), l'assureur propose une indemnisation provisoire, avant que l'offre finale soit présentée cinq mois après que l'assureur ait été informé de la stabilisation de l'état de la victime. En cas de manquement par rapport aux délais ou aux niveaux d'indemnisation, la loi peut émettre des sanctions. En cas de décision de justice, le juge peut décider, indépendamment de l'expertise médicale ou statistique, d'un prolongement des indemnités à verser, sous forme de capital ou d'annuités. Ce processus explique donc les cadences de règlements longs observés sur le marché.

En France, (au moins) entre 2021 et 2022 (Swiss Re, 2022), les coûts d'indemnisation liés à des accidents corporels graves augmentent de 5% par an. Le principe prôné en France est celui de « réparation intégrale » défini par la cour de cassation comme le rétablissement de la situation avant l'accident. Pour une demande d'indemnisation de la part de la victime pour n'importe quel chef de préjudice lié à l'accident, les deux principaux facteurs sont le taux d'AIPP évalué par les médecins, et le dommage portant atteinte aux droits patrimoniaux (lié au patrimoine réduit à la suite des pertes subies) et extra-patrimoniaux (liés au préjudice) , évalué par les avocats et les gestionnaires sinistres.

Les principaux chefs de préjudice, comme le manque à gagner ou le préjudice moral sont indiqués dans la figure 1.19 .

Pour les sinistres corporels, le sinistre donne lieu à un avis de sinistre basé sur la nomenclature Dintilhac, instaurée depuis 2005. Les paiements de ce type de sinistres peuvent être effectués de manière périodique ou en capital. Pour les paiements sous forme de rentes, une fois l'état de la victime stable, les paramètres des annuités versées sont fixés, en fonction des chefs de préjudice. Si le paiement est périodique, les calculs sont effectués sur la base de tables de mortalité avec un taux d'actualisation unique. Pour les paiements des sinistres en capitalisation, le montant est communément calculable de deux manières : avec la méthode BCRIV (Barème de Capitalisation de Référence pour l'Indemnisation des Victimes) ou avec la Gazette du Palais (souvent utilisée par les juridictions). Ces deux méthodes se basent sur les tables les plus récentes de la population générale fournies par l'INSEE. Le dernier barème de capitalisation en date (Planchet et Leroy, 2022) donne la possibilité d'utiliser un taux d'actualisation de 0% ou -1%, davantage favorable aux victimes par rapport au barème de 2020 (0% ou 0,3%), ce qui a pour impact un provisionnement encore plus prudent de la part des organismes d'assurance.

	Direct Victim		Victim's relatives	
	Temporary damage (before stabilisation)	**Permanent damage (after stabilisation)**	**In case of death of the victim**	**In case of victim's survival**
Patrimonial Damage	▪ Actual medical costs (DSA) ▪ Various costs (FD) ▪ Actual loss of income (PGPA)	▪ Future medical costs (DSF) ▪ House accommodation costs (FLA) ▪ Vehicle accommodation costs (FVA) ▪ Third-party assistance (ATP) ▪ Future loss of income (PGPF) ▪ Professional impact (IP) ▪ School, university or training damage (PSU)	▪ Funeral costs (FO) ▪ Relatives' loss of income (PR) ▪ Relatives' various costs	▪ Relatives' loss of income (PR) ▪ Relatives' miscellaneous costs (FD)
Extra-patrimonial Damage	▪ Temporary functional deficiency (DFT) ▪ Pain and suffering (SE) ▪ Temporary aesthetic damage (PET)	▪ Permanent functional deficiency (DFP) ▪ Leisure activities damage (PA) ▪ Permanent aesthetic damage (PEP) ▪ Sexual damage (PS) ▪ Founding damage (PE) ▪ Exceptional permanent damage (PPE)	▪ Loss of consortium (PAC) ▪ Affection damage (PAF)	▪ Affection damage (PAF) ▪ Exceptional extra-patrimonial damage (PEX)
Evolutional Damage	Damage linked to evolutional pathologies			

Figure 1.19 : Les principaux chefs de préjudice à la suite d'un accident corporel (Swiss Re, 2022)

1.3.2 Données ligne à ligne et hétérogénéité de la sinistralité

Il s'agit désormais de se concentrer sur la présentation des données. Celles-ci correspondent à des sinistres automobiles bruts de réassurance et de franchise (*from ground up*) en RC matérielle et corporelle pour une filiale française d'un groupe de location de véhicules (détenant une cellule captive de réassurance). Les triangles agrégés de liquidation des paiements (*paid*) et des charges totales (*incurred*) pour les accidents entre 1998 à 2019 sont disponibles (seules ces données vont être utilisées). Un aperçu à la maille trimestrielle des données ligne à ligne est également disponible, uniquement pour les dernières années de développement (de 2019 à juin 2023), ce qui permet de dresser un aperçu des

sinistres actuels qui composent les triangles. Bien que les calculs du paramètre USP qui vont suivre dans le chapitre 2 et 3 sont basés sur les triangles agrégés de provisionnement, explorer les données individuelles peut permettre d'effectuer des contrôles d'homogénéité et fournir des pistes de réflexion pour l'analyse finale des résultats. La table 1.10 présente les principales variables qui composent la base de données ligne à ligne. Les montants sont en vision cumulée.

Nom de la variable	**Explication**
Case	Numéro de la police d'assurance liée à l'accident
Paid Total	Valeur payée du sinistre
Recoveries Total	Recours (valeurs négatives)
Reserve Total	Montant réservé par les gestionnaires sinistres
Incurred Total	Payés + réserves + recours
Accident Month ou Year	Date de survenance
Report Month ou Year	Date de déclaration à la captive
Quarter_diff	Nombre de trimestres écoulés depuis l'accident
Claim_Lifetime	Date de fermeture – Date d'accident
Report_Time	Date de déclaration - Date d'accident
Closed_Time	Date de fermeture - Date de déclaration
Loss Type	Dommage au véhicule (*Property Damage*) ou Dommage corporel (*Bodily Injury*)
Status	Ouvert ou fermé
Cause/Major Loss Description	Contexte de survenance du sinistre
Section	Voitures (*Cars*) ou vans (*Trucks*)

TABLE 1.10 : Description des variables de la base de données ligne à ligne

Prise en compte uniquement des sinistres de 1998 à 2019 ouverts au 30 juin 2023. Ces données recensent les accidents de voitures et de vans ouverts survenus entre les années 1998-2019 (représentant 16% du nombre total de sinistres de la base et 28% des charges encourues au 30 juin 2023). Cette base d'ouverts concerne 18 920 véhicules et compte 21 279 accidents. Elle comporte en grande majorité des sinistres dommages aux véhicules (98%) et quelques sinistres corporels (2%). Ces sinistres corporels représentent néanmoins beaucoup plus en termes de paiements cumulés (41%), d'*incurred* (23%), de réserves (16%) et de recours (9%). Les accidents de la base des ouverts concernent environ à 60% des voitures et à 40% des vans. Le plus gros sinistre de la base, toujours ouvert, qui a eu lieu en 2008 (dommage corporel) est d'environ 2 millions €. En moyenne, le paiement d'un sinistre est de 1 046€ et la provision dossier/dossier de 2 600 €. Les accidents sans recours sont beaucoup plus nombreux (95%) que ceux avec recours, ce qui signifie que l'assuré est, la majorité du temps, responsable.

Les sinistres encore ouverts au 30 juin 2023 s'étendent du 10 mars 2000 au 27 juin 2023, la plupart des causes sont décrites dans la table 1.11 par les gestionnaires sinistres. Les causes (variables *Cause*) les plus récurrentes d'accident sont le stationnement irrégulier, la perte de contrôle du véhicule, la perte d'objets sur la route, le changement de voie ou la circulation en sens opposé.

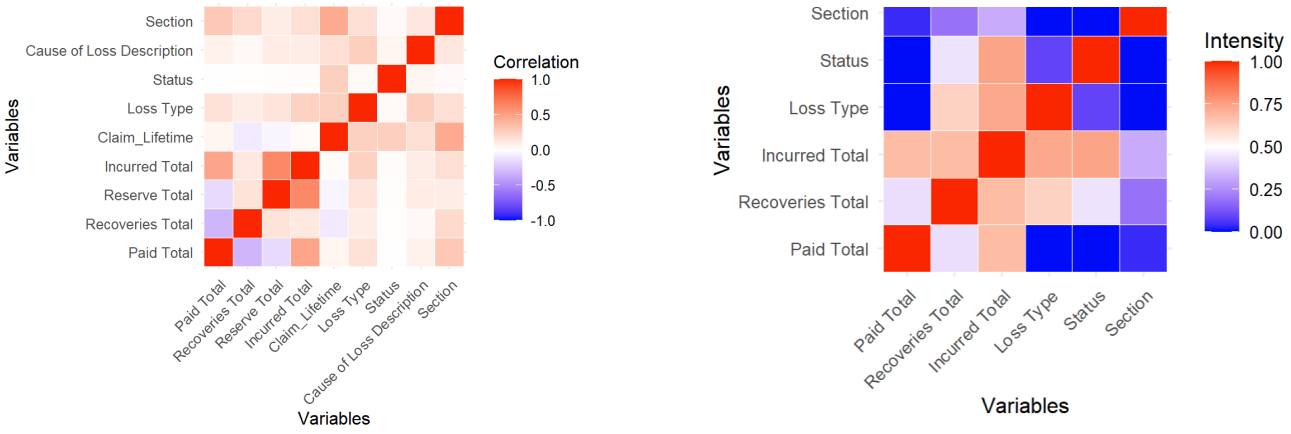
Type de sinistre	Nombre	Fréquence (%)
Un tiers a embouti l'arrière du véhicule de l'assuré	3518	16,5%
Assuré a heurté un tiers stationné	3445	16,2%
Collision hors intersection	3267	15,3%
Un tiers a heurté un assuré stationné	3154	14,8%
Inconnu	2537	11,9%
Collision à l'intersection	2144	10,1%
Accident impliquant un seul véhicule	1300	6,1%
Accident impliquant plusieurs véhicules	828	3,9%
Collision frontale	490	2,4%
Autres valeurs (22)	596	2,8%

TABLE 1.11 : Description par la gestion des sinistres des accidents RC automobile ouverts

La figure 1.20 présente les corrélations existant entre les différentes variables. Cette figure semble indiquer que les montants de sinistres interagissent avec le type de sinistres (matériels ou corporels), la cause du sinistre, ou encore le type de véhicule. Le V de Cramér (CRAMÉR, 1946) est une mesure d'association symétrique entre les variables basées sur une statistique du Chi-2 de Pearson. Il prend ses valeurs entre 0 et 1. Une valeur à 0 signifie une variable indépendante de toute autre et une valeur à 1 signifie que la variable peut être complètement déterminée par une autre. Mathématiquement, la statistique s'écrit

$$V = \sqrt{\frac{\chi^2/n}{\min(k-1, r-1)}},$$

avec $\chi^2 = \sum_{i,j} \frac{\left(n_{ij} - \frac{n_i \cdot n_j}{n}\right)^2}{\frac{n_i \cdot n_j}{n}}$, où $n_i = \sum_j n_{ij}$ est le nombre de fois que la première variable est observée, $n_j = \sum_i n_{ij}$ est le nombre de fois que la seconde variable est observée, pour i et j les modalités variant respectivement de 1 à k et de 1 à r, et n le nombre total d'observations.



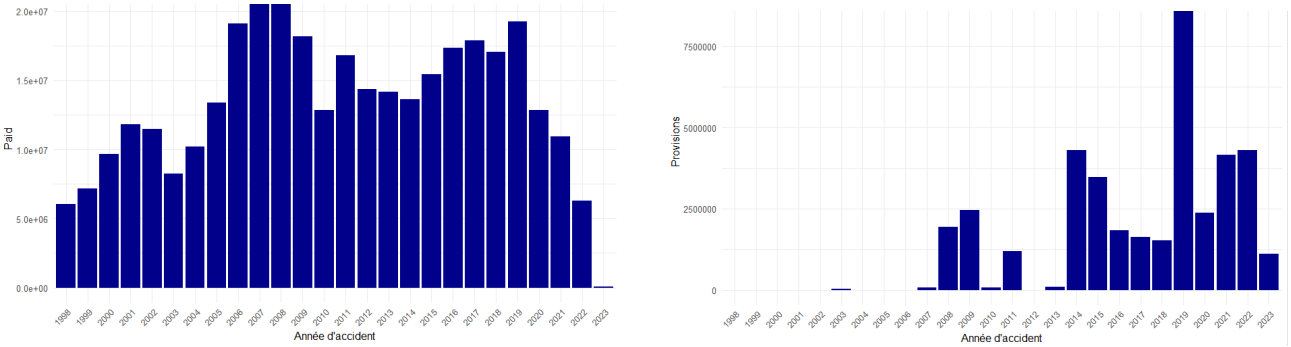
(a) Corrélations des variables liées aux sinistres

(b) V de Cramér des variables

FIGURE 1.20 : Corrélogramme et V de Cramér de la base sinistres

Prise en compte dans la base individuelle des sinistres ouverts et clos. Les figures 1.21 et 1.22a représentent les montants des sinistres payés, des charges, et des recours qui composent la base

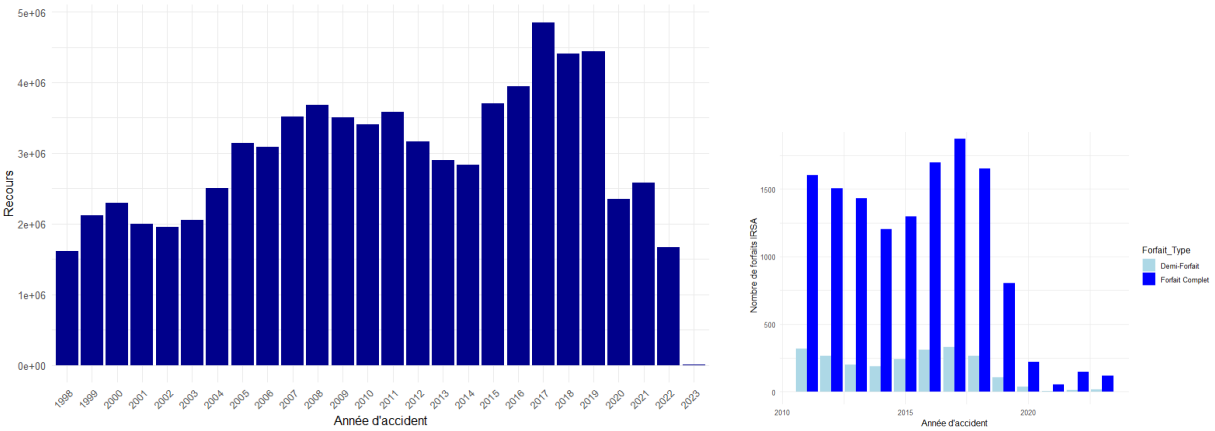
totale des ouverts et fermés (133 784 accidents, 134 021 véhicules impactés). En termes de paiements de sinistres, les années 2007 et 2008 sont les années les plus coûteuses pour la cellule captive.



(a) Montants payés (b) Provisions dossier/dossier

FIGURE 1.21 : Montants payés et provisions dossier/dossier par année d'accident (1998-2023) au 30 juin 2023

En termes de montants de recours dans la figure 1.22, l'année 2017 représente l'année la plus importante. En effet, en se référant aux montants de la table 1.8, l'année 2017 se démarque comme l'année la plus importante en termes de forfaits ou demi-forfaits dans le cadre de la convention IRSA (figure 1.22b).



(a) Recours (en valeur absolue) (b) (Demi-)Forfaits IRSA

FIGURE 1.22 : Recours et forfaits IRSA par année d'accident au 30 juin 2023

Les variables *Closed_Time* et *Claim_Lifetime* peuvent aussi être représentées dans la figure 1.23 pour les données des sinistres clos (puisque la date de clôture n'existe pas pour les sinistres ouverts). La variable *Closed_Time* correspond au nombre de mois entre la date de déclaration du sinistre et sa date de fermeture. La variable *Claim_Lifetime* correspond au nombre de mois entre la date d'accident et la date de fermeture. Pour ces deux variables, la médiane avoisine dix mois. Cela peut s'expliquer par le nombre de forfaits qui composent la base et la rapidité du processus d'indemnisation d'un sinistre une fois déclaré. Les délais peuvent aller jusqu'à 16 ans, ce qui correspond très probablement à des sinistres corporels.

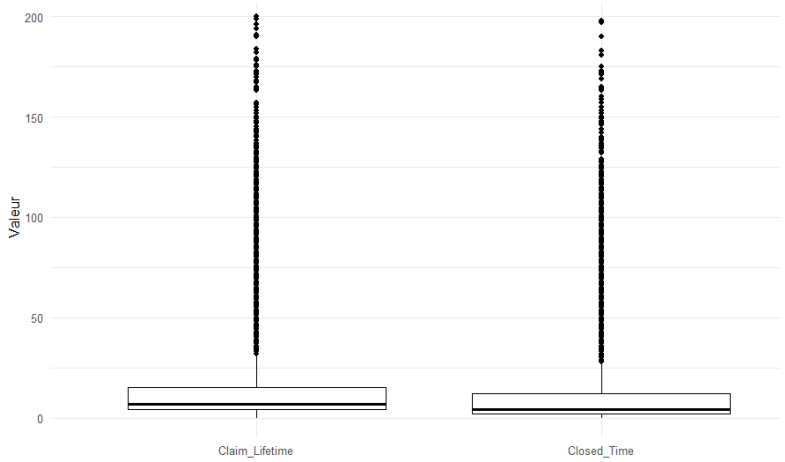


FIGURE 1.23 : Boxplot des variables de délais

En segmentant les sinistres selon différents groupes de caractéristiques à partir des données ligne à ligne (pour les années de développement 2019-2023), les cadences de règlements sont ainsi différentes, comme en témoigne la figure 1.24. Cette hétérogénéité se remarque notamment dans les différences de cadencement entre les dommages corporels (accident avec un piéton ou un cycliste par exemple) beaucoup plus longs et matériels plus courts en termes d'indemnisation.

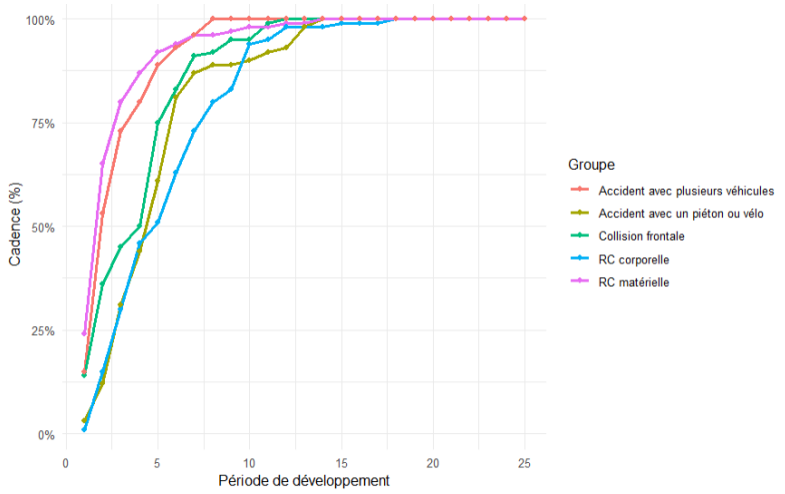


FIGURE 1.24 : Cadences de développement des sinistres à partir de la base individuelle

En conclusion de cette analyse, les données ligne à ligne permettent de comprendre un peu mieux la composition des triangles agrégés utilisés pour les calculs du paramètre USP qui vont suivre. Les résultats permettent de comprendre l'hétérogénéité des données selon les caractéristiques des accidents (notamment la distinction entre RC matérielle et corporelle). Cette hétérogénéité **est une limite au présent travail des calculs du paramètre USP puisque qu'aucune segmentation des données ne peut être effectuée par la suite (le calcul s'effectue sur un unique triangle RC de paiements)**. Les données ligne à ligne ne sont pas utilisées pour le provisionnement par la suite mais bien les données décrites en section 1.3.3. En effet, ne disposant pas de l'entièreté des dates de réévaluation antérieures à 2019 dans la base ligne à ligne, il est impossible de deviner l'évolution de sinistralité pour les années de développement antérieures. De plus, l'absence d'autres covariables pertinentes qui composent la base ligne à ligne, comme l'âge du conducteur ou l'ancienneté de permis (souvent utilisées dans les modèles récents de provisionnement individuel), est une seconde limite à

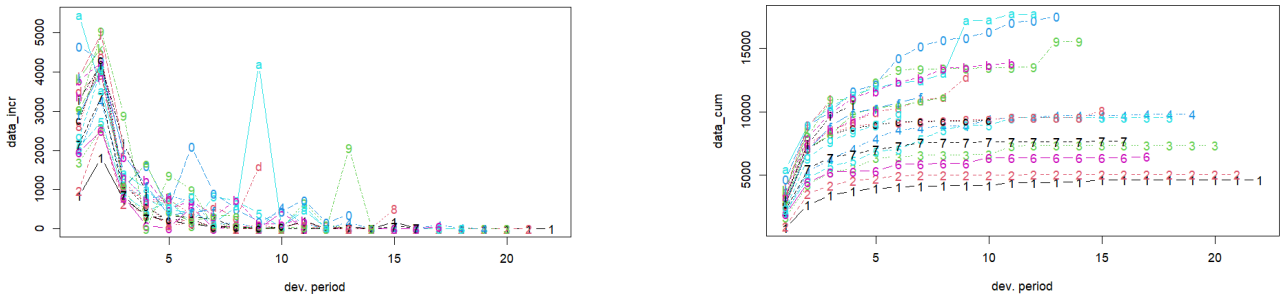
l'exploitation de ces données.

1.3.3 Présentation du triangle de liquidation des paiements

L'objectif du présent travail est, à partir du triangle de liquidation des paiements (issu d'une autre base que les données présentées en section 1.3.2) de mesurer la volatilité du provisionnement, pour les sinistres de 1998 à 2019. En effet, il n'a pas été souhaitable d'aller au-delà (jusqu'en 2023) en considérant la base des données ligne à ligne, pour **trois raisons principales**.

D'abord, des écarts importants de montants sont observés entre les deux bases entre la vision agrégée et ligne à ligne des données. Les hypothèses émises sur ces écarts sont notamment liées à l'indépendance entre les bases des données agrégées et individuelles (origine et traitement des données autonomes), la prise en compte dans les données individuelles des IBNR ou encore les écarts liés aux taux de change. Considérer les données individuelles aurait pu remettre en cause l'entièreté des valeurs du triangle des paiements pour les années de développement antérieures à 2019. En outre, les données individuelles n'étant accessibles que pour les quatre dernières années de développement, l'exploitation n'aurait pu aboutir au vu de la longueur de la branche RC en termes de cadence de règlements (figure 1.16). Enfin, le calcul des paramètres USP imposant de détenir un historique minimum de cinq ans (comme expliqué en section 1.2.4), les données ligne à ligne se révèlent donc être insuffisamment profondes pour répondre à la problématique.

Le *package* `ChainLadder` (R CORE TEAM, 2023) permet d'effectuer les principales analyses de triangles (GESMANN et al., 2023). Les sinistres RC automobile, comme le montrent les montants incrémentaux et cumulatifs, sont des sinistres à développement long. Le suivi par année d'accident permet aussi de repérer celles qui sont atypiques.



(a) Paiements incrémentaux

(b) Paiements cumulés

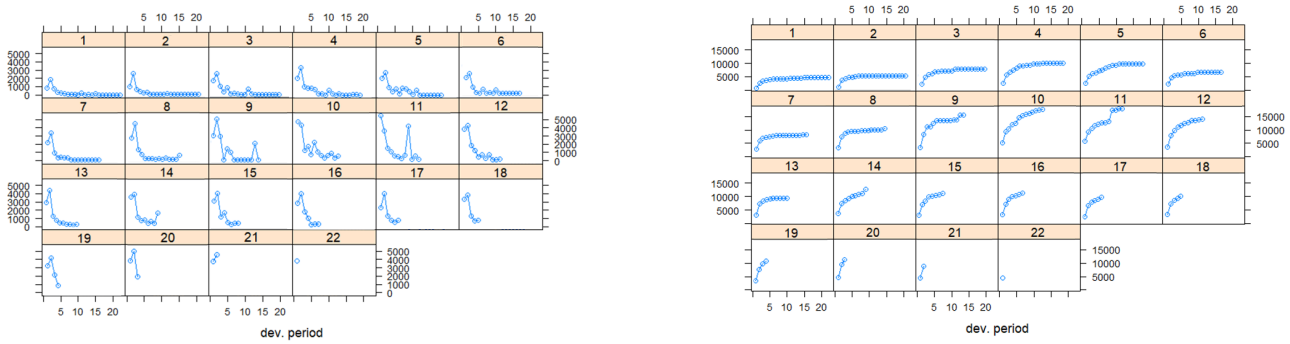
FIGURE 1.25 : Comparaison des développement de sinistres individuels par année d'accident

Sur la figure 1.25, plusieurs remarques peuvent être formulées telles que :

- de manière générale, les années d'accident les plus anciennes ont des montants de sinistres plutôt faibles par rapport aux années d'accident les plus récentes,
- des paiements incrémentaux importants, après cinq années de développement apparaissent, ce qui s'explique par la nature même de la branche d'assurance,
- les années 9, 0 et a (respectivement 2006, 2007 et 2008) sont des années atypiques en termes de développement, ce qui était déjà remarqué avec les données ligne à ligne. Un « plateau » pour les accidents de 2006 et 2008 se dessine en effet entre la huitième et la douzième année de

développement, probablement lié à une consolidation de l'état de santé de(s) victime(s) lors d'un ou plusieurs accidents corporels, avant une recrudescence, quelques années plus tard. Ces cas surviennent fréquemment en RC automobile.

La figure 1.26 indique également que même après dix années de développement, les sinistres survenus lors des années 9 à 11, ne sont pas encore totalement développés. Les sinistres semblent atteindre une cadence proche de 100% à partir de quinze années de développement (constat validé en figure 1.27).



(a) Vision incrémentale par année d'accident (b) Vision cumulée par année d'accident

FIGURE 1.26 : Développement des sinistres par année d'accident

Le rythme de cadencement moyen du triangle est précisé dans la figure 1.27. Le modèle sous-jacent à l'étude est le modèle de Chain-Ladder, dont le principe fondamental est, pour rappel, que les facteurs de développement sont constants, c'est-à-dire que $C_{i,j+1} \approx f_{i,j} C_{i,j}$ où les $f_{i,j}$ représentent les facteurs de développement individuels. La cadence individuelle de règlement correspond à $\rho_{i,j} = \frac{C_{i,j}}{C_{i,I}}$ et la cadence moyenne de développement en découle.

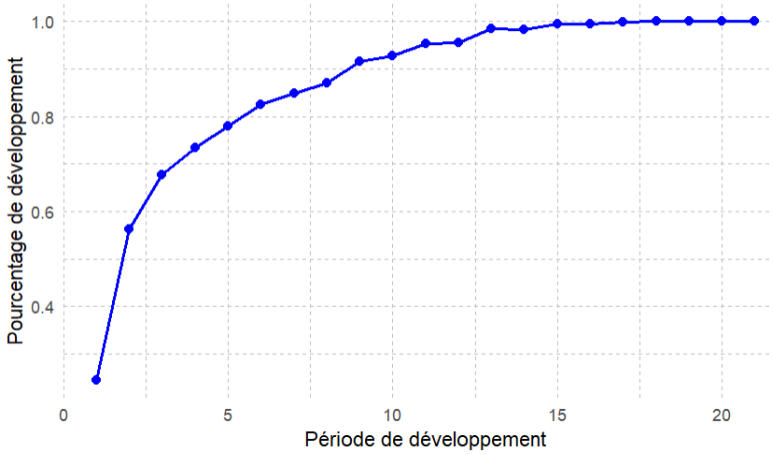


FIGURE 1.27 : Cadence moyenne des règlements des sinistres de 1998 à 2019

Vérification des hypothèses. Il s'agit de vérifier les hypothèses 1.2.1 du modèle de Chain-Ladder. En effet, dans le cadre du calcul du paramètre USP de la méthode 2, il est souhaité que ces prérequis soient vérifiés.

Hypothèse 1.2.1 *(a)*. Elle signifie que les années d'accident sont indépendantes. Il existe un test d'indépendance (MACK, 1993) qui est équivalent à l'absence d'effets calendaires comme l'inflation

(sociale ou économique) affectant les règlements de sinistres au fil des ans, les changements du mode de gestion des sinistres (etc.). Le cadre du test est le suivant

Soit la $j^{\text{ème}}$ diagonale du triangle $D_j = \{C_{j,0}, C_{j-1,1}, \ldots, C_{1,j-1}, C_{0,j}\}$, pour $0 \leq j \leq I$.

Soit $A_j = \{C_{j,1}/C_{j,0}, \ldots, C_{0,j+1}/C_{0,j}\}$. L'ensemble A_j est ordonné pour obtenir l'ensemble $F_k = \{C_{i,k+1}/C_{i,k} \mid 0 \leq i \leq I-k\}$, du facteur de développement le plus petit au plus élevé. On obtient alors respectivement les ensembles LF_k et SF_k correspondant aux facteurs de développement supérieurs (resp. inférieurs) à la médiane.

On définit alors $L = LF_0 + \ldots + LF_{I-2}$ et $S = SF_0 + \ldots + SF_{I-2}$ et respectivement L_j et S_j comme le nombre de composantes de L et S. Pour chaque diagonale A_j, $0 \leq j \leq I-1$, on compte le nombre de L_j et S_j.

Soit $z_j = \min(L_j, S_j)$. Si z_j est plus faible que $(L_j + S_j)/2$, alors cela signifie la présence d'effets calendaires. En supposant L_j comme une loi $\mathcal{B}in(n = L_j + S_j, 0.5)$, les moments de z_j sont

$$\begin{aligned}
\mathbb{E}(z_j) &= \frac{n}{2} - \frac{n-1}{n} \cdot \frac{n}{2^n}, \\
\mathbb{V}(z_j) &= \frac{n(n-1)}{4} - \left(\frac{n-1}{n}\right)^2 \cdot \frac{n(n-1)}{2^n} + \mathbb{E}(z_j) - (\mathbb{E}(z_j))^2.
\end{aligned}$$

On définit ainsi la statistique

$$z = z_1 + \cdots + z_{I-1}.$$

Sous l'hypothèse nulle, le test suppose que z suppose une loi normale. Ainsi,

$$\mathbb{E}(z) = \mathbb{E}(z_1) + \cdots + \mathbb{E}(z_{I-2}), \quad \text{et} \quad \mathbb{V}(z) = \mathbb{V}(z_1) + \cdots + \mathbb{V}(z_{I-2}).$$

Par conséquent, l'hypothèse d'indépendance (nulle) sera rejetée si, au seuil de 5%, z n'est pas comprise entre

$$\mathbb{E}(z) - 2\sqrt{\mathbb{V}(z)} \leq z \leq \mathbb{E}(z) + 2\sqrt{\mathbb{V}(z)}.$$

La figure 1.28 indique que le test est (partiellement) vérifié pour le triangle des données.

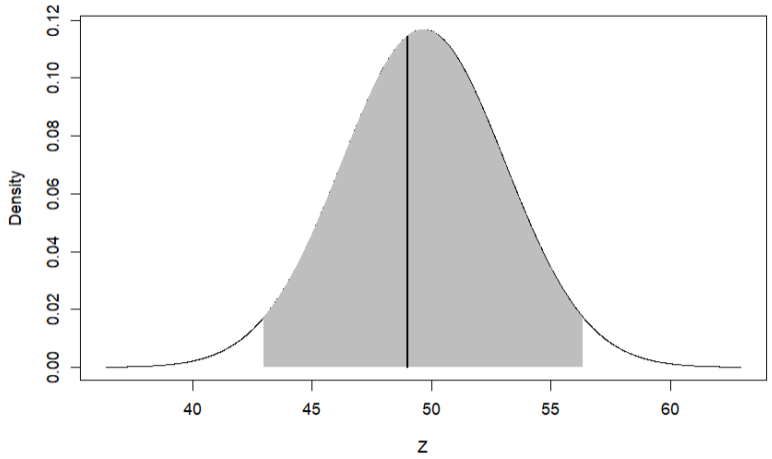


FIGURE 1.28 : Hypothèse 1.2.1 *(a)* du modèle de Mack

Hypothèse 1.2.1 *(b).* Pour l'hypothèse *(b)*, les points $(C_{i,j}, C_{i,j+1})$ doivent être alignés sur une droite passant par l'origine. Il existe aussi un test de corrélation des facteurs de développement qui

permet également de conclure. Si l'hypothèse *(b)* est vérifiée, alors $C_{i,k}/C_{i,k-1}$ et $C_{i,k+1}/C_{i,k}$ ne sont pas corrélés.

Le test d'indépendance utilise le coefficient de rang de Spearman. Sous l'hypothèse nulle, les facteurs de développement adjacents sont non corrélés. L'année de développement k est fixée, et les $C_{i,k+1}/C_{i,k}$ sont ordonnés dans l'ordre croissant. Soit $r_{i,k}$ pour $0 \leq i \leq I-k$ le rang des $C_{i,k+1}/C_{i,k}$ dans cet ordre. On a $1 \leq r_{i,k} \leq I-k$. En raisonnant de la même manière pour $C_{i,k}/C_{i,k-1}$, on obtient les rangs $1 \leq s_{i,k} \leq I-k$. À présent, le coefficient de Spearman, T_k, est alors exprimé par

$$T_k = 1 - 6\sum_{i=0}^{I-k} \frac{(r_{i,k} - s_{i,k})^2}{(I-k-3)(I-k)},$$

avec $-1 \leq T_k \leq 1$. Sous l'hypothèse nulle, $\mathbb{E}(T_k) = 0$ et $\mathbb{V}(T_k) = \frac{1}{(I-k-1)}$. Ainsi, une valeur positive ou négative élevée de T_k suggère que les facteurs de développement, entre les années de développement $k-1$ et k et $k+1$, sont corrélés. En calculant $T_1, \ldots, T_{I-2}$ sous l'hypothèse nulle, T est défini comme

$$T = \frac{\sum_{k=1}^{I-2} T_k}{\sum_{k=1}^{I-2} \frac{1}{\mathbb{V}(T_k)}},$$

et ses deux premiers moments sont

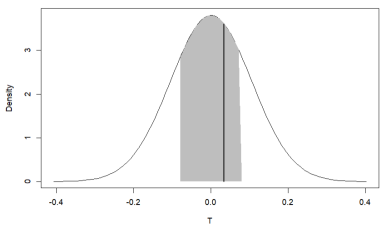
$$\mathbb{E}(T) = \sum_{k=1}^{I-2} \mathbb{E}(T_k) = 0 \quad \text{et} \quad \mathbb{V}(T) = \frac{\sum_{k=1}^{I-2}(I-k-1)^2 \mathbb{V}(T_k)}{\left(\sum_{k=1}^{I-2}(I-k-1)\right)^2} = \frac{1}{\frac{(I-2)(I-3)}{2}}.$$

En supposant les T_k distribués de manière symétrique autour de l'espérance (nulle), T est supposé être une loi normale. Dans un intervalle de confiance raisonnable, l'hypothèse nulle n'est pas rejetée si

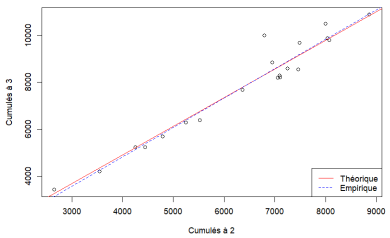
$$-\frac{0.67}{\sqrt{(I-2)(I-3)/2}} \leq T \leq \frac{0.67}{\sqrt{(I-2)(I-3)/2}}.$$

L'hypothèse est également (partiellement) vérifiée, comme le montre la figure 1.29.

Remarque 1. *Pour les hypothèses 1.2.1 (a) et (b), le terme « partiellement » signifie que les hypothèses sont vérifiées sur le triangle, hormis sur les années 2008-2013.*



(a) Test de corrélation des facteurs de développement



(b) C-C *plot* pour la 3e année de développement

FIGURE 1.29 : Hypothèse 1.2.1 *(b)* du modèle de Mack

Hypothèse 1.2.1 - *(c).* Enfin, l'incertitude des estimations du modèle de Mack peut être visualisée dans le diagnostic 1.30. L'incertitude croit avec les années de survenance. Selon les valeurs estimées et selon les années de développement, calendaires ou d'origine, l'ajustement aux données est plutôt bon et aucune tendance particulière dans les résidus ne semble se dégager, ce qui signifie que l'hypothèse 1.2.1 *(c)* est également satisfaite.

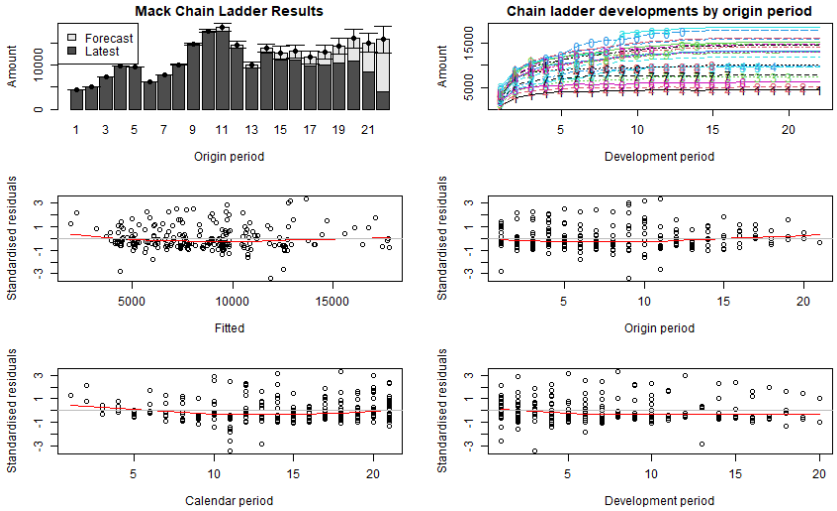


FIGURE 1.30 : Diagnostic du modèle de Mack

Une autre manière de valider l'hypothèse *(c)* est de tracer les résidus standardisés $\epsilon_{(i,j)} = \frac{C_{i,j+1} - C_{i,j} \cdot f}{\sqrt{C_{i,j}}}$. Aucune tendance particulière n'est observée dans les résidus, ce qui confirme que l'hypothèse est vérifiée.

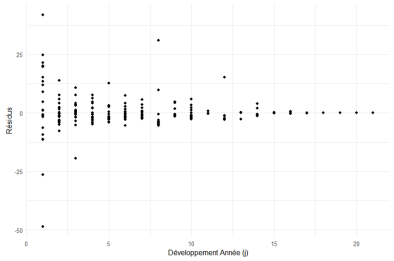
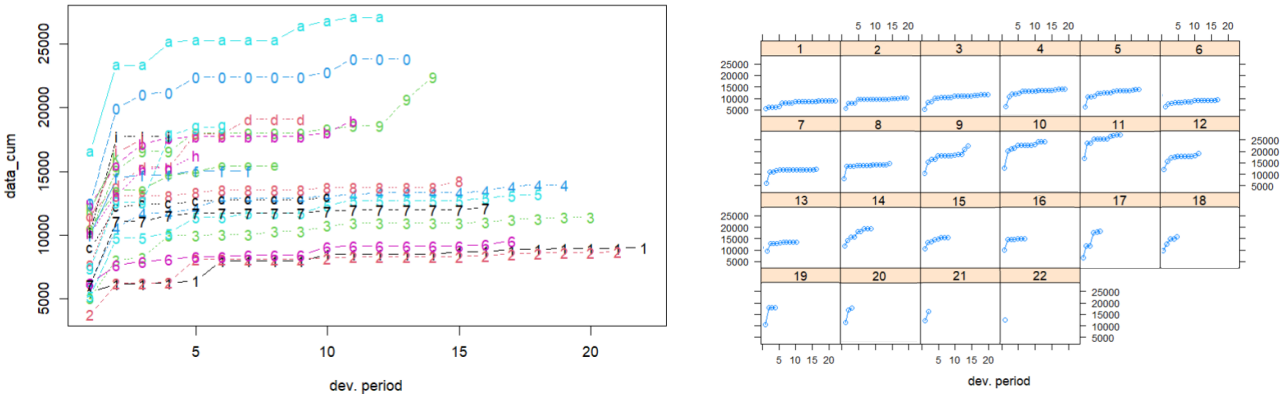


FIGURE 1.31 : Résidus $\epsilon_{(i,j)} = \frac{C_{i,j+1} - C_{i,j} \cdot f}{\sqrt{C_{i,j}}}$

1.3.4 Présentation du triangle de liquidation des charges totales (*incurred*)

La prise en compte des triangles des charges en plus des paiements dans les prédictions des réserves est une information pertinente, notamment en RC automobile, puisque certains sinistres peuvent rester ouverts longtemps, notamment pour les dommages corporels. Ce triangle est composé initialement des montants payés, des provisions dossier/dossier, et des recours. Cependant, pour rester dans les cas les plus standards, il est décidé de ne pas considérer les recours afin de réduire la volatilité liée à des mouvements de sinistres exceptionnels. Pour effectuer le retraitement à partir des données individuelles (sans d'ailleurs avoir la possibilité de faire toute autre modification sur les données du fait de l'absence de données ligne à ligne complètes), une approximation de la proportion des recours parmi les charges totales est estimée par année de développement. De plus, dans le chapitre 2, la méthode utilisée en

section 2.2.3 requiert d'exploiter les provisions dossier/dossier, donc de pouvoir séparer les recours des provisions et paiements. Les montants cumulés de charges sont disponibles en figure 1.32. La figure confirme que les années 2006 à 2008 sont bien des années atypiques.



(a) Vision d'ensemble (b) Par année de survenance

FIGURE 1.32 : Charges cumulées

Cependant, après avoir effectué le retraitement des recours, il est observé que les *incurred* continuent à se développer, d'après la figure 1.33, ce qui signifie que, même après vingt exercices, les réserves ne sont pas complètement nulles, ce qui semble peu probable et représente un biais à l'étude.

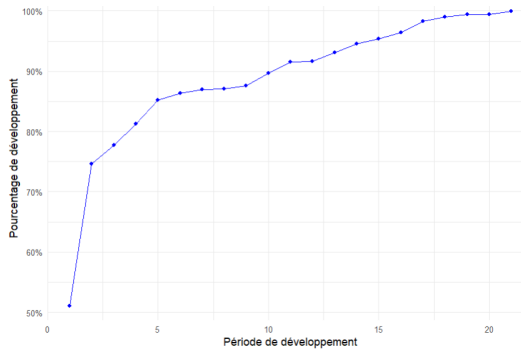


FIGURE 1.33 : Cadence de développement des charges

Ce chapitre a ainsi pour but de présenter le cadre d'évaluation des paramètres USP et les données sous différentes formes, ainsi que leur hétérogénéité. La Directive Solvabilité II ne distingue pas de branches entre la RC matérielle et corporelle, et aucune segmentation n'est effectuée dans les triangles pour calculer les paramètres USP dans le présent travail. Il s'agit là ainsi d'une de ses premières limites. L'absence de données ligne à ligne exhaustives et exactes et le manque de suivi dans le traitement des données constituent des limites non négligeables puisque la qualité des données est un élément crucial dans la demande d'utilisation des USP auprès de l'ACPR, comme vu en section 1.2.1. En ce sens, **les données ne satisfont pas les contraintes d'un dossier USP**. Le travail présente toutefois un intérêt puisqu'il permet de mettre en lumière la difficulté dans la constitution d'un tel dossier à adresser au régulateur. Les calculs des paramètres USP reposant sur des hypothèses arbitraires, le chapitre 2 présente les alternatives (ou approches concurrentes) existantes aujourd'hui aux deux méthodes réglementaires des USP, présentées en section 1.2.4 pour rappel.

Chapitre 2

Panorama et critiques des méthodes USP

L'objectif de ce chapitre est désormais de dresser l'état de l'art des méthodes concurrentes pouvant aboutir au calcul des paramètres USP. Dans le cadre des actes délégués, deux méthodes réglementaires sont aujourd'hui en vigueur. Il convient d'une part de les appliquer, et d'autre part de tenter de les remettre en question.

2.1 Critiques théoriques de la méthode réglementaire 1

L'exposé de cette section, qui se focalise sur la méthode 1 présentée en section 1.2.4, repose sur des travaux existants (Zimbidis, 2021). L'auteur précise le raisonnement mathématique de cette méthode et s'interroge sur l'hypothèse paramétrique 1.2.2 *(c)* selon laquelle y_t suit une loi log-normale. Il propose une revue de la formule en adaptant cette hypothèse aux lois Gamma et Pareto, distributions couramment utilisées en actuariat pour modéliser la sévérité des sinistres. La démonstration pour aboutir à la formule (1.8) est disponible en annexe A.2 et il convient de présenter les alternatives à la méthode 1 selon la distribution paramétrique considérée.

2.1.1 L'approche avec la loi Gamma

Dans le même ordre d'idée qu'en annexe A.2, si l'on suppose $y_t \sim \mathcal{G}a(\mu, \psi)$ alors $f(y) = \frac{\psi^\mu}{\Gamma(\mu)} y^{\mu-1} e^{-\psi y}$, avec $\mu, \psi > 0, y > 0$, et, en relation avec les hypothèses 1.2.2 *(a)* et *(b)*,

$$\begin{cases} \mathbb{E}(y_t) &= \frac{\mu}{\psi} = \beta x_t \\ \mathbb{V}(y_t) &= \frac{\mu}{\psi^2} = \sigma^2[(1-\delta)\bar{x} + \delta x_t^2] \end{cases}.$$

En résolvant le système de deux équations à deux inconnues,

$$\begin{cases} \mu = \frac{\beta x_t}{\frac{\sigma^2}{\beta}[(1-\delta)\bar{x}+\delta x_t]} = \frac{1}{\frac{\sigma^2}{\beta^2}[(1-\delta)\frac{\bar{x}}{x_t^2}+\delta]} \\ \psi = \frac{\beta}{\sigma^2[(1-\delta)\bar{x}+\delta x_t]} \end{cases}.$$

En gardant la même notation qu'en annexe A.2, $\mu = \pi_t^{-1}$ et $\psi = \frac{1}{\beta x_t \pi_t}$, où

$$\pi_t = \frac{\sigma^2}{\beta^2}\left[(1-\delta)\frac{\overline{x}}{x_t} + \delta\right]$$

En remplaçant dans l'expression de la densité, la vraisemblance s'exprime

$$L = \prod_{t=1}^{T} \frac{1}{\Gamma(\pi_t^{-1})}(\beta x_t \pi_t)^{\pi_t} y_t^{\pi_t^{-1}-1} e^{-\frac{y_t}{\beta x_t \pi_t}}.$$

La log-vraisemblance vaut par la suite

$$l = \ln(L) = \sum_{t=1}^{T} \ln\left(\frac{\pi_t \sqrt{\frac{y_t}{x_t}}}{y_t \Gamma(\pi_t^{-1})(\beta x_t \pi_t)^{\pi_t^{-1}}}\right) - \sum_{t=1}^{T} \frac{y_t}{\beta x_t \pi_t}.$$

Par changement de signe, la fonction équivalente à minimiser est

$$l'(\beta,\gamma,\delta) = \sum_{t=1}^{T} \frac{y_t}{x_t} \frac{1}{\beta x_t \pi_t(\gamma,\delta)} \sum_{t=1}^{T} \left[\ln y_t + \ln\Gamma(\pi_t^{-1}) + \ln(\beta x_t \pi_t)\right] - \sum_{t=1}^{T} \ln(\pi_t(\gamma,\delta)). \tag{2.1}$$

La volatilité est ainsi déterminée comme

$$\hat{\sigma} = e^{\hat{\gamma} + \ln\hat{\beta}}.$$

2.1.2 L'approche avec la loi de Pareto

En considérant à présent $y_t \sim \mathcal{P}ar(\mu,\psi)$ de type II, la densité associée est

$$f(y) = \mu\psi^{\mu}(\psi + y)^{-(\mu+1)}, \quad \mu > 2.$$

L'espérance et la variance connues, l'égalité avec les hypothèses 1.2.2 *(a)* et *(b)* permet d'obtenir les paramètres

$$\begin{cases} \mathbb{E}(y_t) = \dfrac{\psi}{\mu - 1} = \beta x_t \\ \mathbb{V}(y_t) = \dfrac{\mu\psi^2}{(\mu-1)^2(\mu-2)} = \sigma^2\left[(1-\delta)\overline{x}x_t + \delta x_t^2\right] \end{cases}.$$

Puis, avec

$$\pi_t = \frac{\sigma^2}{\beta^2}\left[(1-\delta)\frac{\overline{x}}{x_t} + \delta\right],$$

les paramètres sont alors

$$\begin{cases} \mu = \dfrac{2\pi_t}{\pi_t - 1} \\ \psi = \beta x_t\left(\dfrac{\pi_t + 1}{\pi_t - 1}\right) \end{cases}.$$

D'après le lemme 6 (Zimbidis, 2021), $V_t = \ln\left(\frac{\psi+y_t}{\psi}\right) \sim \mathcal{E}(\mu)$, et

$$V_t = \ln(\psi + y_t) - \ln(\psi) = \ln(\psi + y_t) - \ln\left(\beta x_t \frac{\pi_t + 1}{\pi_t - 1}\right) = \ln\left(\beta\frac{\pi_t + 1}{\pi_t - 1} + \frac{y_t}{x_t}\right) - \ln\left(\frac{\pi_t + 1}{\pi_t - 1}\right) - \ln\beta.$$

La vraisemblance de V_t vaut alors

$$L = \prod_{t=1}^{T} f_{V_t}(v_t) = \prod_{t=1}^{T}\left[\frac{2\pi_t}{\pi_t - 1} e^{-\frac{2\pi_t}{\pi_t - 1}v_t}\right] = 2^T\left[\prod_{t=1}^{T}\frac{\pi_t}{\pi_t - 1}\right] e^{-2\sum_{t=1}^{T}\frac{\pi_t}{\pi_t - 1}v_t},$$

et la log-vraisemblance

$$l = \ln(L) = \ln(2^T) + \ln\left[\prod_{t=1}^{T}\frac{\pi_t}{\pi_t - 1}\right] + \ln\left(e^{-2\sum_{t=1}^{T}\frac{\pi_t}{\pi_t - 1}v_t}\right).$$

Le problème revient ainsi à minimiser la fonction

$$l' = 2\sum_{t=1}^{T}\frac{\pi_t}{\pi_t - 1}v_t + \sum_{t=1}^{T}\ln\frac{\pi_t - 1}{\pi_t}$$

$$= 2\sum_{t=1}^{T}\frac{\pi_t(\gamma,\delta)}{\pi_t(\gamma,\delta) - 1}\left[\ln\left(\beta\frac{\pi_t(\gamma,\delta) + 1}{\pi_t(\gamma,\delta) - 1} + \frac{y_t}{x_t}\right) - \ln\left(\frac{\pi_t(\gamma,\delta) + 1}{\pi_t(\gamma,\delta) - 1}\right) - \ln\beta\right] + \sum_{t=1}^{T}\ln\frac{\pi_t(\gamma,\delta) - 1}{\pi_t(\gamma,\delta)},$$

ou encore

$$l'(\beta,\gamma,\delta) = 2\sum_{t=1}^{T}\frac{\pi_t(\gamma,\delta)}{\pi_t(\gamma,\delta) - 1}\ln\left(1 + \frac{\pi_t(\gamma,\delta) - 1}{\beta[\pi_t(\gamma,\delta) + 1]}\frac{y_t}{x_t}\right) + \sum_{t=1}^{T}\ln\frac{\pi_t(\gamma,\delta) - 1}{\pi_t(\gamma,\delta)}. \tag{2.2}$$

La valeur de la volatilité est ainsi estimée comme

$$\hat{\sigma} = e^{\hat{\gamma} + \ln\hat{\beta}}.$$

Ces approches sont appliquées en section 2.3.

2.2 Les alternatives à la méthode réglementaire 2

La méthode 2 (Merz et Wüthrich, 2008), dont le raisonnement et notations intermédiaires aboutissant à l'équation (1.9) sont décrits en annexe A.3, présente des limites. D'abord, elle est totalement dépendante du modèle de Mack Chain-Ladder et suppose donc les hypothèses 1.2.1 toujours vérifiées (ce qui n'est pas toujours le cas en pratique). De plus, la méthode n'envisage l'existence d'aucun facteur de queue, c'est-à-dire qu'elle suppose que les sinistres n'évoluent plus après la dernière année de développement du triangle de liquidation. Enfin, la méthode ne propose pas de distribution du CDR mais seulement des deux premiers moments. La méthode stochastique du bootstrap à un an (ou *re-reserving*), du modèle linéaire généralisé (GLM) et la méthode ECLRM (combinant l'information des charges) sont des alternatives actuelles à la méthode 2 des triangles.

2.2.1 Bootstrap à un an

La méthode du bootstrap à un an (Diers, 2009) consiste à répéter, pour un grand nombre de simulations, une procédure bootstrap afin d'obtenir une distribution du CDR. Cette méthode, décrite dans la figure 2.1, est très utile dans une logique de solvabilité, car elle permet d'obtenir des quantiles. Sur cette figure, l'expression « $\text{Pertes}^{(b)}_{\text{réserves}} = \text{BE}_0 - \text{Prest}^{(b)}_1 - \text{BE}^{(b)}_1$ » correspond à l'équation (1.2) généralisée à toutes les années d'accident. On retrouve également l'erreur de processus et d'estimation utilisées pour calculer la MSEP de l'équation (1.1). L'écart-type de la distribution du CDR obtenue reflète le risque de réserve. Dans le cadre du calcul du paramètre USP, il correspond à la racine carrée de la MSEP.

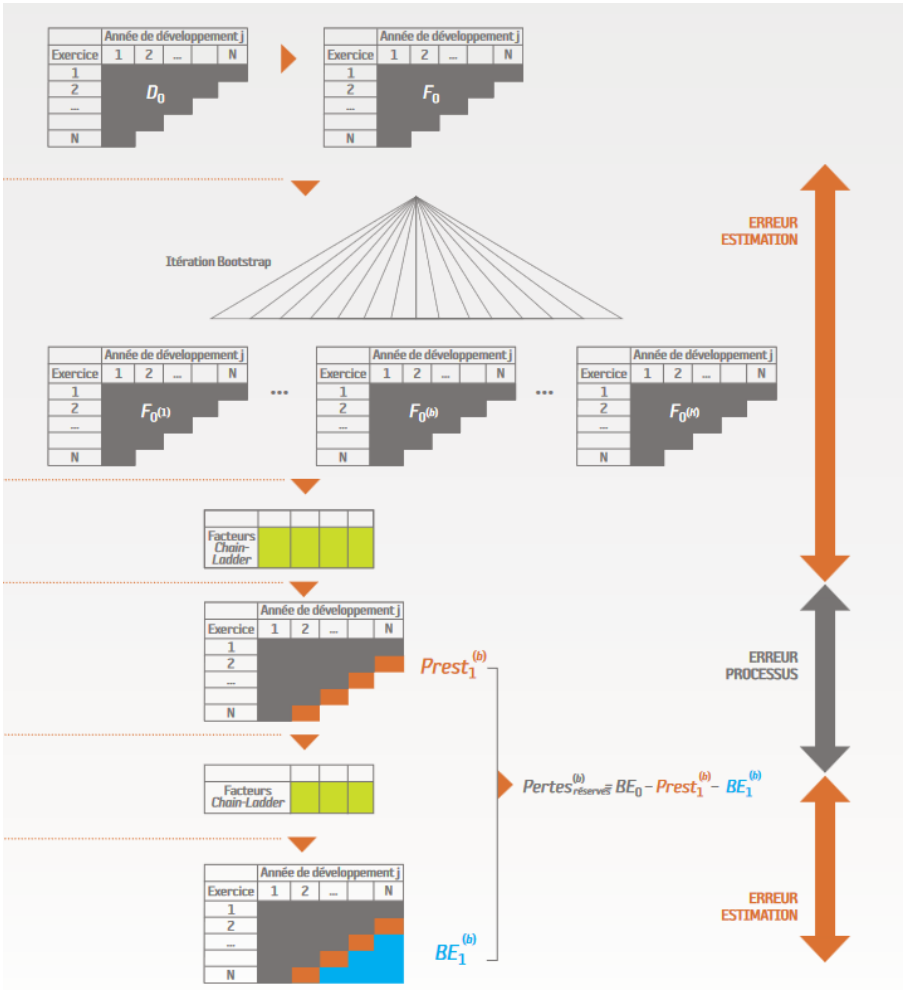


Figure 2.1 : Bootstrap à un an (Pillaudin, 2013)

On rappelle les algorithmes 1 et 2 du bootstrap et de sa version à un an. Dans l'algorithme 1, les $\mu_{i,j}$ correspondent aux montants incrémentaux ajustés avec les facteurs de lien de Chain-Ladder. En particulier, dans le cas Poisson, avec x_i le montant d'ultimes attendu par année d'accident, et y_j la part d'ultime par année de développement, $\mathbb{E}(Y_{i,j}) = \mu_{i,j} = x_i y_j$ et $\mathbb{V}(Y_{i,j}) = \phi x_i y_j$, où ϕ représente le paramètre de sur-dispersion (England et Verrall, 2002). Entre les deux algorithmes, la seule différence réside dans le fait que le processus stochastique s'applique à la diagonale $I+1$, estimée par la suite jusqu'à l'ultime par la méthode Chain-Ladder. Le bootstrap classique, simple à mettre en pratique, permet d'obtenir une distribution des réserves à l'horizon de liquidation totale du portefeuille (ultime), mais n'est pas adapté dans le cadre du calcul du SCR risque de réserve.

Algorithme 1 Procédure de bootstrap

1: Calcul des résidus de Pearson

$$r_{i,j} = \frac{Y_{i,j} - \mu_{i,j}}{\sqrt{\mathrm{V}(\mu_{i,j})}}.$$

2: Itération B fois des étapes suivantes :

1. rééchantillonnage des résidus avec remise $\tilde{r}_{i,j}^{(b)}$,
2. calcul des montants cumulés par inverse de $r_{i,j}$,

$$Y_{i,j}^{(b)} = \tilde{r}_{i,j}^{(b)} \sqrt{\mu_{i,j}} + \mu_{i,j},$$

3. cumul du triangle et déduction de $R^{(b)}$.

Algorithme 2 Procédure de bootstrap à un an

1: Estimation des $\hat{R}_i^I$.

2: Calcul des résidus

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

3: Itération B fois des étapes suivantes :

1. rééchantillonnage des résidus avec remise $\tilde{r}_{i,j}^{(b)}$,
2. calcul des nouveaux facteurs de lien individuels

$$\tilde{f}_{i,j}^{(b)} = \hat{f}_j + \sqrt{\frac{\hat{\sigma}_j^2}{Y_{i,j}}} \tilde{r}_{i,j}^{(b)},$$

3. calcul des nouveaux facteurs de lien moyennés

$$\tilde{f}_j^{I,(b)} = \frac{\sum_{i=0}^{I-j-1} C_{i,j} \tilde{f}_{i,j}^{(b)}}{\sum_{i=0}^{I-j-1} C_{i,j}},$$

4. simulation de la nouvelle diagonale $\tilde{C}_{i,I-i+1}^{I+1}$ selon la distribution choisie (Poisson sur-dispersé, Gamma, log-normale, etc.),
5. recalcul des facteurs de liens avec la diagonale $I+1$

$$\tilde{f}_j^{I+1,(b)} = \frac{\sum_{i=0}^{I-j-1} C_{i,j} f_{i,j}^{(b)} + C_{I-j,j} \tilde{f}_{I-j}^{I+1,(b)}}{\sum_{i=0}^{I-j-1} C_{i,j}},$$

6. estimation de $\tilde{R}_i^{I+1,(b)} = \tilde{C}_{i,I-i+1}^{(b)} \prod_{j=I-i+1}^{I-1} \tilde{f}_j^{I+1,(b)} - \tilde{C}_{i,I-i+1}^{(b)}$,
7. calcul de $\widehat{CDR}(I+1) = \sum_{i=0}^{I} \widehat{CDR}_i(I+1)$ à l'aide de la formule (1.2).

2.2.2 Modèle linéaire généralisé Poisson sur-dispersé

Cette autre approche concurrente à la méthode réglementaire 2 relative au calcul du paramètre USP est proposée dans la littérature (Cavastracci et Tripodi, 2018). Elle fournit une formule fermée (England et Verrall, 2002) pour estimer la volatilité à un an via l'utilisation d'un modèle linéaire généralisé (GLM) Poisson sur-dispersé (ODP). Le modèle Poisson sur-dispersé, sur lequel se base l'estimation de la MSEP, est usuel. Il existe une version détaillée (Wüthrich et Merz, 2015). Pour des détails plus généraux sur les modèles linéaires généralisés avec les quasi lois, le lecteur est invité à consulter d'autres ouvrages (Denuit et al., 2019).

Le modèle Poisson sur-dispersé. Dans leurs travaux (England et Verrall, 2002), les auteurs se basent sur le triangle de liquidation incrémental des paiements $Y_{i,j}$ comme illustré en figure 2.2. Dans le cadre des GLM, on suppose les $Y_{i,j}$ comme variable réponse. Les variables explicatives du modèle sont les années d'accident et de développement, agissant comme variables qualitatives. Pour rappel, pour tout exercice comptable $t \in \{0, ..., I\}$, avec les montants cumulés, les relations à garder à l'esprit sont

$$
\begin{aligned}
C_{i,j} &= C_{i,j-1} + Y_{i,j}, \\
C_{i,J} &= \sum_{k=0}^{J} Y_{i,k}, \\
C_{i,J}^{(t)} &= \sum_{k=0}^{t-i} Y_{i,k} + \sum_{k=t-i+1}^{J} \hat{Y}_{i,k}, \\
\hat{R}_i^{(t)} &= \sum_{k=t-i+1}^{J} \hat{Y}_{i,k}.
\end{aligned}
$$

i / j	0	1	$\cdots$	j	$\cdots$	J
1	Y_{10}	Y_{11}	$\cdots$	Y_{1j}	$\cdots$	Y_{1J}
2	Y_{20}	Y_{21}	$\cdots$			
$\vdots$	$\vdots$					
i	Y_{i0}		Y_{ij}			
$\vdots$	$\vdots$					
I	Y_{I0}					

Figure 2.2 : Triangle des paiements incrémentaux (Cavastracci et Tripodi, 2018)

Les hypothèses du modèle GLM sont énoncées ci-dessous. Les hypothèses 2.2.1 *(c)* et *(d)* sous-entendent que la fonction de lien du modèle g est la fonction logarithme.

Hypothèse 2.2.1. *(a) Les Y_{ij} sont stochastiquement indépendants tels que $Y_{ij} \sim \mathcal{P}ois\left(\frac{\mu_{ij}}{\phi}\right)$.*

(b) La densité appartient à la famille exponentielle

$$
f(y; \theta_{ij}, \phi) = \exp\left\{ \frac{\omega_{ij}}{\phi} \left[y\theta_{ij} - b(\theta_{ij}) \right] + c(y; \theta_{ij}, \phi) \right\},
$$

où ω_{ij} représente les poids, θ_{ij} le paramètre naturel, et ϕ le paramètre de dispersion.

(c) L'espérance s'écrit $\mathbb{E}(Y_{ij}) = g^{-1}(x_{ij}^{\top}\beta) = h(x_{ij}^{\top}\beta) = b'(\theta_{ij}) = \mu_{ij} = e^{c+a_i+b_j}$.

(d) La variance s'écrit $\mathbb{V}(Y_{ij}) = \frac{\phi}{\omega_{ij}} b''(\theta_{ij}) = \frac{\phi}{\omega_{ij}} \mathbb{V}(\mu_{ij}) = \phi \mu_{ij}$.

Dans le modèle, $X \in \mathbb{R}^{n \times p}$ représente la matrice *design* du modèle

$$X = \left[\begin{array}{c|c} \mathbf{1}_{n \times 1} & Id_{n \times (p-1)} \end{array} \right],$$

où $n = \frac{I(I+1)}{2}$ est le nombre d'observations, $p = I + J = 2I - 1$ le nombre de paramètres du modèle et $n - p$ le nombre de degrés de libertés. Ainsi,

$$g\left(\mathbb{E}[Y \mid X = x]\right) = \log(\mu_{ij}) = c + a_i + b_j.$$

Les variables $x_{i,j}$ représentent les réalisations des variables explicatives du modèle contenues dans X, correspondant à des vecteurs unitaires pour l'encodage des années d'accident et de développement. β le vecteur des paramètres estimés (par maximum de vraisemblance), g la fonction de lien (continue inversible), et $\mathbb{V}(\mu_{ij}) = b''\left(b'^{-1}(\mu_{ij})\right)$ est la fonction de variance. $\eta = X\beta$ représente le prédicteur linéaire avec $\beta = (c, a_1, \ldots, a_I, b_0, \ldots, b_J)^\top$, c l'intercept du modèle, a les paramètres d'années de survenance et b ceux liés aux années de développement (il est supposé $a_1 = b_0 = 0$) ; le paramètre de sur-dispersion ϕ est quant à lui estimé par

$$\hat{\phi} = \frac{1}{n-p} \sum_{i+j \leq t} \omega_{ij} \frac{(y_{ij} - \hat{\mu}_{ij})^2}{\mathbb{V}(\hat{\mu}_{ij})}.$$

Le modèle de Poisson sur-dispersé est adapté dans les scénarii où la variance est plus élevée que la moyenne. Pour estimer β, l'utilisation d'une fonction de quasi-vraisemblance offre une alternative solide lorsque les données ne suivent pas strictement les hypothèses requises pour les modèles de Poisson traditionnels, fournissant des estimations plus fiables dans de tels contextes. Cette fonction de quasi-vraisemblance du modèle est s'exprime comme

$$K(y; \beta, \phi) := \sum_{i+j \leq t} \omega_{ij} \int_{y_{ij}}^{\mu_{i,j}} \frac{y_{ij} - s}{\phi \mathbb{V}(s)} ds = \sum_{i+j \leq t} \frac{\omega_{ij}}{\phi} \left[y_{ij} \log \frac{\mu_{ij}}{y_{ij}} - \mu_{ij} + y_{ij} \right].$$

Pour mesurer la qualité d'ajustement du modèle, les résidus, la statistique de Pearson et la déviance sont utilisés et s'expriment respectivement comme

$$r_{ij} = \frac{y_{ij} - \hat{\mu}_{ij}}{\sqrt{\mathbb{V}(\hat{\mu}_{ij})/\omega_{ij}}} = \frac{y_{ij} - \hat{\mu}_{ij}}{\sqrt{\hat{\phi}\hat{\mu}_{ij}}}, \quad \text{si l'on suppose} \quad \omega_{ij} = 1,$$

$$\chi^2 = \sum_{i+j \leq t} \omega_{ij} \frac{(y_{ij} - \hat{\mu}_{ij})^2}{\mathbb{V}(\hat{\mu}_{ij})},$$

$$D(\hat{\mu}; y) = -2 \sum_{i+j \leq t} \omega_{ij} \left[y_{ij}(\hat{\theta}_{ij} - \theta^*_{ij}) - (b(\hat{\theta}_{ij}) - b(\theta^*_{ij})) \right] = -2\hat{\phi} K(y; \beta, \phi), \tag{2.3}$$

où $\hat{\theta}_{ij} = b'^{-1}(\hat{\mu}_{ij})$ et $\theta^*_{ij} = b'^{-1}(y_{ij})$.

L'estimateur du maximum de vraisemblance est noté $\tilde{\beta} = (\tilde{c}, \tilde{a}_1, \ldots, \tilde{a}_I, \tilde{b}_0, \ldots, \tilde{b}_J)^\top$. Pour i, j, lorsque $i + j > t$, les prédictions sont telles que

$$\begin{aligned}
\mathbb{E}(\hat{Y}_{ij}) &= \hat{\mu}_{ij} = h(\hat{\eta}_{ij}) = h(\hat{c} + \hat{a}_i + \hat{b}_j) = e^{\hat{c}+\hat{a}_i+\hat{b}_j}, \\
\mathbb{V}(\hat{Y}_{ij}) &= \hat{\phi}\mathbb{V}(\hat{\mu}_{ij}) = \phi\hat{\mu}_{ij}, \\
\mathbb{E}(\hat{R}_i) &= \sum_{j=t-i+1}^{J} \hat{\mu}_{ij}, \\
\mathbb{E}(\hat{R}) &= \sum_{i+j>t} \hat{\mu}_{ij}.
\end{aligned}$$

Proposition 2.4. *Les erreurs quadratiques de prédiction (MSEP) des estimateurs $\tilde{Y}_{ij}$, $\tilde{R}_i$ et $\tilde{R}$ s'expriment respectivement comme*

$$MSEP(\tilde{Y}_{ij}) = \hat{\phi}\hat{\mu}_{ij} + \left[h'(\hat{\eta}_{ij})\right]^2 \widehat{\mathbb{V}}(\hat{\eta}_{ij}),$$

$$MSEP(\tilde{R}_i) = \hat{\phi} \sum_{j=t-i+1}^{J} \hat{\mu}_{ij} + \sum_{j=t-i+1}^{J} \hat{\mu}_{ij}^2 x_{ij}^\top \widehat{\mathbb{V}}(\tilde{\beta}) x_{ij} + \sum_{\substack{j_1,j_2=t-i+1 \\ j_1\neq j_2}}^{J} \hat{\mu}_{ij_1}\hat{\mu}_{ij_2} x_{i_1,j_1}^\top \widehat{\mathbb{V}}(\tilde{\beta}) x_{i_2,j_2},$$

$$MSEP(\tilde{R}) = \hat{\phi} \sum_{i+j>t} \hat{\mu}_{ij} + \sum_{i+j>t} \hat{\mu}_{ij}^2 x_{ij}^\top \widehat{\mathbb{V}}(\tilde{\beta}) x_{ij} + \sum_{\substack{i_1+j_1>t \\ i_2+j_2>t \\ (i_1,j_1)\neq(i_2,j_2)}} \hat{\mu}_{i_1,j_1}\hat{\mu}_{i_2,j_2} x_{i_1,j_1}^\top \widehat{\mathbb{V}}(\tilde{\beta}) x_{i_2,j_2}.$$

La vision à un an avec le modèle Poisson sur-dispersé. Comme étudié en section 1.2.3, le CDR est une mesure à court terme du provisionnement. Son expression dans le modèle est spécifiée en fonction des facteurs de Chain-Ladder.

Proposition 2.5. *Dans le modèle linéaire généralisé, le CDR est estimé par*

$$\widehat{CDR}_{i,t+1} = \underbrace{\hat{C}_{i,J}^{(t)}}_{\text{ultime en } t} - \underbrace{\hat{C}_{i,J}^{(t)} \left(1 + \frac{\xi_{i,t-i+1} + \zeta_{i,t-i+1}}{\hat{C}_{i,t-i+1}^{(t)}}\right) \prod_{j=t-i+1}^{J-1} \left(1 + \alpha_j^{(t)} \frac{\xi_{t-j,j+1} + \zeta_{t-j,j+1}}{\hat{C}_{t-j,j+1}^{(t)}}\right)}_{\text{ultime en } t+1}.$$

Remarque 2. *Dans la proposition 2.5, les coefficients (α_j), dont l'expression est précisée en annexe A.4.2, s'interprètent comme le poids donné à l'année d'accident j dans le calcul du facteur de développement suivant. Les coefficients ξ et ζ correspondent aux résidus associés respectivement à la variance de l'erreur de processus et d'estimation tels que $\xi_{t-j,j+1} = Y_{t-j,j+1} - e^{c+a_{t-j}+b_{j+1}}$ et $\zeta_{t-j,j+1} = e^{c+a_j+b_{j+1}} - e^{\hat{c}+\hat{a}_j+\hat{b}_{j+1}}$.*

De la même manière qu'à la vision à l'ultime, et en vue de calculer l'erreur quadratique de prédiction du CDR, les notations suivantes sont introduites

$$\hat{q}_{k+1} = \frac{\frac{e^{\hat{b}_{k+1}}}{\sum_{j=0}^{k+1} e^{\hat{b}_j}} \left(e^{\hat{a}_{t-k}} + \alpha_k^{(t)} \sum_{i=t-k+1}^{t} e^{\hat{a}_i}\right)}{\sum_{i=t-J+1}^{I} e^{\hat{a}_i}}, \quad t - I \leq k \leq J-1,$$

$$\hat{r}_{k+1}^{(t)} = \frac{e^{\hat{b}_{k+1}}}{\sum_{j=0}^{k+1} e^{\hat{b}_j}} = 1 - \frac{1}{f_k^{(t)}}, \quad k = 0, \ldots, J-1,$$

$$\hat{s}_{i,k+1} = \begin{cases} \hat{r}_{t-i+1} & \text{si } t-k=i \\ \hat{r}_{k+1}\alpha_k^{(t)} & \text{si } 2 \leq t-k < i \end{cases}, \quad k = t-i, \ldots, J-1.$$

Ces notations décrivent les différentes contributions au modèle :

- les coefficients q correspondent aux poids de la volatilité globale par rapport aux années de développement,
- les coefficients r correspondent au poids du paramètre de la k-ième année de développement (qui diminue à mesure que l'année de développement augmente),
- les coefficients s font référence à la volatilité des années d'accident (dépendent des valeurs de r).

À la fin de l'article, les auteurs proposent une approche concurrente (à la méthode 2) remplaçant la formule (1.9) pour le calcul de la MSEP (aussi bien pour toutes les années de survenance que pour une année de survenance en particulier).

Proposition 2.6.

$$MSEP\left(\sum_{i=t-J+1}^{I} \widehat{CDR}_{i,t+1}\right) = \left(\sum_{i=t-J+1}^{I}\sum_{j=0}^{J} \hat{\mu}_{ij}\right)^2 \times \left[\hat{\phi}\sum_{k=t-I}^{J-1} \frac{\hat{q}_{k+1}^2}{\hat{\mu}_{t-k,k+1}} + \hat{q}^\top X_{(t+1)}^\top \widehat{\mathbb{V}}(\hat{\beta}) X_{(t+1)}\hat{q}\right],$$

Les preuves des propositions 2.4, 2.5 et 2.6 sont disponibles dans l'annexe, respectivement dans les sections A.4.1, A.4.2 et A.4.3. Ainsi, de manière analogue à l'équation (1.9) de Merz & Wüthrich, l'approche avec le modèle linéaire généralisé permet d'estimer l'erreur à un an du CDR, et donc d'obtenir les paramètres USP pour les années d'accident passées. La section 2.2.3 est une troisième alternative à la méthode 2.

2.2.3 La méthode ECLRM, combinant charges et paiements

La méthode ECLRM (*Extended-Complementary-Loss-Ratio Method*) est une méthode basée sur les triangles de paiements et des charges (Dahms et al., 2009). Elle tire en partie son inspiration des premiers travaux sur la combinaison des sources d'information que sont les charges (*incurred*) et les montants payés (Wüthrich et Merz, 2015), et notamment du modèle pionnier de Munich Chain-Ladder (Quarg et Mack, 2004), dont les principaux résultats sont rappelés en préliminaire. La méthode ECLRM se base sur l'idée que les provisions dossier/dossier représentent une mesure de risque d'exposition, que les paiements et les changements d'*incurred* de la période de développement suivante sont proportionnels aux montants de provisions de l'année en cours, et que les années d'accident sont indépendantes. Une présentation académique donne un exemple d'application de la méthode (Dahms, 2021). Avec respectivement $C_{i,j}^{Pa}$ et $C_{i,j}^{In}$ les montants cumulés des paiements et des charges, les autres notations employées dans le modèle sont

$$\begin{aligned}
Y_{i,j}^{Pa} &= C_{i,j}^{Pa} - C_{i,j-1}^{Pa}, && \text{incréments des paiements,} \\
Y_{i,j}^{In} &= C_{i,j}^{In} - C_{i,j-1}^{In}, && \text{incréments des } \textit{incurred}, \\
R_{i,j} &= C_{i,j}^{In} - C_{i,j}^{Pa} = R_{i,j-1} + Y_{i,j}^{In} - Y_{i,j}^{Pa}, && \text{provisions dossier/dossier,} \\
C_{i,j} &= (C_{i,j}^{Pa}, C_{i,j}^{In}), && \text{couples de montants cumulés,} \\
Y_{i,j} &= (Y_{i,j}^{Pa}, Y_{i,j}^{In}), && \text{couples d'incréments,} \\
B_k &= \{C_{i,j}; 0 \leq i, 0 \leq j \leq k\}, && \text{information jusqu'au temps } k.
\end{aligned}$$

Un mot sur le modèle de Munich Chain-Ladder. Les hypothèses du modèle Munich Chain-Ladder sont les suivantes.

Hypothèse 2.2.2. *(a) Il existe deux constantes, λ^P et λ^I, telles que*

$$\mathbb{E}\left[\mathrm{Res}\left(\frac{C^{Pa}_{i,s+1}}{C^{Pa}_{i,s}}\middle|\mathcal{P}_i(s)\right)\middle|\mathcal{B}_i(s)\right] = \lambda^P \cdot \mathrm{Res}\left(Q^{-1}_{i,s}\middle|\mathcal{P}_i(s)\right),$$

et

$$\mathbb{E}\left[\mathrm{Res}\left(\frac{C^{In}_{i,s+1}}{C^{In}_{i,s}}\right)\middle|\mathcal{B}_i(s)\right] = \lambda^I \cdot \mathrm{Res}\left(Q^{-1}_{i,s}\middle|\mathcal{I}_i(s)\right),$$

où

$$Q_i := \frac{C^{Pa}_i}{C^{In}_i} = \frac{C^{Pa}_{i,t}}{C^{In}_{i,t}}, \quad t \in \{0,\dots,I\},$$

$$\mathcal{P}_i(s) := \{C^{Pa}_{i,1},\dots,C^{Pa}_{i,s}\}, \quad \mathcal{I}_i(s) := \{C^{In}_{i,1},\dots,C^{In}_{i,s}\},$$

$$\mathcal{B}_i(s) := \{C^{Pa}_{i,1},\dots,C^{Pa}_{i,s},C^{In}_{i,1},\dots,C^{In}_{i,s}\},$$

et

$$\mathrm{Res}(X \mid \mathcal{C}) := \frac{X - \mathbb{E}(X \mid \mathcal{C})}{\sigma(X \mid \mathcal{C})}.$$

L'hypothèse 2.2.2 est équivalente à l'existence de facteurs de développement

$$f^{Pa}_{MCL}(i,s) = \mathbb{E}\left(\frac{C^{Pa}_{i,s+1}}{C^{Pa}_{i,s}}\middle|B_i(s)\right) = f^P_s + \lambda^P \cdot \frac{\sigma\left(\frac{C^{Pa}_{i,s+1}}{C^{Pa}_{i,s}}\middle|C^{Pa}_i(s)\right)}{\sigma\left(Q^{-1}_{i,s}C^{Pa}_i(s)\right)} \cdot \left(Q^{-1}_{i,s} - \mathbb{E}\left(Q^{-1}_{i,s}C^{Pa}_i(s)\right)\right),$$

et

$$f^{In}_{MCL}(i,s) = \mathbb{E}\left(\frac{C^{In}_{i,s+1}}{C^{In}_{i,s}}\middle|B_i(s)\right) = f^I_s + \lambda^I \cdot \frac{\sigma\left(\frac{C^{In}_{i,s+1}}{C^{In}_{i,s}}\middle|C^{In}_i(s)\right)}{\sigma\left(Q^{-1}_{i,s}C^{In}_i(s)\right)} \cdot \left(Q^{-1}_{i,s} - \mathbb{E}\left(Q^{-1}_{i,s}C^{In}_i(s)\right)\right).$$

Ce modèle est utilisé dans la section 3.5.3 comme référence de marché.

Le modèle ECLRM. Les provisions $R_{i,j}$ sont des prédicteurs des $Y_{i,j+1}$. Le modèle suppose les hypothèses suivantes.

Hypothèse 2.2.3. *(a) Il existe f_j, g_j et une matrice de covariance définie positive 2×2 $\Sigma_j = (\sigma^{m,n}_j)_{m,n=1,2}$, pour tout $j = 0,\dots,J-1$ tels que*

$$\mathbb{E}[Y_{i,j+1} \mid B_j] = (R_{i,j}f_j, R_{i,j}g_j),$$

(b) $\mathrm{Cov}(Y_{i,j+1}, Y_{i,j+1} \mid B_j) = R_{i,j}\Sigma_j$,

(c) $\mathbb{E}[R_{i,j+1} \mid B_j] = R_{i,j}(1 + g_j - f_j) = R_{i,j}h_j$,

(d) $\mathbb{V}(R_{i,j+1} \mid B_j) = R_{i,j}(\sigma^{1,1}_j - 2\sigma^{1,2}_j + \sigma^{2,2}_j)$.

Dans le modèle, les estimateurs des facteurs de développement sont donnés par

$$\hat{f}_j^I = \frac{\sum_{i=0}^{j-1} Y_{i,j+1}^{Pa}}{\sum_{i=0}^{j-1} R_{i,j}}, \quad \hat{f}_j^{I+1} = \frac{\sum_{i=0}^{j} Y_{i,j+1}^{Pa}}{\sum_{i=0}^{j} R_{i,j}}, \quad \hat{g}_j^I = \frac{\sum_{i=0}^{j-1} Y_{i,j+1}^{In}}{\sum_{i=0}^{j-1} R_{i,j}}, \quad \hat{g}_j^{I+1} = \frac{\sum_{i=0}^{j} Y_{i,j+1}^{In}}{\sum_{i=0}^{j} R_{i,j}},$$

$$\hat{h}_j^I = 1 + \hat{g}_{j+1}^I - \hat{f}_{j+1}^I, \quad \hat{h}_j^{I+1} = 1 + \hat{g}_j^{I+1} - \hat{f}_j^{I+1}.$$

De même, les estimateurs des coefficients de volatilité sont donnés par

$$\begin{cases} \hat{\sigma}_j^{1,1} = \dfrac{1}{I-j-1} \displaystyle\sum_{i=0}^{I-j-1} R_{i,j} \left(\frac{Y_{i,j+1}^{Pa} - \hat{f}_j^I R_{i,j}}{R_{i,j}} \right)^2, \\ \hat{\sigma}_j^{2,2} = \dfrac{1}{I-j-1} \displaystyle\sum_{i=0}^{I-j-1} R_{i,j} \left(\frac{Y_{i,j+1}^{In} - \hat{g}_j^I R_{i,j}}{R_{i,j}} \right)^2, \\ \hat{\sigma}_j^{1,2} = \dfrac{1}{I-j-1} \displaystyle\sum_{i=0}^{I-j-1} R_{i,j} \left(\frac{Y_{i,j+1}^{Pa} - \hat{f}_j^I R_{i,j}}{R_{i,j}} \right) \left(\frac{Y_{i,j+1}^{In} - \hat{g}_j^I R_{i,j}}{R_{i,j}} \right). \end{cases}$$

Les prédictions du modèle sont données par

$$\begin{aligned} \hat{R}_{i,j} &= R_{i,I-i} \prod_{k=i}^{j-1} \hat{h}_k^I, \\ (\hat{Y}_{i,j}^{Pa})^I &= R_{i,I-i} \prod_{k=i}^{j-2} \hat{h}_k^I \hat{f}_{j-1}^I, \\ (\hat{Y}_{i,j}^{In})^I &= R_{i,I-i} \prod_{k=i}^{j-2} \hat{h}_k^I \hat{g}_{j-1}^I, \\ \hat{C}_{i,J} &= C_{i,I-i}^{In} + \sum_{j=I-i+1}^{J} (\hat{Y}_{i,j}^{In})^I = C_{i,I-i}^{Pa} + \sum_{j=I-i+1}^{J} (\hat{Y}_{i,j}^{Pa})^I, \end{aligned}$$

et de manière analogue à la section 2.2.2, le modèle fournit une formule de l'erreur quadratique du CDR.

Proposition 2.7.

$$\widehat{MSEP}_{\widehat{CDR}_i((I+1)|D_I)}(0) = \sum_{j,m=I-i+1}^{J} \left(\widehat{Y}_{i,j}^{Pa,I} \left(\widehat{Y}_{i,m}^{Pa,I} \right) \left[\frac{\delta_{I-i}^{-1} \widehat{\alpha}_{j,m,I-i}}{\sum_{n=0}^{I-i-1} R_{n,I-i}} + \sum_{k=I-i+1}^{\min(j,m)-1} \frac{\delta_k \widehat{\alpha}_{j,m,k}}{\sum_{n=0}^{I-k-1} R_{n,k}} \right] \right),$$

$$\begin{aligned} \widehat{MSEP}_{\sum_{i=1}^{I} \widehat{CDR}_i((I+1)|D_I)}(0) &= \sum_{i=1}^{I} \widehat{MSEP}_{\widehat{CDR}_i((I+1)|D_I)}(0) \\ &\quad + 2 \sum_{1 \le i < n \le I} \sum_{j,m=I-i+1}^{J} \left(\widehat{Y}_{i,j}^{Pa,I} \widehat{Y}_{n,m}^{Pa,I} \left[\frac{\widehat{\alpha}_{j,m,I-i}}{\sum_{l=0}^{I-i} R_{l,I-i}} + \sum_{k=I-i+1}^{\min(j,m)-1} \frac{\delta_k \widehat{\alpha}_{j,m,k}}{\sum_{l=0}^{I-k-1} R_{l,k}} \right] \right), \end{aligned}$$

avec

$$\delta_j = \frac{R_{I-j,j}}{\sum_{i=0}^{I-j} R_{i,j}} \in [0;1],$$

et

$$\hat{\alpha}_{j,m,k} = \begin{cases} \frac{\hat{\sigma}_k^{1,1}}{\hat{f}_k^2} & \text{pour } m = j = k+1, \\ \left(\frac{\hat{\sigma}_k^{1,2} - \hat{\sigma}_k^{1,1}}{\hat{f}_k \hat{h}_k}\right)^2 & \text{pour } m > j = k+1 \text{ ou } j > m = k+1, \\ \left(\hat{\sigma}_k^{1,1} - 2\hat{\sigma}_k^{1,2} + \hat{\sigma}_k^{2,2}\right)/(\hat{h}_k)^2 & \text{pour } m > j > k+1 \text{ ou } j > m > k+1. \end{cases}$$

En calculant les prédictions des réserves et l'erreur quadratique du CDR, on peut, comme les méthodes précédentes, aboutir au calcul de paramètres USP. La preuve de la proposition 2.7 est disponible dans l'annexe en section A.5.1.

2.3 Application des méthodes

Dans cette section, il convient de mettre en pratique les méthodes des actes délégués (méthode log-normale pour la méthode 1 et méthode de Merz & Wüthrich pour la méthode 2), les alternatives présentées dans les sections 2.1 et 2.2 (approches Gamma/Pareto pour la méthode 1 et approches par bootstrap, GLM et ECLRM pour la méthode 2) ainsi que d'étudier si une ou plusieurs méthodes USP se distingue(nt) parmi les autres avec un résultat plus précis.

2.3.1 Application de la méthode réglementaire 1 et des approches concurrentes Gamma et Pareto

Construction de la méthode. Dans cette méthode historique, les approches paramétriques log-normale, Gamma et Pareto de la méthode 1, décrites dans les sections 1.2.4 et 2.1, sont appliquées. Il convient de rappeler comment sont calculés x_t et y_t (Ndao, 2018). Les figures 2.3 et 2.4 fournissent un exemple de calcul des deux composantes, il convient de raisonner de la même manière pour les autres exercices.

$$x_{2015} = \text{Prov}_{2014,0} + \text{Prov}_{2013,1} + \text{Prov}_{2012,2},$$

$$y_{2015} = \text{Prest}_{2014,1} + \text{Prest}_{2013,2} + \text{Prest}_{2012,3} + \text{Prov}_{2014,1} + \text{Prov}_{2013,2} + \text{Prov}_{2012,3}.$$

	0	1	2	3	4
2012	$Prest_{2012,0}$	$Prest_{2012,1}$	$Prest_{2012,2}$	$Prest_{2012,3}$	$Prest_{2012,4}$
2013	$Prest_{2013,0}$	$Prest_{2013,1}$	$Prest_{2013,2}$	$Prest_{2013,4}$	
2014	$Prest_{2014,0}$	$Prest_{2014,1}$	$Prest_{2014,2}$		
2015	$Prest_{2015,0}$	$Prest_{2015,1}$			
2016	$Prest_{2016,0}$				

FIGURE 2.3 : Triangle des prestations ou paiements incrémentaux (Ndao, 2018)

La table 2.1 fournit les valeurs calculées des x_t et y_t sur les données. L'algorithme sélectionné pour déterminer les paramètres à chaque exercice δ et γ qui minimisent les équations (1.8) , (2.1), (2.2) est l'algorithme L-BFGS-B (Byrd et al., 1994). Cet algorithme, présenté en annexe A.7, a pour avantage sa simplicité (lorsque le calcul des dérivées partielles secondes est non-trivial, une approximation est proposée), sa stabilité (l'erreur est réduite au fil des itérations) et son adéquation avec les résolutions

	0	1	2	3	4
2012	$Prov_{2012,0}$	$Prov_{2012,1}$	$Prov_{2012,2}$	$Prov_{2012,3}$	$Prov_{2012,4}$
2013	$Prov_{2013,0}$	$Prov_{2013,1}$	$Prov_{2013,2}$	$Prov_{2013,4}$	
2014	$Prov_{2014,0}$	$Prov_{2014,1}$	$Prov_{2014,2}$		
2015	$Prov_{2015,0}$	$Prov_{2015,1}$			
2016	$Prov_{2016,0}$				

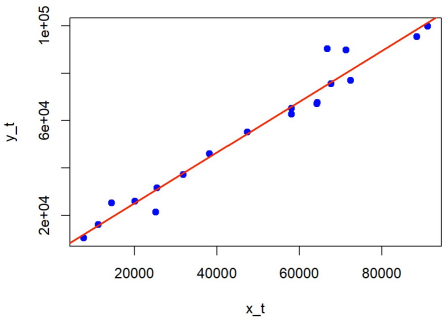
FIGURE 2.4 : Triangle cumulé des provisions (NDAO, 2018)

de fonctions sous contraintes inférieurement et supérieurement bornées (ce qui est le cas ici, du fait de la définition de δ).

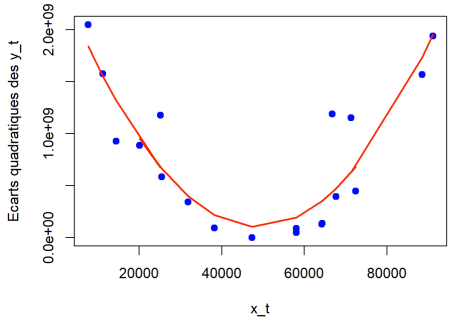
Année calendaire	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
x_t	7618	11225	14393	25208	20064	25419	31798	38253	47304	58070	58056	64227	64333	66756	67675	72473	71379	88581	91207
y_t	10502	16063	25281	21442	25963	31549	37260	46056	55146	62694	65238	67045	67542	90240	75595	76909	89707	95384	99780

TABLE 2.1 : Calcul des x_t et y_t (milliers €)

Vérification des hypothèses. La vérification des hypothèses 1.2.2 est illustrée dans les figures 2.5 et 2.6. Pour l'hypothèse 1.2.2 *(a)*, la régression linéaire fournit un R^2 de 96% avec un intercept statistiquement non significatif et une p-value de l'ordre de 10^{-13} pour la F-statistique associée. Pour l'hypothèse 1.2.2 *(b)*, le modèle donne un R^2 de 81% et une p-value de l'ordre de 10^{-6} pour la F-statistique. Ces deux premières hypothèses sont donc validées. Pour l'hypothèse 1.2.2 *(c)*, le test d'Anderson-Darling donne une *p-value* de 6,5% alors que celui de Shapiro-Wilk indique une *p-value* de 5,9%, ce qui n'indique pas clairement si l'hypothèse de log-normalité des données est significativement vérifiée. La figure 2.6b vérifie quant à elle l'hypothèse 1.2.2 *(d)*.



(a) Hypothèse *(a)*

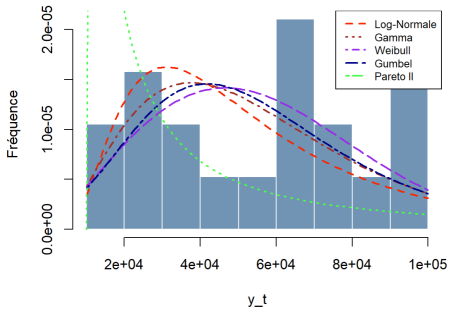


(b) Hypothèse *(b)*

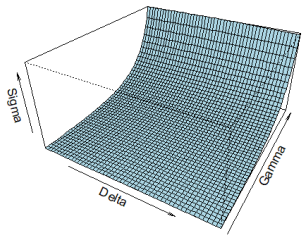
FIGURE 2.5 : Hypothèses 1.2.2 *(a)* et *(b)* de la méthode 1

Qualité d'ajustement des lois aux données. D'autres distributions ajustées, comme les distributions Gamma et Pareto (introduites dans la section 2.1) qui sous-tendent le calcul des USP, sont indiquées en figure 2.6a. Pour distinguer si une distribution se démarque plus qu'une autre, les Q-Q *plot* sont présentés en figures 2.7 et 2.8. Ces figures montrent que les lois Gamma et Weibull semblent le mieux s'ajuster aux données. La loi Pareto s'ajuste très mal aux données. Quant à la loi log-normale, elle s'ajuste bien aux données pour les valeurs les plus faibles puis surestime les réserves.

Enfin, la table 2.2 apporte également des détails sur les critères AIC, BIC et les log-vraisemblances, pour chaque distribution ajustée. Pour rappel, la meilleur ajustement est celui qui minimise les critères AIC/BIC et maximise la log-vraisemblance. D'après cette table, dans l'ordre, les lois Weibull, Gamma puis log-normale sont les plus adéquates.

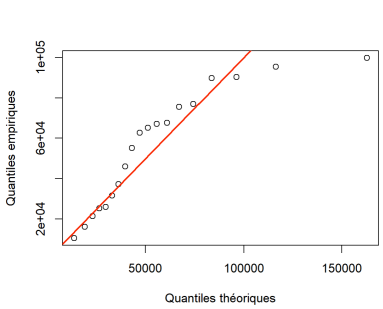


(a) Hypothèse *(c)*

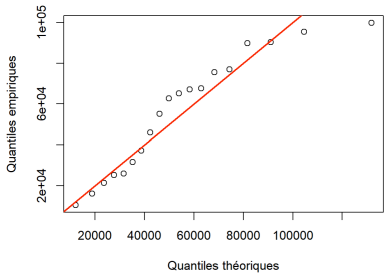


(b) Hypothèse *(d)*

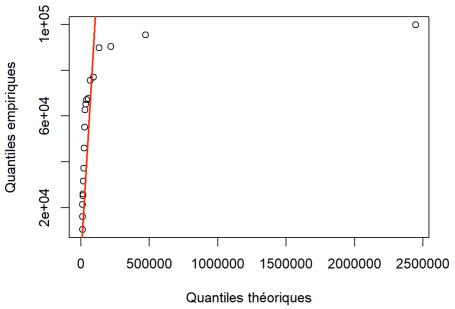
FIGURE 2.6 : Hypothèses 1.2.2 *(c)* et *(d)* de la méthode 1



(a) Q-Q *plot* log-normal

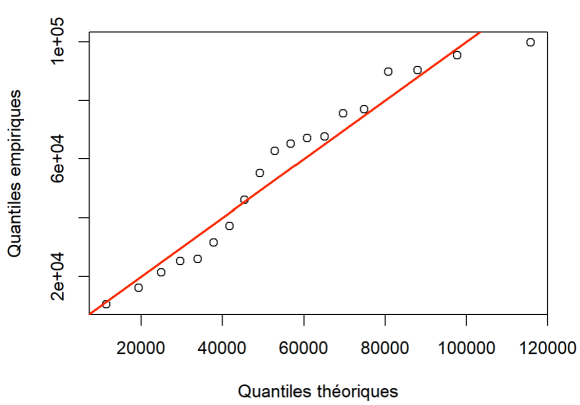


(b) Q-Q *plot* Gamma

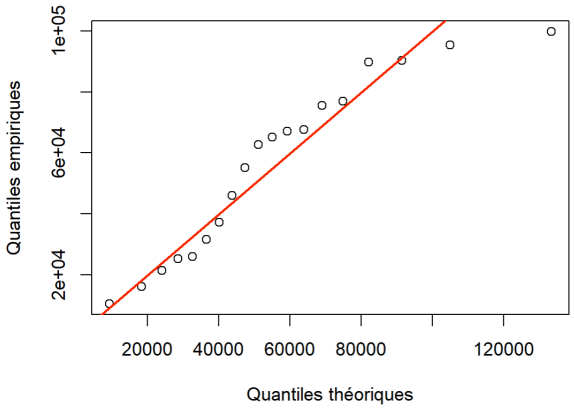


(c) Q-Q *plot* Pareto II

FIGURE 2.7 : Ajustement des lois proposées pour le calcul des USP (ZIMBIDIS, 2021)



(a) Q-Q *plot* Weibull



(b) Q-Q *plot* Gumbel

FIGURE 2.8 : Ajustement aux lois Weibull et Gumbel

Paramètres USP obtenus avec méthode 1, Gamma et Pareto. Le fait que la distribution Gamma soit plus adaptée aux y_t par rapport à la loi log-normale laisse penser que les paramètres USP obtenus avec l'approche Gamma proposée dans la section 2.1.1 sont plus robustes. La table 2.3 fournit les paramètres USP obtenus avec les approches proposées (ZIMBIDIS, 2021). Avec l'approche Gamma, les paramètres USP valent plus du double que par rapport aux approches log-normale standard et Pareto. En terme de charge en capital, toutes choses étant égales par ailleurs, si la captive suit l'approche avec la loi Gamma, la charge en capital allouée au SCR primes et réserves sera plus importante en conséquence.

Distribution	Log-vraisemblance	AIC	BIC
Log-Normale	-222,9	449,8	451,7
Gamma	-221,7	447,4	449,3
Weibull	-220,9	445,8	447,7
Gumbel	-222,1	448,2	450,0
Pareto II	-231,1	466,2	468,1

TABLE 2.2 : Critères de comparaison des différentes distributions

Année calendaire	Log-Normale	Gamma	Pareto
2004	16,91%	41,72%	40,24%
2005	17,11%	42,08%	36,22%
2006	17,03%	41,89%	33,06%
2007	16,76%	41,22%	30,66%
2008	16,66%	40,97%	27,27%
2009	16,48%	40,50%	24,55%
2010	16,08%	39,56%	22,74%
2011	15,93%	39,18%	21,06%
2012	15,64%	38,48%	20,16%
2013	15,44%	37,99%	19,26%
2014	15,80%	38,87%	19,94%
2015	15,72%	38,67%	18,97%
2016	15,58%	38,32%	18,30%
2017	15,72%	38,65%	17,99%
2018	15,59%	38,35%	17,33%
2019	15,52%	38,16%	16,66%

TABLE 2.3 : Paramètres USP pour les distributions log-normale, Gamma et Pareto

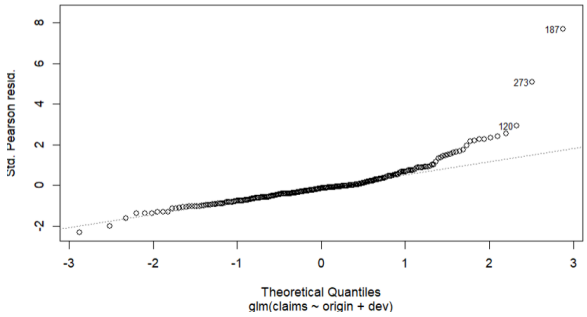
2.3.2 Application de la méthode 2 réglementaire et des approches concurrentes bootstrap à un an, modèle linéaire généralisé (GLM) et ECLRM

La méthode 2 réglementaire : Merz & Wüthrich. Comme mentionné dans la section 1.2.4, la méthode est construite en se basant sur le triangle cumulé des paiements, les estimations fournies par Chain-Ladder et sur l'équation (1.9). L'hypothèse 1.2.3 est de plus vérifiée sur l'ensemble du triangle de liquidation.

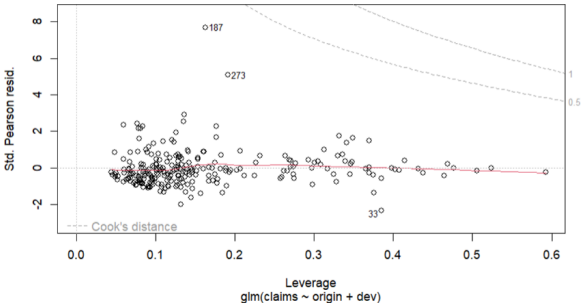
Le bootstrap à un an. Comme rappelé dans la section 2.2.1, le paramètre USP calculé par cette méthode correspond au ratio entre l'écart-type de la distribution du CDR, et les IBNR estimés par bootstrap. Le *package* `ChainLadder` permet d'obtenir la distribution du CDR aisément. Pour rappel, le CDR correspond à l'écart entre les IBNR estimés initialement et les IBNR ré-estimés avec une diagonale supplémentaire.

Analyse des résidus du modèle généralisé Poisson sur-dispersé. Les résultats de l'analyse du GLM sont indiqués en figure 2.9 pour le triangle des paiements complet. D'après la figure 2.9a, qui compare les quantiles des résidus de Pearson standardisés avec une distribution normale, les points suivent globalement la diagonale théorique, ce qui permet de juger que les résidus sont, dans l'ensemble, bien normaux (en pratique, il faudrait peut-être envisager un second modèle pour les valeurs extrêmes). Dans la figure 2.9b, aucune observation ne semble apporter davantage de bruit que d'information pertinente au modèle, donc tous les points peuvent être conservés. Les distances de Cook, qui mesurent l'influence des observations sur le pouvoir prédictif du modèle, sont en effet toutes inférieures à 0,5, ce qui signifie qu'aucune valeur *outlier* ne doit forcément être écartée. Enfin, les graphiques 2.9d et 2.9c manifestent une légère hétéroscédasticité dans le modèle, ce qui va à l'encontre de l'hypothèse de variance constante des résidus. Cependant, dans le cadre des modèles sur-dispersés

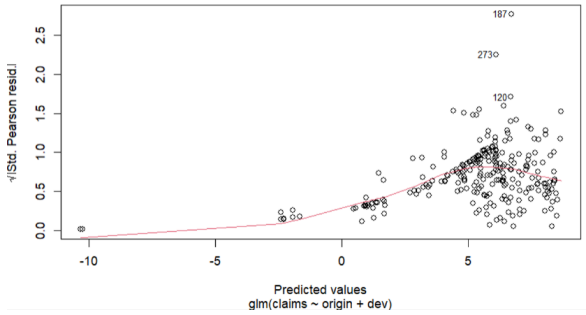
comme celui-ci, ces graphiques sont tout à fait communs.



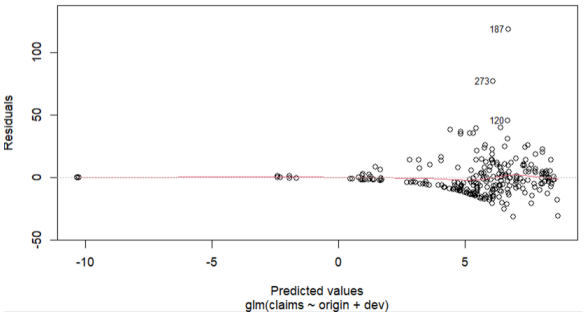
(a) Q-Q *plot* de normalité



(b) Distances de Cook



(c) *Scale-location* plot



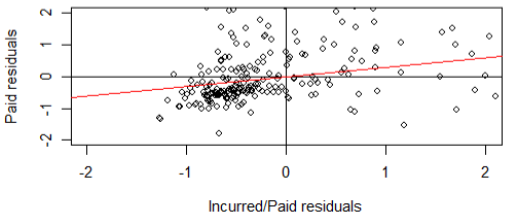
(d) Résidus du GLM

FIGURE 2.9 : Diagnostic du GLM

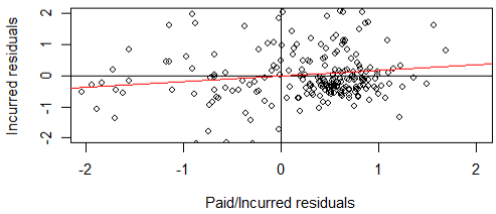
Scores du modèle linéaire généralisé. Le modèle GLM offre un pseudo R^2 de McFadden de 86%, une déviance (décrite en équation (2.3)) totale de 52 711, et une dispersion de 284.

Contribution des charges dans la méthode ECLRM. La mise en place de la méthode ECLRM décrite en section 2.2.3 s'effectue conjointement avec le triangle des provisions déduit du triangle charges présenté section 1.3.4. Tout comme pour la méthode 2, le paramètre USP obtenu dans la méthode ECLRM est calculé comme le ratio entre l'erreur quadratique du CDR (d'où la proposition 2.7) et les réserves estimées. La méthode permet d'obtenir une projection à l'ultime du triangle de paiements. Le code associé est fourni en annexe A.6.

Remarque 3. *Pour revenir à l'hypothèse 2.2.2 du modèle pionnier de la même famille que la méthode ECLRM, la figure indique que l'hypothèse est vérifiée, puisque* $\lambda^P = 0.30$ *représente le coefficient directeur de la figure 2.10a et* $\lambda^I = 0.18$ *représente le coefficient directeur de la figure 2.10b.*



(a) Résidus standardisés des payés par rapport au ratio charges/payés



(b) Résidus standardisés des charges par rapport au ratio payés/charges

FIGURE 2.10 : Hypothèses 2.2.2 du modèle Munich Chain-Ladder

Qualité d'ajustement des différentes méthodes. Les écarts relatifs entre les réserves estimées, fournies en annexe A.8 et les valeurs réellement observées sont disponibles dans la table 2.4 pour les différents modèles. Globalement, les estimations de provisions avec les méthodes de Chain-Ladder (méthode sous-jacente de Merz & Wüthrich, pour rappel), bootstrap et GLM semblent les plus fiables alors que la méthode ECLRM dévie d'environ la moitié du niveau réel des réserves.

Année calendaire	Chain-Ladder	Bootstrap	GLM	ECLRM
2003	5%	5%	5%	-43%
2004	10%	10%	10%	-37%
2005	5%	6%	5%	-55%
2006	5%	5%	5%	-56%
2007	-1%	-1%	-1%	-37%
Moyenne	**4,8%**	**4,8%**	**4,8%**	**-45,6%**

TABLE 2.4 : Écarts relatifs des réserves estimées (à l'ultime) selon les différents modèles par rapport aux valeurs observées (en milliers €)

Paramètres USP obtenus avec les méthodes réglementaire 2, bootstrap à un an, GLM et ECLRM. Les paramètres USP sont donnés en table 2.5 et en figure 2.11. Des écarts relatifs des réserves plus faibles permettent de fiabiliser ces derniers pour les trois premières méthodes. La colonne « Merz & Wüthrich (Chain-Ladder) » correspond à la méthode 2 réglementaire des actes délégués. Les paramètres avec la méthode ECLRM sont plus erratiques. Comme expliqué dans la section 1.3.4,

le reliquat de recours restant dans le triangle des charges peut causer un biais significatif au calcul des USP, il est donc décidé de ne pas considérer les résultats de cette méthode crédibles. Enfin, comme expliqué en section 1.2.4, l'EIOPA impose de conserver le paramètre le plus élevé parmi les deux méthodes réglementaires des actes délégués. Par exemple, pour l'exercice 2019, le paramètre USP est de 15,52%. Ainsi, parmi les méthodes 2, l'approche standard de Merz & Wüthrich, du bootstrap à un an et du modèle linéaire généralisé donnent des résulats plus ou moins équivalents et plus précis que la méthode ECLRM. Les paramètres USP estimés sont supérieurs au coefficient de 9%, fixé par la formule standard. Ces coefficients plus élevés peuvent s'expliquer s'expliquer du fait de la présence de sinistres corporels graves, sinistres difficilement supportables par une captive et nécessitant une charge en capital à détenir plus élevée.

Année calendaire	Merz & Wüthrich (Chain-Ladder)	GLM	Bootstrap	ECLRM
2004	20,22%	14,45%	14,82%	23,10%
2005	16,90%	12,04%	12,27%	25,25%
2006	14,48%	11,68%	12,29%	25,32%
2007	12,71%	11,16%	11,23%	25,89%
2008	9,50%	14,45%	17,20%	17,93%
2009	10,39%	15,13%	17,71%	20,48%
2010	10,88%	15,70%	18,19%	28,39%
2011	11,55%	13,31%	18,21%	22,82%
2012	12,30%	13,29%	21,23%	24,65%
2013	13,90%	13,45%	21,81%	24,31%
2014	16,81%	12,20%	16,39%	20,67%
2015	15,76%	11,79%	14,76%	23,18%
2016	16,53%	12,98%	15,03%	13,55%
2017	14,83%	12,88%	15,16%	15,82%
2018	14,31%	11,81%	14,08%	16,94%
2019	12,86%	10,56%	11,86%	16,70%

TABLE 2.5 : Comparaison des méthodes USP pour les années 2004 à 2019 par rapport à la méthode réglementaire 2

Les méthodes bootstrap à un an, GLM et ECLRM sont ainsi des alternatives actuelles à la méthode des actes délégués (Merz et Wüthrich, 2008). Un autre modèle (Boumezoued et Devineau, 2011), non présenté dans le présent travail, incluant un facteur de queue (peu crédible ici car l'historique est jugé suffisement profond) aurait pu être également une possibilité à exploiter. Concernant le modèle généralisé Poisson, d'autres travaux sur le provisionnement existent, selon des approches log-normale (Kremer, 1982), Gamma (Mack, 1991a), ou encore binomiale négative (Verrall, 2000). Cependant, ces dernières ne proposent pas de formalisme pour calculer la MSEP à un an. Un avantage de la méthode de *re-reserving* est que, basée sur des simulations Monte-Carlo, elle permet d'obtenir une distribution du CDR. Comme la version générale du bootstrap, obtenir la distribution permet d'obtenir les quantiles, notamment la VaR, qui a toute son utilité dans un cadre Solvabilité II. Cette méthode est toutefois lourde en temps de calcul. Au-delà du calcul de paramètres USP, d'autres modèles plus poussés comme le modèle interne partiel bayésien (décrit en annexe A.9) existent également sur le marché. Le choix restreint de distributions paramétriques dans la méthode 1 et la diversité des approches étudiées comme alternatives à la méthode 2 permettent légitimement de se focaliser sur de potentielles améliorations à apporter à la méthode 2, dans le chapitre 3.

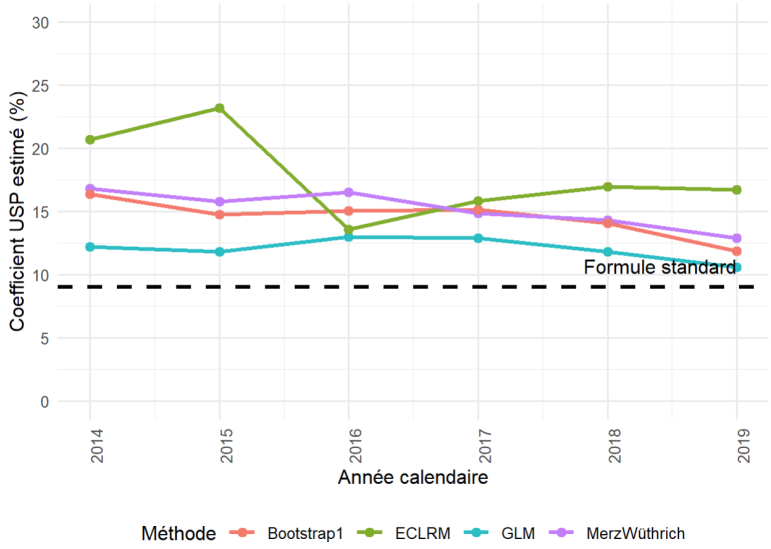


Figure 2.11 : Paramètres USP estimés avec les méthodes alternatives à la méthode 2

Ces dernières années, des techniques de *Machine Learning*, plus précisément d'apprentissage profond (*Deep Learning*) appliquées au provisionnement non-vie semblent gagner en popularité. A titre d'exemple, une extension du modèle linéaire généralisé, présenté dans la section 2.2.2, embarque le modèle dans un réseau de neurones (Gabrielli, 2019). Il convient donc de se demander à présent dans le chapitre 3 s'il est possible d'exploiter de telles techniques d'apprentissage dans le contexte des calculs du paramètre USP de la captive.

Chapitre 3

Vers une version *Deep Learning* concurrente à Mack Chain-Ladder pour le calcul de l'USP

Dans la pratique, la modèle de Chain-Ladder reste la référence chez les assureurs en matière de provisionnement. Cependant, les récentes techniques de *Machine Learning* appliquées au provisionnement émergent ces dernières années. Avec une collecte de données de plus en plus importante, les actuaires réussissent à provisionner les sinistres à la maille individuelle. Malgré l'émergence des modèles de provisionnement ligne à ligne, le présent travail étudie les modèles de *Machine Learning* basés sur des triangles de liquidation (pour les raisons évoquées dans la section 1.3.3). Il convient dans ce chapitre de s'intéresser à ces récentes évolutions, en étudiant en particulier un modèle de *Deep Learning* construit sur des réseaux de neurones récurrents (RNN), et en comparant ses performances à des modèles déjà employés sur le marché. Selon ces performances, le modèle choisi pourrait avoir des répercussions sur le paramètre USP nouvellement calculé et ainsi sur la charge en capital allouée au risque de réserve de la captive.

3.1 Le choix du modèle de provisionnement

Motivation et palette des modèles de marché disponibles. Un récent état des lieux de la diversité des modèles de provisionnement non-vie disponibles sur le marché (Titon et Talbi, 2024) permet de distinguer notamment, au-delà des méthodes individuelles (basées sur les sinistres ligne à ligne) et agrégées (basées sur les triangles de liquidation), les modèles paramétriques des modèles non-paramétriques (beaucoup plus flexibles, comme les arbres de regression (Wüthrich, 2018), les réseaux de neurones (Taylor, 2019). Parmi cette dernière classe de modèles, une piste est intéressante : « *another way to look at past loss histories is to use recurrent neural networks (RNNs), a very popular class of NNs introduced by (J. J. Hopfield, 1982). S. Hochreiter and J. Schmidhuber, 1997 introduced long short-term memory (LSTM) networks, a class of RNNs, to avoid gradient explosion* ». L'absence d'utilisation des données ligne à ligne plus précises peut même dans certains cas finalement être comblée, puisque le *white paper* conclut son étude par : « *In conclusion, at this stage, we have no proof that individual models are better, in terms of best estimate, than conventional aggregate methods. Moreover, their implementation may prove more complex* ».

Pour surmonter les potentielles contraintes liées au manque de données ligne à ligne ou à des hypothèses de lois parfois trop contraignantes, et en vue de rester fidèle aux hypothèses 1.2.3 de la

méthode réglementaire 2, il est opportun d'étudier une méthode combinant à la fois les hypothèses 1.2.1 de Mack, sans supposer d'hypothèses contraignantes de distribution en amont sur les montants, et en exploitant le pouvoir prédictif du *Deep Learning* via l'utilisation de réseaux de neurones récurrents. Le modèle qui va être présenté innove en la matière parmi les modèles de provisionnement évoqués. Il propose le calcul de nouveaux facteurs de développement et de coefficients de volatilité, qui vont pouvoir être incorporés dans une équation (1.9) « améliorée », potentiellement plus robuste. Enfin, au-delà de l'approche déterministe de Chain-Ladder, la méthode présentée par la suite offre une distribution des réserves via des simulations de bootstrap, utiles à des fins réglementaires.

Le modèle Mack-Net, un modèle de *Deep Learning*. Le modèle Mack-Net (Ramos-Pérez et al., 2022) est disponible via le *package* `MackNet` de R core team (2023) et développé par Ramos-Pérez (2024). D'autres méthodes de la même famille que le modèle Mack-Net basées sur des réseaux de neurones existent, comme Wüthrich (2018), Baudry et Robert (2019) mais ces méthodes exploitent, à nouveau, certaines covariables des données ligne à ligne. Comme évoqué en fin de chapitre 2, Gabrielli (2019) et Gabrielli et al. (2019) proposent également des modèles avec réseaux de neurones embarquant le modèle GLM décrit en section 2.2.2. La méthode Mack-Net ne suppose quant à elle aucune distribution sous-jacente pour les montants de sinistres. L'idée qui sous-tend ce choix des auteurs provient du fait que le changement de comportements des assurés, des programmes de réassurance, de la réglementation ou encore le nombre d'assurés peut, d'une année à l'autre, modifier considérablement les distributions des montants.

Enfin, un modèle *DeepTriangle* basé sur les RNN existe aussi sur le marché (Kuo, 2018). L'approche est seulement appliquée aux données de l'auteur. Le mémoire de Goma (2020), dont la contribution finale consiste en la création d'une librairie (non-rendue publique, généralisant à d'autres données la solution proposée par l'auteur) permet de mieux comprendre ce modèle. Enfin, dans le cadre d'une problématique sur le calcul du paramètre USP, ce modèle ne propose aucune formule généralisée de facteurs de développement ou de variance comme dans le modèle Mack-Net, ce qui apporte à ce dernier une valeur ajoutée pertinente. Comme évoqué en section 3.5.2, le modèle *DeepTriangle* est seulement exploitable sur une profondeur de dix ans ce qui le rend moins légitime comme modèle de marché de comparaison.

Ainsi, en se basant sur une approche novatrice pouvant rivaliser avec l'approche traditionnelle de Mack Chain-Ladder, comme précisé dans le *white paper*, et tout en respectant les hypothèses sous-jacentes souhaitées pour le calcul du paramètre USP dans la méthode réglementaire 2, le modèle Mack-Net vaut la peine d'être exploré.

Une fois le modèle sélectionné et justifié, il convient de faire un rappel théorique sur l'architecture des réseaux de neurones ainsi que de leur version avec RNN.

3.2 Utilisation des réseaux de neurones et réseaux de neurones récurrents (RNN)

3.2.1 Architecture d'un réseau de neurones

Les réseaux de neurones (notés aussi NN) représentent depuis quelques années des approches incontournables pour capturer les comportements dans les jeux de données de grande taille et l'apprentissage de fonctions non-linéaires ou non triviales.

Un neurone. En considérant

$$\mathbf{x} = (x_1, \dots, x_p) \in \mathbb{R}^p$$

une observation à p *features*,

$$\mathbf{w} = (w_1, \dots, w_p) \in \mathbb{R}^p$$

un vecteur de pondération de l'observation, b l'intercept du réseau, z désigne la fonction de pré-activation (schématisée en vert sur chaque neurone des figures 3.1 et 3.3) définie par

$$z(\mathbf{x}) = \sum_{j=1}^{p} w_j x_j + b = \mathbf{w}^\top \mathbf{x} + b,$$

et a désigne la fonction de post-activation (schématisée en rouge sur chaque neurone des figures 3.1 et 3.3) définie par

$$a(\mathbf{x}) = \phi(z(\mathbf{x})),$$

où ϕ est un hyperparamètre correspondant à une fonction d'activation (souvent non-linéaire).

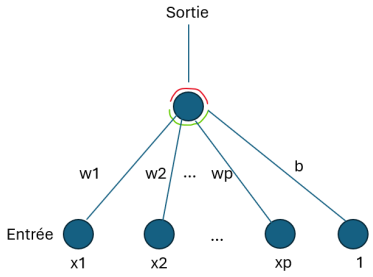


FIGURE 3.1 : Un neurone

Il existe de nombreuses fonctions d'activation, comme dans la figure 3.2, parmi elles :

- la fonction identité $\phi(x) = x$,
- la fonction *ReLU* $\phi(x) = x^+ = \max(x, 0)$,
- la fonction sigmoïde $\phi(x) = \frac{1}{1+\exp(-x)}$,
- la tangente hyperbolique $\phi(x) = \tanh(x) = \frac{\exp(x)-\exp(-x)}{\exp(x)+\exp(-x)}$.

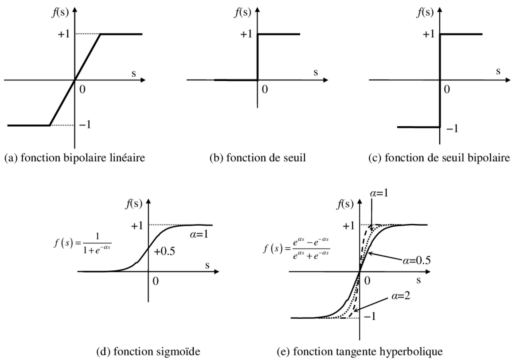


FIGURE 3.2 : Fonctions d'activation usuelles (BECHOUCHE, 2013)

Un réseau de neurones. Un réseau de neurones, décrit dans la figure 3.3, est une fonction

$$NN : \mathbb{R}^{n_1} \to \mathbb{R}^{n_L}.$$

Dans le cas d'une régression, $n_L = 1$ et la fonction de prédiction s'écrit

$$NN(\mathbf{x}) = \phi^{[L]} \circ z^{[L]} \circ \phi^{[L-1]} \circ z^{[L-1]} \circ \cdots \circ \phi^{[1]} \circ z^{[1]}(\mathbf{x}).$$

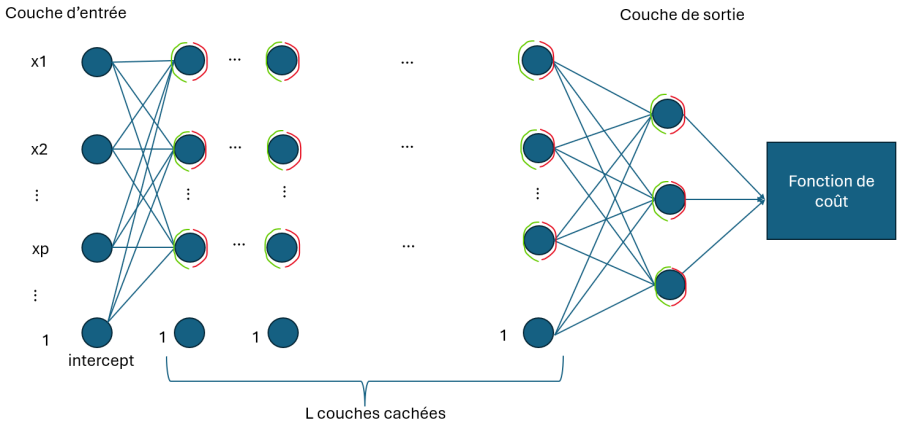


FIGURE 3.3 : Un réseau de neurones

L correspond au nombre de couches cachées, dont chacune possède une matrice de poids

$$\mathbf{W^{[l]}} = (w_{i,j}^{[l]}) \in \mathbb{R}^{n_l \times n_{l-1}},$$

un intercept

$$\mathbf{b^{[l]}} \in \mathbb{R}^{n_l},$$

n_l désignant le nombre de neurones dans chaque couche. Le nombre de couches cachées correspond à la profondeur du réseau, n_l la largeur de la couche l (correspondant à deux hyperparamètres), et $w_{i,j}^{[l]}$ est associé au poids accordé à la connexion entre le neurone $x_i^{[l-1]}$ et $x_j^{[l]}$ de deux couches successives. La i-ème pré-activation de la l-ème couche s'écrit

$$z^{[l]}(\mathbf{x})_i = \sum_{j=1}^{p} w_{ij}^{[l]} a^{([l-1]}(\mathbf{x})_j + b_i^{[l]}, \quad 1 \leq i \leq n,$$

où n correspond au nombre de neurones dans chaque couche cachée, alors que la i-ème post-activation est

$$a^{[l]}(\mathbf{x})_i = \phi^{[l]}(z^{[l]}(\mathbf{x})_i),$$

avec :

- $w_{ij}^{[l]}$ est le poids de $a^{[l-1]}(\mathbf{x})_j$ dans la somme pondérée $z^{[l]}(\mathbf{x})_i$ (notons $a^{[0]}(\mathbf{x})_j = x_j$, $1 \leq j \leq p$),
- $b_i^{[l]}$ est l'intercept (biais) dans la somme pondérée de $z^{[l]}$,
- $\phi^{[l]}$ est la fonction d'activation de la l-ème couche. Souvent : $\phi^{[l]} = \phi$, $1 \leq l \leq L$.

Par la suite, on note $\mathbf{z^{[l]}}(\mathbf{x}) = (z^{[l]}(\mathbf{x})_1, \dots, z^{[l]}(\mathbf{x})_n)^\top$, on peut donc écrire que

$$\mathbf{z^{[l]}}(\mathbf{x}) = \mathbf{W^{[l]}}\mathbf{a^{[l-1]}}(\mathbf{x}) + \mathbf{b^{[l]}},$$

avec

$$W^{[l]} = \begin{pmatrix} w_{1,1}^{[l]} & \cdots & w_{1,n}^{[l]} \\ \vdots & \ddots & \vdots \\ w_{n,1}^{[l]} & \cdots & w_{n,n}^{[l]} \end{pmatrix}, \quad 2 \leq l \leq L.$$

Il faut remarquer que pour la première couche, $\mathbf{W^{[1]}}$ est de taille $n \times p$. Enfin, pour un problème de prédiction, la couche de sortie contient un unique neurone. La sortie du réseau peut être positive (avec la fonction $ReLU$ par exemple), négative ou positive (avec la fonction identité par exemple), etc. selon l'espace d'arrivée de la fonction d'activation souhaité.

La fonction de coût du réseau. La fonction de coût (de perte, ou d'erreur)

$$\mathcal{L} : \mathbb{R}^{n_L} \times \mathbb{R}^{n_L} \to \mathbb{R}.$$

permet de mesurer la qualité d'une valeur prédite par le réseau par rapport à la valeur réellement observée.

Le risque d'un modèle h qui en découle correspond à l'espérance de la fonction de coût

$$\mathcal{R}(h) = \mathbb{E}(\mathcal{L}(y, NN(\mathbf{x}))).$$

Le modèle recherché NN est tel que

$$NN \in \arg\min_{h \in \mathcal{F}} \mathbb{E}(\mathcal{L}(y, h(\mathbf{x}))).$$

où $\mathcal{F}$ désigne un sous-espace de fonctions de $\mathbb{R}^p$ à valeurs dans $\mathbb{R}^{n_L}$. Pour une régression, la fonction de coût souvent utilisée est $\mathcal{L}(y, \hat{y}) = (y - \hat{y})^2$ ou une variante. D'une manière générale, la fonction de perte est souvent convexe pour simplifier la résolution du problème d'optimisation. En pratique, ce problème s'appelle la minimisation du risque empirique.

Le théorème d'approximation universelle. Ce théorème (d'existence) énonce que toute fonction continue peut être approchée avec un degré de précision illimité par un réseau de neurones. Mathématiquement, cela signifie que l'ensemble des fonctions de prédiction NN est dense pour une classe de fonctions particulières. Deux exemples de théorèmes émanant de l'approximation universelle sont proposés dans le théorème 1 et 2 et issus respectivement de Pinkus (1999) et Kidger et Lyons (2020).

La fonction de prédiction NN du réseau de neurones se modélise ainsi comme

$$\text{Entrée} : \mathbf{x} \xrightarrow{\mathbf{W^{[1]}x+b^{[1]}}} \mathbf{z^{[1]}}(\mathbf{x}) \xrightarrow{\phi^{[1]}(\cdot)} \mathbf{a^{[1]}}(\mathbf{x}) \to \cdots \to \mathbf{z^{[L+1]}}(\mathbf{x}) \xrightarrow{\phi^{[L+1]}(\cdot)} \mathbf{a^{[L+1]}}(\mathbf{x}) \to \text{Sortie},$$

où $(\mathbf{W}, \mathbf{b}) = (\mathbf{W^{[1]}}, \mathbf{b^{[1]}}, \dots, \mathbf{W^{[L]}}, \mathbf{b^{[L]}}, \mathbf{W^{[L+1]}}, \mathbf{b^{[L+1]}})$ sont les paramètres du réseau.

Théorème 1. ***Largeur arbitraire du réseau.***

Soit $K \subset \mathbb{R}^p$ un compact. On note $\mathcal{C}(K, \mathbb{R}^d)$ l'ensemble des fonctions continues de K dans $\mathbb{R}^d$ et

$$\mathcal{L}^q(\mathbb{R}^p, \mathbb{R}^d) := \left\{ f : \mathbb{R}^p \to \mathbb{R}^d \ t.q. \ \int_{\mathbb{R}^p} |f(\mathbf{x})|^q \, dx < +\infty \right\}, \quad q \geq 1,$$

avec $|\cdot|$ la norme euclidienne dans $\mathbb{R}^p$. L'ensemble des fonctions de prédiction pour des réseaux à une seule couche cachée, noté

$$G := \left\{ g(\mathbf{x}) := \sum_{i=1}^{n} w_i^{[2]} \phi \left(\sum_{j=1}^{p} w_{i,j}^{[1]} x_j + b_i^{[1]} \right) \middle| n \in \mathbb{N}^*, w_{i,j}^{[1]}, b_i^{[1]}, w_i^{[2]} \in \mathbb{R}, 1 \leq i \leq n, 1 \leq j \leq p \right\},$$

est dense dans $\mathcal{C}(K, \mathbb{R})$, i.e., pour toute fonction $f \in \mathcal{C}(K, \mathbb{R})$, pour tout $\epsilon > 0$, il existe une fonction $g \in G$ telle que

$$\sup_{x \in K} |f(\mathbf{x}) - g(\mathbf{x})| < \epsilon.$$

Théorème 2. ***Profondeur L arbitraire du réseau.***

Soit l'ensemble des fonctions de prédiction $\mathcal{F}_{p,m,k}^{\phi}$ d'un réseau de neurones tel que :

- *la couche d'entrée et de sortie sont constituées respectivement de p et m neurones,*
- *chaque couche cachée est constituée de k neurones et sa fonction d'activation est définie comme $\phi : \mathbb{R} \to \mathbb{R}$,*
- *tous les neurones de la couche de sortie ont la fonction d'activation d'identité.*

Alors,

1. *si $K \subset \mathbb{R}^p$ est compact, si ϕ est continue, non-affine et différentiable en au moins une valeur telle que sa dérivée est non-nulle, alors $\mathcal{F}_{p,m,p+m+2}^{\phi}$ est dense dans $\mathcal{C}(K, \mathbb{R}^m)$,*
2. *si ϕ correspond à la fonction ReLU, alors $\mathcal{F}_{p,m,p+m+1}^{\phi}$ est dense dans $\mathcal{L}^q(\mathbb{R}^p, \mathbb{R}^m)$ par rapport à la norme $\mathcal{L}^q$.*

Mécanisme de *backpropagation*. Comme le théorème d'approximation universelle ne prouve seulement que l'existence d'un réseau de neurones optimal, la rétropropagation du gradient (*backpropagation*) permet quant à elle d'entraîner un réseau de neurones pour obtenir les paramètres du réseau $(\mathbf{W}, \mathbf{b}) = (\mathbf{W^{[1]}}, \mathbf{b^{[1]}}, \ldots, \mathbf{W^{[L]}}, \mathbf{b^{[L]}}, \mathbf{W^{[L+1]}}, \mathbf{b^{[L+1]}})$. Elle correspond à la minimisation du risque empirique.

Soit $\widehat{NN}^{(\mathbf{W},\mathbf{b})}(\cdot) = \hat{y}$ la fonction de prédiction,

$$\mathcal{R}^{\mathcal{D}_N}(\mathbf{W}, \mathbf{b}) = \sum_{i=1}^{n} \mathcal{L}\left(y_i, \widehat{NN}^{(\mathbf{W},\mathbf{b})}(x_i)\right),$$

avec $\mathcal{D}_N$ l'ensemble des données d'entraînement.

Dans $\mathbb{R}^p$, l'algorithme consiste à trouver les paramètres minimisant $\mathcal{R}^{\mathcal{D}_N}(\mathbf{W}, \mathbf{b})$. Cela passe par le calcul de

$$\frac{\partial \mathcal{L}(y, \hat{y})}{\partial w_{ij}^{[l]}} \quad \text{et} \quad \frac{\partial \mathcal{L}(y, \hat{y})}{\partial b_i^{[l]}}.$$

Il s'agit d'une méthode *backward* : après la propagation en avant des données dans le réseau, on calcule les dérivées de la couche l avant celles de $l-1$. $\hat{y}$ étant une composition de fonction, la règle de la chaîne intervient dans les calculs $(f \circ g)'(x) = f'(g(x))g'(x)$.
L'algorithme 3 décrit la procédure de *backpropagation* par descente de gradient stochastique (James et al., 2013) de $\mathcal{R}^{\mathcal{D}_N}(\mathbf{W}, \mathbf{b})$. Par exemple, un critère d'arrêt peut être un nombre d'itérations maximum ou un arrêt dans la décroissance dans la fonction de coût $\mathcal{L}$ au fil des itérations. Les dérivations en cascade sont illustrées dans la figure 3.4 (dans notre cas, en relation avec la figure, $m = 1$).

Algorithme 3 Algorithme de *backpropagation* par descente de gradient stochastique

Entrée : Ensemble d'entraînement $\mathcal{D}_N = \{(x_i, y_i), 1 \leq i \leq n\}$, ensemble de test $\mathcal{T}_K$, taux d'apprentissage η, taille de mini-batch m, structure du réseau (L, n, ϕ), initialisation des paramètres $(\mathbf{W}, \mathbf{b}) = \{\mathbf{W^{[1]}}, \mathbf{b^{[1]}}, \ldots, \mathbf{W^{[L+1]}}, \mathbf{b^{[L+1]}}\}$
while un critère d'arrêt n'est pas atteint **do**
  Sélection d'un mini-batch $D_m \subset \mathcal{D}_N$ de taille m
  Calcul *forward*
  for $(x_i, y_i) \in D_m$ **do**
    Calcul des activations $z^{[l]}(x_i), a^{[l]}(x_i)$ pour $1 \leq l \leq L+1$
  end for
  Calcul *backward*
  for $(x_i, y_i) \in D_m$ **do**
    Calcul des dérivées $\frac{\partial \mathcal{L}(y_i, \hat{y}_i)}{\partial W^{[l]}}, \frac{\partial \mathcal{L}(y_i, \hat{y}_i)}{\partial b^{[l]}}$ pour $1 \leq l \leq L+1$
    où $\hat{y}_i = \hat{NN}^{(\mathbf{W},\mathbf{b})}(x_i)$
  end for
  Mise à jour des paramètres
  for $l = 1$ to $L+1$ **do**
    $\mathbf{W^{[l]}} \leftarrow \mathbf{W^{[l]}} - \eta \sum_{i=1}^{m} \frac{\partial \mathcal{L}(y_i, \hat{y}_i)}{\partial \mathbf{W^{[l]}}}$
    $\mathbf{b^{[l]}} \leftarrow \mathbf{b^{[l]}} - \eta \sum_{i=1}^{m} \frac{\partial \mathcal{L}(y_i, \hat{y}_i)}{\partial \mathbf{b^{[l]}}}$
  end for
end while
***Output* :** Les paramètres $(\mathbf{W}, \mathbf{b})$

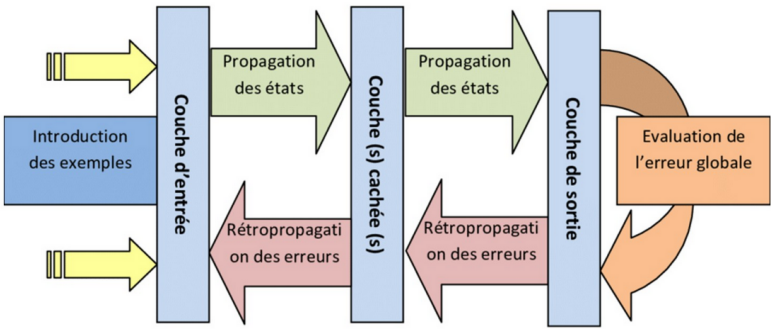


FIGURE 3.4 : Dérivations en cascade lors de la *backpropagation* (DJUIKEM, 2024)

Le problème des *vanishing gradients*. Ce problème est rencontré lors de l'apprentissage du réseau de neurones. Lorsque des réseaux de neurones complexes sont étudiés, le gradient calculé lors de la *backpropagation* pour les premières couches du réseau risque de disparaître, c'est-à-dire qu'à mesure que l'algorithme remonte dans le réseau, le gradient tend à diminuer. Lorsque la fonction d'activation est la fonction sigmoïde ou tanh, ce problème est prépondérant (contrairement à la fonction $ReLU$), car le gradient diminue exponentiellement. Cela peut engendrer un arrêt dans la mise à jour des poids et des temps de calculs beaucoup trop élevés car les couches antérieures sont entrainées beaucoup moins rapidement que les couches ultérieures. Les RNN et cellules $LSTM$ permettent notamment de contourner ce problème.

L'algorithme ADAM. Cet algorithme (ROYER, 2022) est une variante de l'algorithme 3. Il s'agit d'une méthode de réduction de variance pour les gradients stochastiques en prenant en compte des

paquets (*batch*) de manière simultanée. On définit une époque lorsque le réseau traite tous les *batch* qui composent l'ensemble des données. Concrètement, une époque équivaut à un passage complet à travers les données d'entraînement en vue d'un meilleur ajustement des poids. Pour rappel, la k-ième itération dans l'algorithme de descente de gradient est de la forme

$$\mathbf{W}^{[\mathbf{k+1}]} = \mathbf{W}^{[\mathbf{k}]} - \eta_k \nabla \mathcal{R}^{\mathcal{D}_N}(\mathbf{W}, \mathbf{b})$$

En notant $\mathbf{W}^{[\mathbf{k+1}]} = \mathbf{W}^{[\mathbf{k}]} - \eta \mathbf{g_k}$ où η correspond au taux d'apprentissage, et $\mathbf{g_k}$ un estimateur stochastique du gradient associé à un *batch* d'indices, on décompose l'équation avec le schéma suivant

$$\mathbf{W}^{[\mathbf{k+1}]} = \mathbf{W}^{[\mathbf{k}]} - \eta_k \mathbf{m_k} \odot \mathbf{v_k}.$$

Le pseudo-code de l'algorithme 4 est issu de Kingma et Ba (2017). Tel que formulé dans l'article, il est fourni ci-dessous de manière générique.

Algorithme 4 Algorithme ADAM

Entrée: α : Pas de l'apprentissage
Entrée: ϵ : Terme de régularisation pour éviter les divisions par zéro (exemple : $\epsilon = 10^{-8}$)
Entrée: $\beta_1, \beta_2 \in [0, 1)$: Taux de dégradation exponentielle pour les estimations des moments
Entrée: $f(\theta)$: Fonction objectif stochastique avec les paramètres θ
Entrée: $\theta_\mathbf{0}$: Vecteur des paramètres initiaux

$\mathbf{m_0} \leftarrow 0$ ▷ Initialisation du premier vecteur de moment
$\mathbf{v_0} \leftarrow 0$ ▷ Initialisation du second vecteur de moment
$t \leftarrow 0$
while θ_t n'a pas convergé **do**
  $t \leftarrow t + 1$
  $\mathbf{g_t} \leftarrow \nabla f_t(\theta_{t-1})$ ▷ Calcul des gradients à l'étape t
  $\mathbf{m_t} \leftarrow \beta_1 \cdot \mathbf{m_{t-1}} + (1 - \beta_1) \cdot \mathbf{g_t}$ ▷ Mise à jour du moment de premier ordre
  $\mathbf{v_t} \leftarrow \beta_2 \cdot \mathbf{v_{t-1}} + (1 - \beta_2) \cdot \mathbf{g_t}^2$ ▷ Mise à jour du moment de second ordre
  $\hat{\mathbf{m}}_\mathbf{t} \leftarrow \mathbf{m_t}/(1 - \beta_1^t)$ ▷ Correction du biais pour le premier moment
  $\hat{\mathbf{v}}_\mathbf{t} \leftarrow \mathbf{v_t}/(1 - \beta_2^t)$ ▷ Correction du biais pour le second moment
  $\theta_\mathbf{t} \leftarrow \theta_\mathbf{t-1} - \alpha \cdot \hat{\mathbf{m}}_\mathbf{t}/(\sqrt{\hat{\mathbf{v}}_\mathbf{t}} + \epsilon)$ ▷ Mise à jour des paramètres
end while
return $\theta_\mathbf{t}$

Le but de l'algorithme est de minimiser $\mathbb{E}(f(\theta))$ par rapport aux paramètres θ, avec $\mathbf{g_t}$ le vecteur des dérivées partielles de f_t les réalisations de la fonction objectif à chaque pas de temps. $\mathbf{m_t}$ représente la moyenne mobile du gradient et $\mathbf{v_t}$ la moyenne mobile du carré des gradients qui permet la mise à l'échelle des mises à jour avec prise en compte de la variance non-centrée des gradients. Les opérations sont toutes appliquées élément par élément. L'avantage d'ADAM est qu'il permet à la fois de faire face au problème de gradients nuls (*sparse gradients*) qui peuvent arriver notamment lorsque des techniques de régularisation sont appliquées aux données, mais aussi de se prémunir contre les problèmes de non-stationnarité, c'est-à-dire lorsque l'on observe des changements trop brusques du gradient. Cet algorithme est une référence dans le domaine du *Deep Learning*, car il donne plus d'importance aux dernières itérations de l'algorithme.

Méthode d'extinction et paramètre de *dropout*. La méthode d'extinction est une technique de régularisation stochastique, qui permet de limiter le surapprentissage (*overfitting*) du modèle. Comme présenté dans la figure 3.5, une couche d'extinction « éteint » certains neurones de manière aléatoire

pendant l'entraînement du modèle, afin de considérer l'information contenue dans le maximum de neurones. Le taux d'extinction (ou paramètre *dropout*) correspond à la probabilité pour chaque neurone d'être écarté de l'entraînement. Ainsi pour un réseau à n neurones, il existe 2^n combinaisons de réseaux possibles après application du taux d'extinction.

Mathématiquement, la méthode d'extinction consiste à ajouter du bruit aux couches cachées. Avec le paramètre *dropout* p, les fonctions pré/post-activation s'écrivent

$$r_i^{(l)} \sim \mathcal{B}(p),$$
$$\tilde{a}_i^{[l]} = r_i^{[l]} \odot a_i^{[l]},$$
$$z_i^{[l+1]} = w_i^{[l+1]} \tilde{a}_i^{[l]} + b_i^{[l+1]},$$
$$a_i^{[l+1]} = \phi(z_i^{[l+1]}).$$

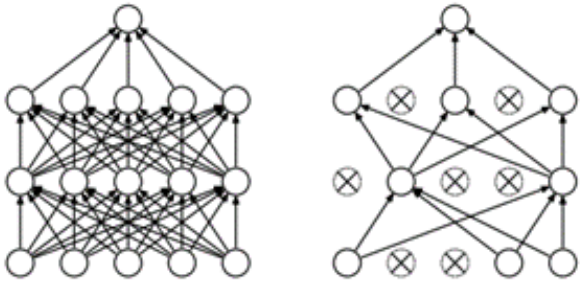


FIGURE 3.5 : Amincissement d'un réseau avec taux d'extinction (BUDHIRAJA, 2016)

3.2.2 Les RNN et les cellules LSTM

Les réseaux de neurones récurrents (*RNN*) sont une classe de réseaux de neurones dont l'information dépend du temps (séquence, ou série temporelle). Ces séquences, comme dans la figure 3.6, reposent sur l'ordre des données qui se suivent dans le temps (CNRS, 2022). Chaque état d'une séquence n'a de sens que par rapport à l'état qui le précède (si on mélange les éléments d'une séquence celle-ci n'a plus vraiment de sens).

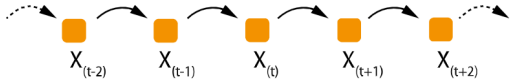


FIGURE 3.6 : Une séquence (CNRS, 2022)

Il existe différents types de RNN, décrits dans la figure 3.7 : les RNN « séquence à séquence »(l'entrée et de sortie du réseau sont de type séquence), les RNN « séquence à vecteur »(l'entrée et la sortie du réseau sont respectivement de type séquence et vecteur, etc.), les RNN « encodeur-décodeur » et les RNN « vecteur à série ». Par exemple, l'annotation d'image par un réseau récurrent consiste à capturer en entrée du réseau une image, le réseau constitue ensuite un vecteur des caractéristiques de l'image puis génère en sortie une légende descriptive de l'image. Les RNN encodeur-décodeur désignent enfin des réseaux série vers série décalées dans le temps (ces types de réseaux sont très utiles par exemple pour la traduction automatique d'un texte dans son ensemble, plutôt qu'une traduction mot à mot). Les RNN présentés en section 3.3 sont de type séquence à vecteur.

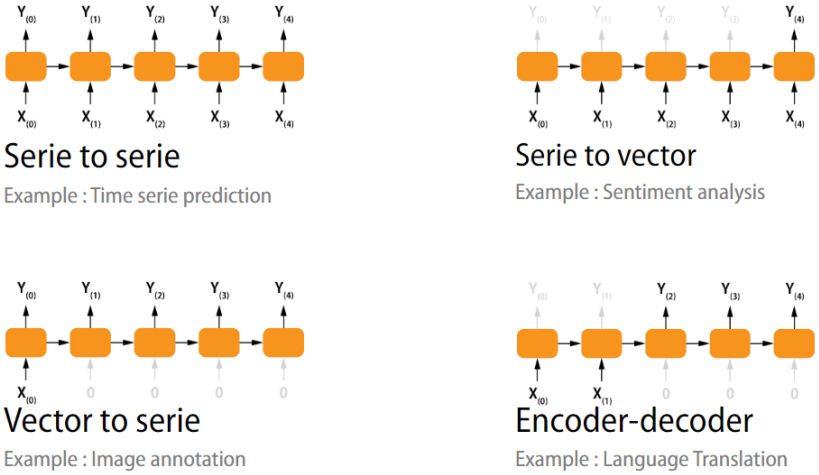


FIGURE 3.7 : Différents types de RNN (CNRS, 2022)

Un neurone récurrent. Les architectures d'un neurone classique par rapport à un neurone récurrent sont celles décrites en figure 3.8, où w_y, b et y sont des scalaires, et $\mathbf{W_x}$ un vecteur. L'observation $\mathbf{x}$ entre classiquement dans le neurone, l'activation se met en place puis à sa sortie boucle pour rentrer à nouveau dans le neurone.

(a) Neurone classique

(b) Neurone récurrent

FIGURE 3.8 : Comparaison d'un neurone classique et récurrent (CNRS, 2022)

Dans sa version dépliée (en figure 3.9), y_{t_1} correspond à la sortie du neurone de la première composante x_{t_1} de la séquence $\mathbf{x}$. Elle est réutilisée à l'itération suivante avec un poids (scalaire) w_y.

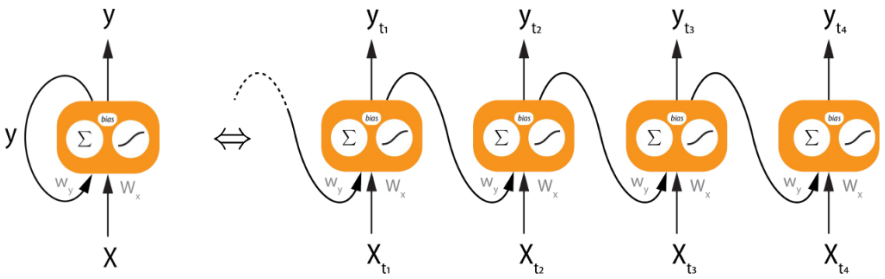
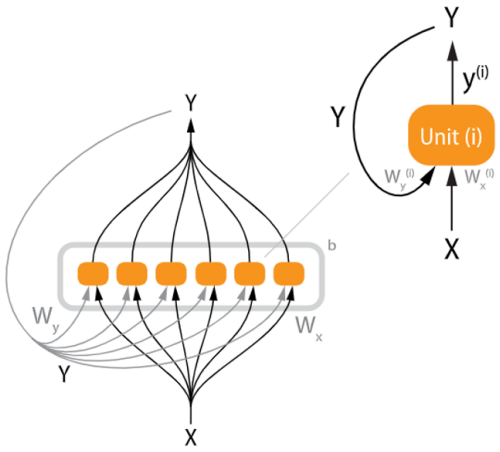
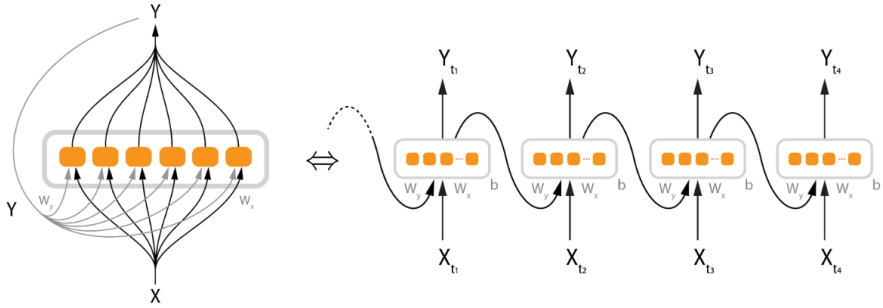


FIGURE 3.9 : Version dépliée d'un neurone récurrent (CNRS, 2022)

Une cellule RNN. Les neurones récurrents se regroupent ensuite de manière à constituer une cellule (couche) récurrente composée d'unités (neurones) comme en figure 3.10a et à agir « en escadrille ».



(a) Cellule récurrente composée d'unités



(b) Escadrille d'une cellule récurrente

FIGURE 3.10 : Une cellule récurrente (CNRS, 2022)

Dans la cellule récurrente, $\mathbf{W_x}$ et $\mathbf{W_y}$ désignent des tenseurs (des matrices), respectivement de taille nombre d'unités × dimension du vecteur d'entrée et nombre d'unités × nombre d'unités, et $\mathbf{b}$ et $\mathbf{Y}$ sont quant à eux des vecteurs. Le *RNN* offre enfin une flexibilité au niveau de la sortie en permettant

l'ajout d'un biais, et, formellement, on note l'équation suivante pour $\mathbf{Y_t}$

$$\mathbf{Y_t} = \phi(\mathbf{W_x}^T \cdot \mathbf{X_t} + \mathbf{W_y}^T \cdot \mathbf{Y_{t-1}} + \mathbf{b}). \tag{3.1}$$

De la même manière, le fonctionnement déplié d'une cellule récurrente est ainsi résumé comme en figure 3.11. Dans cet exemple, les entrées $\mathbf{X_{t_0}}$, $\mathbf{X_{t_1}}$ et $\mathbf{X_{t_2}}$ correspondent à trois vecteurs. Lorsque la première itération débute, la mémoire du *RNN* est initialisée au vecteur nul. À la sortie de la cellule, le vecteur $\mathbf{Y_{t_0}}$ est dupliqué, d'une part pour devenir la mémoire du *RNN* et d'autre part pour activer, lors de la seconde itération, la seconde entrée $\mathbf{X_{t_1}}$ (et ainsi de suite jusqu'à la fin de la troisième itération). Il est donc à noter que la cellule récurrente itère le même nombre de fois que le nombre de vecteurs d'entrée. Contre-intuitivement, bien que le nombre de vecteurs d'entrée et de sortie de la cellule soit identique, les vecteurs de sortie n'ont pas la même dimension que ceux de l'entrée. La taille des vecteurs de sortie dépend du nombre d'unités de la cellule récurrente (dans l'exemple, la cellule est composée de six unités, les vecteurs de sortie ont donc six composantes).

Les inconvénients des RNN. Les cellules récurrentes peuvent engendrer divers problèmes :

- une mémoire uniquement à court terme pour la prédiction, avec uniquement l'utilisation de $\mathbf{Y_{t-1}}$ pour prédire $\mathbf{Y_t}$ comme dans l'équation (3.1), ce qui peut provoquer l'oubli d'informations pertinentes notamment au sein de grandes séries temporelles,
- le problème de *vanishing gradients* évoqué en 3.2.1,
- une convergence faible, en conséquence.

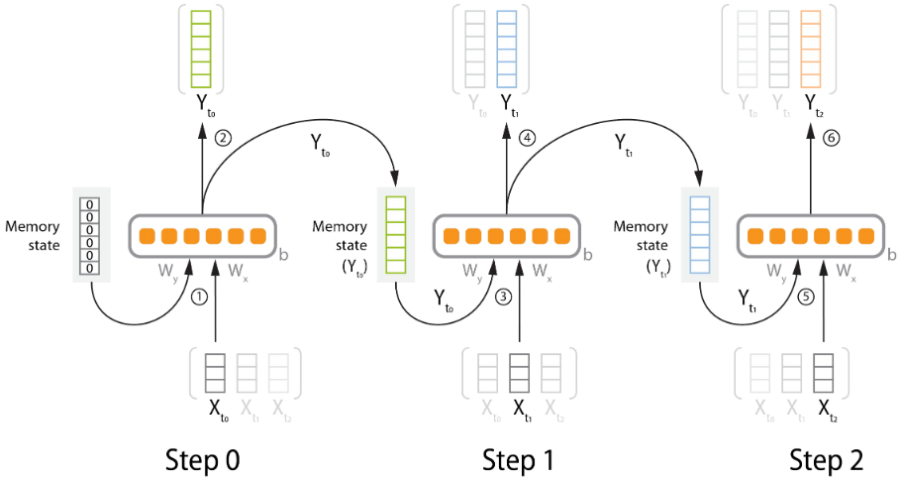


FIGURE 3.11 : Version dépliée d'une cellule d'un *RNN* (CNRS, 2022)

Pour résoudre le problème de la mémoire court terme, il existe un type de cellules récurrentes permettant la combinaison d'une mémoire immédiate et d'une mémoire de long terme : les cellules LSTM (*Long Short-Term Memory*).

Une cellule LSTM. Comme indiqué en figure 3.11, la mémoire courte (*memory state*, celle d'une cellule récurrente classique) est le résultat de l'itération précédente. Dans une cellule LSTM, la sortie de la cellule $\mathbf{Y_t}$ est dupliquée en deux exemplaires : d'une part conservée en mémoire court terme, la sortie est aussi destinée à alimenter la mémoire de long terme (*carry state*, conservée de manière

beaucoup plus étendue dans le temps, et qui disparaît beaucoup plus doucement). La mémoire de la cellule se constitue donc de deux états. Une cellule LSTM prend donc trois entrées : l'*input*, la mémoire court terme et la mémoire long terme. Pour piloter ces trois entrées, des « vannes »(*gate*, associées à des produits d'Hadamard, produits élément par élément) contrôlent le débit d'importance de chacune des deux mémoires et de l'*input*. Au total, quatre ensembles (sous-jacents à chaque cellule) de neurones pilotent ces vannes et l'*input* avec des poids, biais et fonction d'activation associées. Les fonctions d'activation, selon leur ensemble d'arrivée, permettent de transmettre, supprimer ou ajouter de l'information (par exemple, la fonction sigmoïde permet de normaliser la donnée entre 0 et 1 et d'extraire la pertinence de chaque information). Les portes jouent chacune un rôle différent :

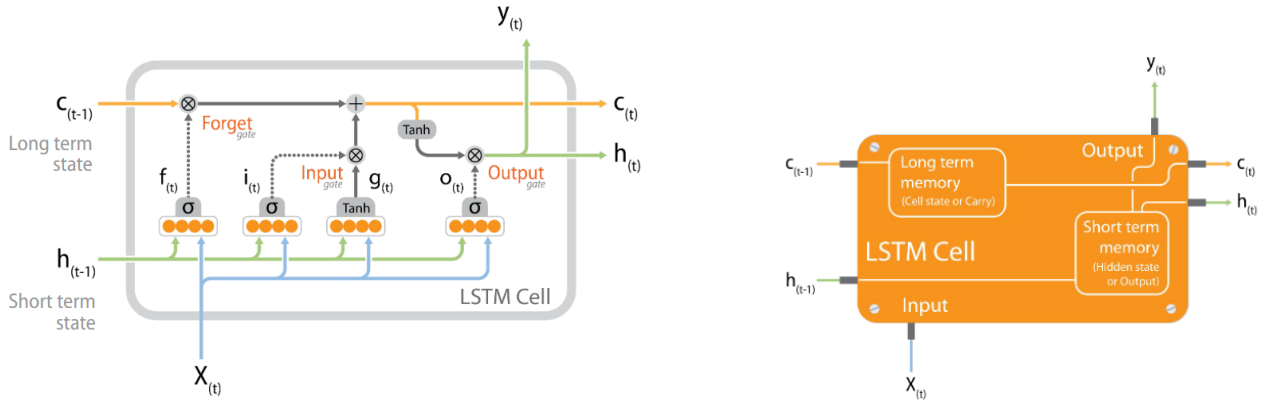
- une porte d'entrée (*input gate*) : son rôle est de décider des *inputs* à accepter au sein du neurone,
- une porte d'oubli (*forget gate*) : son rôle est de décider des informations qui peuvent être oubliées dans la mémoire,
- une porte de sortie (*output gate*) : son rôle est de permettre la sortie de la cellule LSTM et d'enrichir la mémoire de long terme $\mathbf{c_{t-1}}$.

Une version simplifiée avec moins de paramètres que les cellules LSTM existe également, les GRU (*Gated recurrent units*, exploitées notamment dans le modèle de l'annexe A.10). Elles comportent seulement deux portes :

- une porte de mise à jour (*update gate*) : son rôle est de filtrer si une nouvelle entrée doit être ajoutée à la mémoire cachée de la cellule ,
- une porte de réinitialisation (*reset gate*) : son rôle est de filtrer les informations pertinentes actuelles de l'état de la cellule.

L'avantage des GRU est que leur architecture et temps de calcul moins lourds permettent de réduire le risque de surapprentissage. Cependant, les cellules LSTM procurent généralement de meilleures performances.

Tout comme les cellules récurrentes standards, le nombre de composantes de chaque vecteur de sortie d'une couche LSTM correspond au nombre d'unités qui la composent, et le nombre de vecteurs de sortie correspond au nombre de séquences fournies en entrée (pour les réseaux « série vers vecteurs »). Pour ce type de RNN, seul le dernier vecteur importe, les autres vecteurs « intermédiaires » de sortie n'apporte nullement à la prédiction. Dans le cas d'un RNN « série vers série », la totalité des vecteurs (même ceux intermédiaires, c'est-à-dire toute la séquence) est exploitée.



(a) Circuit d'une cellule LSTM

(b) Composantes d'une cellule LSTM

FIGURE 3.12 : Cellule LSTM (CNRS, 2022)

Les notations de la figure 3.12 sont

$\mathbf{f_t} = \sigma(\mathbf{W_{xf}}^T\mathbf{X_t} + \mathbf{W_{hf}}^T\mathbf{h_{t-1}} + \mathbf{b_f})$,	vecteur d'activation de la porte d'oubli
$\mathbf{i_t} = \sigma(\mathbf{W_{xi}}^T\mathbf{X_t} + \mathbf{W_{hi}}^T\mathbf{h_{t-1}} + \mathbf{b_i})$,	vecteur d'activation de la porte d'entrée
$\tilde{\mathbf{c_t}} = \tanh(\mathbf{W_{xc}}^T\mathbf{X_t} + \mathbf{W_{hc}}^T\mathbf{h_{t-1}} + \mathbf{b_c})$,	vecteur d'entrée courant
$\mathbf{o_t} = \sigma(\mathbf{W_{xo}}^T\mathbf{X_t} + \mathbf{W_{ho}}^T\mathbf{h_{t-1}} + \mathbf{b_o})$,	vecteur d'activation de la porte de sortie
$\mathbf{c_t} = \mathbf{f_t} \odot \mathbf{c_{t-1}} + \mathbf{i_t} \odot \tilde{\mathbf{c_t}}$,	mémoire long terme (vecteur de l'état de la cellule)
$\mathbf{y_t} = \mathbf{h_t} = \mathbf{o_t} \odot \tanh(\mathbf{c_t})$,	mémoire court terme (état caché) ou vecteur de sortie
$\mathbf{X_t} \in \mathbb{R}^d$	vecteur d'entrée
$\odot$	produit d'Hadamard
σ	fonction sigmoïde
$\mathbf{W}$	matrice des poids
$\mathbf{b}$	vecteur de biais

L'interprétation de ces quantités est la suivante :

- $\mathbf{f_t}$ correspond à une opération d'oubli de l'information sachant l'état de la cellule et l'*input*,
- $\mathbf{i_t}$ joue un rôle de filtre pour l'ajout dans la mémoire actuelle $\mathbf{c_t}$,
- $\tilde{\mathbf{c_t}}$ correspond à une opération d'enrichissement de l'information sachant $\mathbf{X_t}$,
- $\mathbf{o_t}$ correspond au filtre de sélection dans la mémoire des informations pertinentes pour définir l'état de la cellule à l'instant t,
- $\mathbf{h_t}$ correspond à l'état de la cellule à l'instant t.

Parmi ces notations, $\mathbf{X}_t \in \mathbb{R}^d$ où d est la dimension de l'entrée, $\mathbf{f_t} \in \mathbb{R}^h$ où h représente la dimension de l'état caché et $\mathbf{W}_{xf} \in \mathbb{R}^{h\times d}$, $\mathbf{W}_{hf} \in \mathbb{R}^{h\times h}$, et $\mathbf{b}_f \in \mathbb{R}^h$. Enfin, $\mathbf{i}_t \in \mathbb{R}^h$, $\tilde{\mathbf{c}}_\mathbf{t} \in \mathbb{R}^h$, $\mathbf{o}_t \in \mathbb{R}^h$, $\mathbf{c}_t \in \mathbb{R}^h$, $\mathbf{h}_t \in \mathbb{R}^h$. Les produits d'Hadamard représentent les produits élément par élément, c'est-à-dire qu'entre deux vecteurs de même dimension, chaque composante des deux vecteurs sont multipliées entre elles.

Une autre différence entre une cellule LSTM et une cellule récurrente standard est que

$\mathbf{h_t} = \mathbf{o_t} \odot \tanh(\mathbf{c_t})$	dans une cellule LSTM,
$\mathbf{h_t} = \tanh(\mathbf{W_c}[\mathbf{h_{t-1}}, \mathbf{X_t}] + \mathbf{b_c})$	dans une cellule récurrente classique.

Ainsi, les cellules LSTM sont des solutions envisagées pour résoudre le problème des *vanishing gradients*, car seule l'information la plus appropriée est sélectionnée grâce à la mémoire de long terme durant la *backpropagation*. Elles sont disponibles via le *package* `Keras` (Allaire et Chollet, 2024). Ce *package* est notamment utilisé pour l'implémentation du modèle Mack-Net (Ramos-Pérez et al., 2022). Il convient par la suite de présenter le modèle et de voir que la construction d'une telle architecture de modèle est adaptée à la prédiction des réserves pour des triangles de liquidation, en considérant les montants de sinistres comme des séries temporelles au cours des années de développement.

3.3 Le modèle Mack-Net

3.3.1 Structure du modèle

A l'origine, Ramos-Pérez et al. (2022) utilisent la base de données de la NAIC (2015) (*National Association of Insurance Commissioners*, la même que dans le modèle de Kuo (2018), présenté en annexe A.10). Le modèle Mack-Net se base sur l'entraînement de *RNN* série vers série. Les données d'entraînement sont les triangles supérieurs hormis la dernière diagonale observée (comme sur la figure 3.15). La dernière diagonale observée correspond quant à elle aux données de validation. Les variables explicatives du modèle s'écrivent

$$\mathbf{X_1} = \left(Y^{*,Pa}_{i,j-1}, Y^{*,Pa}_{i,j-2}, \dots, Y^{*,Pa}_{i,j-F}\right) = \left(\frac{Y^{Pa}_{i,j-1}}{P_i}, \dots, \frac{Y^{Pa}_{i,j-F}}{P_i}\right),$$

$$\mathbf{X_2} = (DY^*_{j-1}, \dots, DY^*_{j-F}) = \left(\frac{DY_{j-1}}{I}, \dots, \frac{DY_{j-F}}{I}\right),$$

$$\mathbf{X_3} = (R^*_{j-1}, \dots, R^*_{j-F}) = \left(\frac{\sum_{i=0}^{I-j+2} D^*_{i,j-1}}{\sum_{i=0}^{I-j+2} \frac{C^{In}_{i,j-1}}{P_i}}, ..., \frac{\sum_{i=0}^{I-j+(I-1)} D^*_{i,j-F}}{\sum_{i=0}^{I-j+(I-1)} \frac{C^{In}_{i,j-F}}{P_i}}\right).$$

Ces trois variables forment des séries temporelles qui permettent de prédire le triangle inférieur de liquidation. La dimension de la matrice ($\mathbf{X_1}$, $\mathbf{X_2}$, $\mathbf{X_3}$) est T×3, où T correspond à la taille du triangle moins 2. F correspond au nombre d'années de survenance passées exploitées dans chaque réseau de neurones. Dans le modèle, F vaut (I-2). La variable réponse Z est définie par

$$Z_{i,j} = Y^{*,Pa}_{i,j} = \frac{Y^{Pa}_{i,j}}{P_i},$$

où, pour rappel (en reprenant les notations de la section 2.2.3) :

- $C^{Pa}_{i,j}$ représentent les paiements cumulés,
- $C^{In}_{i,j}$ représentent les montants d'*incurred* cumulés,
- P_i représente un vecteur d'exposition,
- $D^*_{i,j}$ vaut $\frac{C^{Pa}_{i,j}}{P_i}$,
- $Y^{Pa}_{i,j}$ représente les paiements incrémentaux,
- DY_j représente les années de développement.

Ainsi, $\mathbf{X_3}$ correspond aux ratios des paiements cumulés et des *incurred* selon le niveau d'exposition au risque pour l'année i. Les paiements et années de développement sont aussi normalisés, car l'exposition permet de mettre à la même échelle les cadences de règlements, les variations du volume d'activité pour toutes les années d'accident et ainsi faciliter l'apprentissage. La variable Z permet de prédire les paiements futurs.

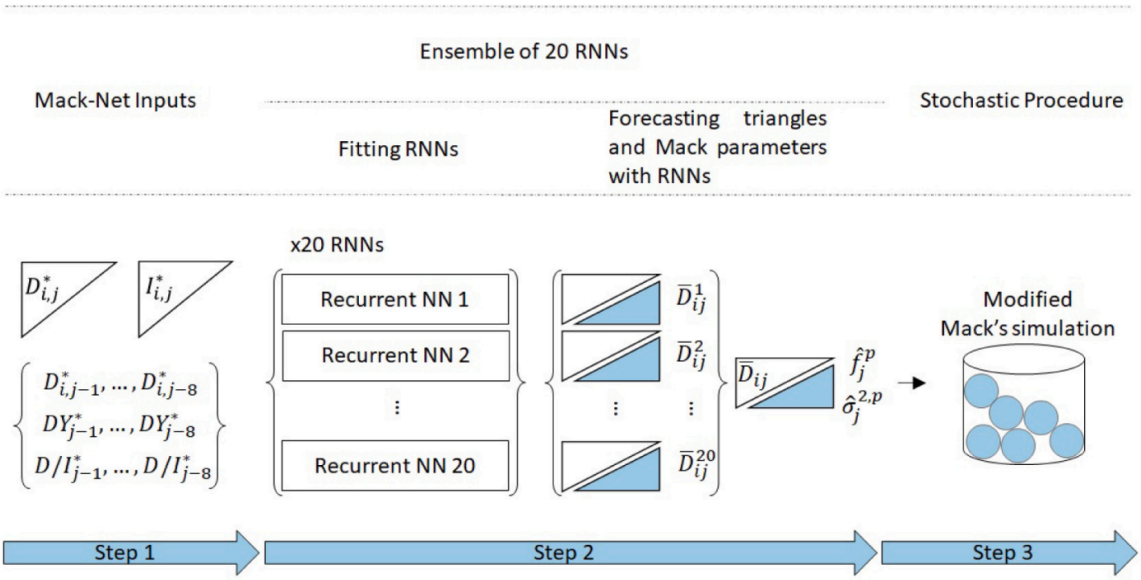


FIGURE 3.13 : Exemple de déroulement du modèle Mack-Net pour un triangle de provisionnement 10×10 (RAMOS-PÉREZ et al., 2022)

Hypothèse 3.3.1. *(a) Comme dans le cadre du calcul deu paramètre USP et de la méthode 2, le modèle Mack-Net suppose les hypothèses 1.2.1 de Mack vérifiées.*

D'après l'auteur (E. Ramos-Peréz), le fait que les hypothèses 1.2.1 ne soient pas vérifiées n'implique pas forcément de mauvaises prédictions des RNN (source : échanges par mails). Le but du modèle Mack-Net est de trouver une meilleure estimation des facteurs de développement et des coefficients de volatilité. Pour le reste, le modèle suit les mêmes approches que celles de MACK (1993) et ENGLAND et VERRALL (2002).

Comme décrit dans la figure 3.13, le modèle Mack-Net prend en entrée les séries des paiements ou des *incurred*, les années de développement et les ratios entre les deux triangles. Ensuite, les variables sont entraînées dans K RNN qui fournissent chacun un triangle de montants cumulés prédit $\overline{D}^k$ où $k \in \{1, ..., K\}$. Chaque triangle prédit permet d'obtenir un triangle moyenné $\overline{D}$, comme base pour le calcul de nouveaux facteurs de développement et de coefficients de volatilité « Mack-Net »définis par

$$\hat{f}_j^p = \frac{\sum_{i=I-j+2}^{I} \overline{D}_{i,j}}{\sum_{i=I-j+2}^{I} \overline{D}_{i,j-1}},$$

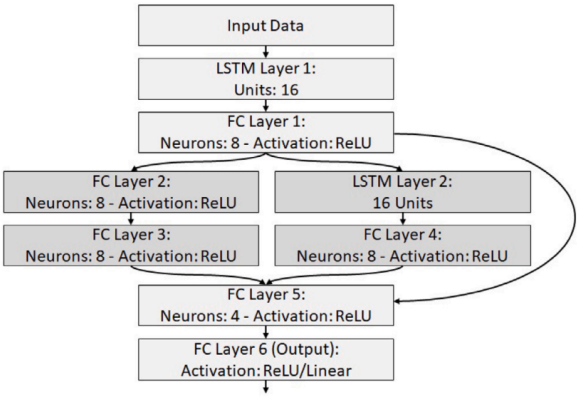
$$\hat{\sigma}_j^{2,p} = \frac{1}{I-j-1} \sum_{i=0}^{I} \overline{D}_{i,j} \left(\frac{\overline{D}_{i,j}}{\overline{D}_{i,j-1}} - \overline{f}_j \right)^2,$$

avec $\overline{f}_j = \frac{\sum_{i=0}^{I} \overline{D}_{ij}}{\sum_{i=0}^{I} \overline{D}_{ij-1}}$ et $\overline{D}_{ij} = \frac{\sum_{k=1}^{K} \overline{D}_{ij}^k}{K}$.

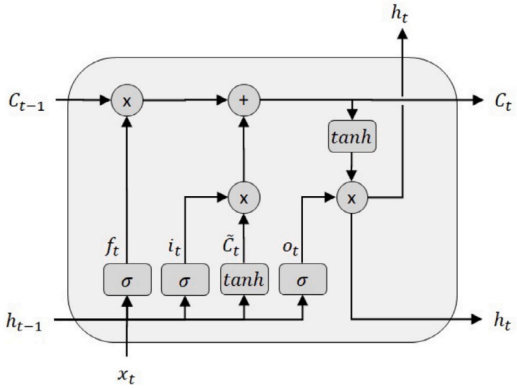
Chaque réseau de neurones possède la même architecture décrit dans la figure 3.14a. Il est composé de plusieurs couches denses et d'une cellule LSTM. Cependant, l'application de la librairie `Keras` induisant de l'aléa lors de l'initialisation des poids des réseaux et l'utilisation du paramètre *dropout* justifient le fait d'entraîner plusieurs RNN puis d'en faire la moyenne. Les RNN étant classiquement utilisés sur de plus grosses bases de données qu'un simple triangle de liquidation, calculer la moyenne permet en outre de réduire le bruit lié aux poids initiaux et à la taille relativement faible du triangle.

Il faut remarquer également qu'il existe une connexion entre la première couche dense et la cinquième couche dense. Cette connexion, schématisée en figure 3.14a, qui agit comme « raccourci », permet de ne pas entraîner le modèle sur les couches intermédiaires (FC 3 et 4) ce qui permet d'éviter deux sources de problèmes : la complexité de la structure du réseau pouvant entraîner du surapprentissage et le problème des *vanishing gradients* rencontré lors de la *backpropagation*. Comme vu en 3.2.1, ce dernier problème apparaît à cause du calcul du gradient, notamment lors de l'application de la règle de la chaîne qui réduit la capacité de propagation du signal dans le réseau. La cellule LSTM à seize unités de chaque réseau de la figure 3.14b correspond à celle décrite en 3.2.2. En particulier, $\mathbf{W_f}, \mathbf{W_i}, \mathbf{W_c}, \mathbf{W_o}, \mathbf{b_f}, \mathbf{b_i}, \mathbf{b_c}, \mathbf{b_o}$ correspondent aux paramètres (matrices de poids et vecteurs de biais) des RNN et σ correspond toujours à la fonction sigmoïde. Plus globalement, les notations du modèle sont les suivantes

$$
\begin{aligned}
\mathbf{f_t} &= \sigma(\mathbf{W_f}[\mathbf{h_{t-1}}, \mathbf{x_t}] + \mathbf{b_f}),\\
\mathbf{i_t} &= \sigma(\mathbf{W_i}[\mathbf{h_{t-1}}, \mathbf{x_t}] + \mathbf{b_i}),\\
\tilde{\mathbf{c}}_\mathbf{t} &= \tanh(\mathbf{W_c}[\mathbf{h_{t-1}}, \mathbf{x_t}] + \mathbf{b_c}),\\
\mathbf{c_t} &= \mathbf{f_t c_{t-1}} + \mathbf{i_t}\tilde{\mathbf{c}}_\mathbf{t},\\
\mathbf{o_t} &= \sigma(\mathbf{W_o}[\mathbf{h_{t-1}}, \mathbf{x_t}] + \mathbf{b_o}),\\
\mathbf{h_t} &= \mathbf{o_t}\tanh(\mathbf{c_t}).
\end{aligned}
$$



(a) Architecture d'un RNN du modèle Mack-Net



(b) Cellule LSTM du modèle Mack-Net

FIGURE 3.14 : Architecture des RNN et cellules LSTM dans le modèle Mack-Net (RAMOS-PÉREZ et al., 2022)

Le modèle Mack-Net, dans sa version stochastique, propose enfin des simulations des réserves par bootstrap. Les auteurs supposent que

$$\mathbb{E}[{C^{Pa}}_{i,j}] = \hat{f}_j^p \overline{D}_{i,j-1} \text{ et } \mathbb{V}({C^{Pa}}_{i,j}) = \hat{\sigma}_j^{2p} \overline{D}_{i,j-1}.$$

En comparaison avec l'algorithme 1, des résidus du bootstrap sont calculés de deux manières, avec ou sans ajustement

$$\hat{r}_{i,j}^p = \frac{\sqrt{\overline{D}_{i,j-1}}\left(\frac{\overline{D}_{i,j}}{\overline{D}_{i,j-1}} - \overline{f}_j\right)}{\hat{\sigma}_j^{2,p}},$$

$$\hat{r}_{i,j}^{p} = \sqrt{\frac{N}{N-p}} \frac{\sqrt{\overline{D}_{i,j-1}} \left(\frac{\overline{D}_{i,j}}{\overline{D}_{i,j-1}} - \overline{f}_j \right)}{\hat{\sigma}_j^{2,p}},$$

où p est le nombre de facteurs de développement. Ces résidus utilisent à la fois les observations du triangle supérieur de liquidation, mais également les prédictions du triangle $\overline{D}$ inférieur estimé par les réseaux de neurones, contrairement au bootstrap de Mack qui ne se base que sur l'observation passée. Les facteurs de développement simulés comme dans l'étape 3 de l'algorithme 1 sont donnés par

$$\tilde{f}_j^{B,p} = \frac{\sum_{i=0}^{I-j+1} \overline{D}_{ij-1} f_{ij}^{B,p}}{\sum_{i=0}^{I-j+1} \overline{D}_{ij-1}},$$

où

$$f_{ij}^{B,p} = \hat{f}_j^p + r_{i,j}^{B,p} \frac{\hat{\sigma}_j^p}{\sqrt{\overline{D}_{ij-1}}},$$

pour obtenir finalement

$$\hat{C^{Pa}}_{ij} = \overline{D}_{ij-1} \tilde{f}_j^{B,p}.$$

Tout l'intérêt du choix du modèle Mack-Net pour répondre à la problématique repose sur sa capacité à fournir de nouveaux coefficients $\hat{f}_j^p$ et $\hat{\sigma}_j^{2,p}$ à partir du triangle prédit $\overline{D}$ en vue de calculer une nouvelle formule (1.9) pour les paramètres USP.

Les hyperparamètres du modèle Mack-Net. Pour optimiser l'entraînement d'un réseau de neurones, l'initialisation des poids est une étape non négligeable. Les cellules LSTM sont initialisées aléatoirement avec les pondérations de Glorot (Glorot et Bengio, 2010) selon une loi $\mathcal{U}\left(-\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\right)$ avec n le nombre d'entrées du modèle. Les cellules récurrentes sont quant à elles initialisées avec l'approche orthogonale (Saxe et al., 2014) : dans chaque cellule, les poids initiaux correspondent à une matrice orthogonale aléatoire, i.e. telle que $\mathbf{W}^T\mathbf{W} = I$.

La dégradation des pondérations (*weight decay*) dans l'algorithme 4, est également utilisée dans le modèle. Il s'agit d'une méthode de régularisation L^2 de type Ridge (Hoerl et Kennard, 1970) permettant également d'éviter le surapprentissage du modèle. Elle consiste à ajouter un paramètre de pénalité à la fonction d'erreur de la forme

$$\lambda \sum_i W^{[i]2}$$

où le paramètre λ est un hyperparamètre. Enfin, le modèle propose un paramètre nommé AR permettant d'ignorer, si le paramètre est affecté à 0, la variable d'entrée $\mathbf{X_2}$ du modèle qui correspond à la variable auto-régressive. Les autres paramètres (*dropout*, nombre d'époques, nombre de RNN, taux d'apprentissage et fonction d'activation de sortie) ont déjà été définis dans la section 3.2.1 et au fil du chapitre *supra*.

3.3.2 Mise en place du modèle Mack-Net

Séparation des données. Pour entraîner le modèle Mack-Net (et plus généralement la majorité des modèles de provisionnement sur triangles agrégés), les auteurs distinguent trois ensembles de données, décrits dans la figure 3.15 :

- l'ensemble d'entraînement, sur lequel l'apprentissage est effectué, correspondant au triangle supérieur privé de sa dernière diagonale (cellules bleues),
- l'ensemble de validation, sur lequel est effectuée l'optimisation des hyperparamètres du modèle, correspondant à la dernière diagonale du triangle supérieur, à l'aide d'une métrique d'entraînement (cellules rouges),
- l'ensemble de test, sur lequel est calculée la performance du modèle sur de nouvelles données, correspondant à la première diagonale prédite (dit autrement, la première diagonale du triangle inférieur), à l'aide d'une métrique de test (cellules vertes).

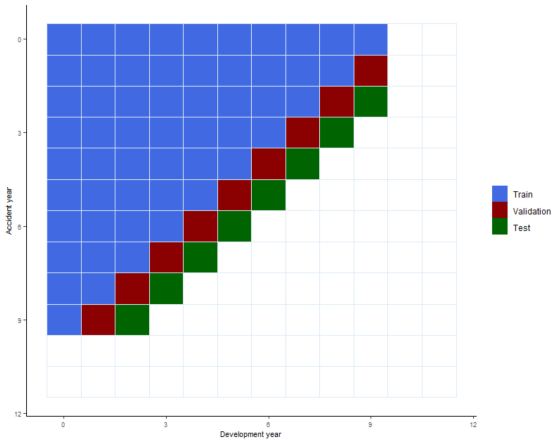


FIGURE 3.15 : Séparation des données d'entraînement, validation et test (PITTARELLO, 2023)

Métriques d'entraînement utilisées. Disposant des triangles des paiements, celui des *incurred* et d'une mesure (vecteur) d'exposition par année d'accident, correspondant au nombre total de jours assurés par la captive pour tous les véhicules sur une année civile multiplié par un coefficient d'abattement (*rater*), le modèle Mack-Net peut ainsi être mis en place. L'optimisation des hyperparamètres lors du *tuning* du modèle Mack-Net est une étape clé. Dans une démarche d'optimisation du modèle, le modèle est entraîné avec différentes configurations d'hyperparamètres, puis la performance est testée par rapport à la diagonale de validation (figure 3.15) . Une erreur de performance est utilisée pour optimiser les hyperparamètres : dans le cas du modèle initial proposé par les auteurs, il s'agit de la RMSE (*root-mean-square error*) entre la dernière diagonale observée et les valeurs prédites par le RNN sur le triangle $\overline{D}$. Pour trouver les hyperparamètres, la procédure est la suivante, en appliquant un *random search* :

- entraînement du modèle avec différentes gammes d'hyperparamètres,
- sauvegarde des résultats de l'erreur de test,
- sélection de la configuration d'hyperparamètres minimisant l'erreur de validation.

Le présent travail reposant sur le calcul du paramètre USP et l'étude du risque de réserve à horizon un an, il convient d'explorer et choisir également d'autres mesures qui reflètent cette incertitude. Deux métriques à minimiser, propres au provisionnement (BALONA et RICHMAN, 2020) peuvent être utilisées :

- l'*AvE* (*Actual versus Expected*), comme mesure de la qualité des prédictions à un an (première diagonale prédite),
- le CDR, comme mesure de stabilité de la méthode considérée, d'une année calendaire à l'autre.

Pour rappel, en relation avec l'équation (1.2), le CDR pour l'année k s'écrit comme

$$\begin{aligned} CDR_i^{I+1} &= R_{i,j*}^{I+1} - R_{i,j*}^{I} + \left(Y_{i,j*} - \hat{Y}_{i,j*}^{I}\right) \\ &= R_{i,j*}^{I+1} - R_{i,j*}^{I} + AvE_{i,j*}^{I+1}, \end{aligned}$$

où $j^* = I - i$.

En d'autres termes,

$$CDR = AvE + \Delta IBNR,$$

donc au global, la métrique CDR est aussi intéressante comme métrique d'entraînement puisqu'elle pénalise les méthodes qui estiment des variations d'IBNR importantes entre deux années calendaires. Pour rappel, pour un organisme d'assurance, le CDR est une mesure de profit ou de perte par rapport aux années de survenance des accidents passés. Par exemple, en se fiant à la figure 3.15 pour estimer les hyperparamètres du modèle associé au triangle supérieur arrêté à la diagonale rouge, on considérera donc les variations entre le triangle arrêté un an avant (dernière diagonale en bleue) et le triangle arrêté à la diagonale rouge. En particulier, afin de pondérer en fonction des années d'accident dans le calcul des réserves, Balona et Richman (2020) utilisent les scores suivants

$$CDR_{\text{score}} = \sqrt{\frac{\sum_{i=1}^{I} |Y_{i,j*}| \cdot (CDR_i^{I+1})^2}{\sum_{i=1}^{I} |Y_{i,j*}|}} \text{ et } \quad AvE_{\text{score}} = \sqrt{\frac{\sum_{i=1}^{I} |Y_{i,j^*}| \cdot (AvE_i^{I+1})^2}{\sum_{i=1}^{I} |Y_{i,j^*}|}}.$$

Ces métriques sont à considérer comme des RMSE et vont être utilisées comme métrique d'entraînement pour trouver les hyperparamètres optimaux.

Enfin, pour concilier à la fois la précision de la méthode de provisionnement et sa stabilité, on pourrait imaginer, parmi les valeurs de la grille choisie, une autre erreur de validation lors de l'entraînement du modèle Mack-Net, comme par exemple la métrique suivante, nommée *Combined Score* dans la suite du travail

$$\alpha.\frac{CDR_{\text{score}} - \mathbb{E}(CDR_{\text{score}})}{\max(CDR_{\text{score}}) - \min(CDR_{\text{score}})} + (1-\alpha).\frac{AvE_{\text{score}} - \mathbb{E}(AvE_{\text{score}})}{\max(AvE_{\text{score}}) - \min(AvE_{\text{score}})},$$

avec $\alpha \in [0;1]$ un poids défini arbitrairement par l'actuaire.

Métriques de test utilisées. En plus des métriques déjà mentionnées comme métriques d'entraînement, d'autres plus classiques peuvent être utilisées, comme la MAE (*Mean Absolute Error*) ou la MAPE (*Mean Absolute Percentage Error*) à un an, c'est-à-dire en comparant les valeurs estimées et observées de la première diagonale du triangle inférieur, ou à l'ultime, en comparant les dernières colonnes des triangles de liquidation. Mathématiquement, ces mesures s'écrivent

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \quad \text{et} \quad MAPE = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{y_i - \hat{y}_i}{y_i}\right|,$$

où n est la dimension du triangle *backtesté*.

Résultats de l'entraînement du modèle. Sur le triangle complet, la recherche aléatoire s'effectue parmi une grille de valeurs d'hyperparamètres décrite en table 3.1 sur 1 000 combinaisons possibles.

Hyperparamètres	**Grille de valeurs**
Taux d'apprentissage η	0.005, 0.01, 0.02, 0.1
dropout	0.05, 0.1, 0.2, 0.3, 0.4, 0.5
Époques	(40,...,50)
RNN entrainés	2, 4, 6, 8, 10, 12, 14
weight decay	0, 0.001, 0.01
Fonction d'activation de sortie ϕ	identité, $ReLU$
AR	0,1

η	***dropout***	**Époques**	RNN	***weight decay***	ϕ	**AR**	**Erreur de test**
0.02	0.1	42	2	0	$ReLU$	1	15.67%

TABLE 3.1 : Grille et sélection des hyperparamètres pour le triangle complet

Toutes choses étant égales par ailleurs, tracer graphiquement l'erreur de validation en fonction des paramètres comme en figure 3.16 peut indiquer la grille à considérer.

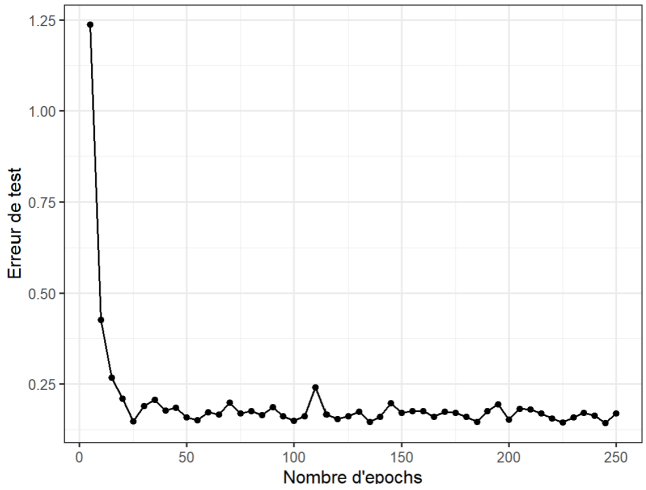


FIGURE 3.16 : Erreur de validation du modèle Mack-Net initial en fonction du nombre d'époques

3.4 Présentation des résultats

3.4.1 Paramètres USP et réserves obtenues

Avec la métrique initiale du modèle. Les réserves du modèle Mack-Net sont données dans la table 3.2. Comme dans la table 3.10, les variations des réserves d'une année sur l'autre pour le modèle Mack-Net sont beaucoup plus importantes, et ce pour une raison évidente : l'erreur de validation du modèle initialement proposée par les auteurs ne pénalise pas suffisamment les variations d'IBNR trop élevées, d'une année calendaire à l'autre. Les paramètres USP, estimés via l'équation (1.9) (où les $\hat{Q}_j$ sont nouvellement calculés avec le modèle Mack-Net) se trouvent dans la table 3.3. Du fait de la métrique ignorant l'impact du CDR lors de l'entraînement, des paramètres peuvent être non satisfaisants pour un organisme d'assurance, comme pour l'année 2009 (23,19%).

Année	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
Mack	11704	13428	16300	21042	29534	25377	26104	32056	36732	30597	23234	26017	29366	30840	36658	40050
Mack-Net	17874	13399	21009	21033	19644	15271	24197	25268	27253	31448	36267	39980	46581	47383	58966	47841

TABLE 3.2 : Comparaison des réserves par méthode de Mack et Mack-Net pour les années calendaires 2004 à 2019 (milliers €)

Année	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
M&W	20,22%	16,90%	14,48%	12,71%	9,50%	10,39%	10,88%	11,55%	12,30%	13,90%	16,81%	15,76%	16,53%	14,83%	14,31%	12,86%
Mack-Net	18,10%	17,11%	12,84%	18,18%	20,62%	23,19%	15,56%	18,61%	12,61%	13,75%	10,18%	10,39%	9,91%	11,17%	9,00%	7,29%
ECLRM	23,10%	25,25%	25,32%	25,89%	17,93%	20,48%	28,39%	22,82%	24,65%	24,31%	20,67%	23,18%	13,55%	15,82%	16,94%	16,70%

TABLE 3.3 : Comparaison des paramètres USP entre Merz & Wüthrich, ECLRM et Mack-Net pour les années calendaires 2004 à 2019

Avec la métrique $\mathbf{CDR_{score}}$. Il convient donc, dans un second temps de modifier la définition de l'erreur de validation permettant de trouver les hyperparamètres optimaux et de minimiser le CDR_{score}, introduit en section 3.3.2 (et illustré en table 3.4). Ainsi, les scores obtenus permettent d'obtenir un plus faible CDR_{score}, au détriment de la dégradation des indicateurs comme l'AvE_{score}.

CDR_{score}	AvE_{score}	MAE^{1yr}	$MAPE^{\text{1yr}}$	MAE^{ult}	$MAPE^{\text{ult}}$
729,8	504,7	141,6	1,61%	470,2	5,7%

TABLE 3.4 : Performances moyennes des modèles Mack-Net avec les hyperparamètres minimisant le CDR_{score} (milliers €)

.

Les paramètres USP obtenus finaux sont décrits en table 3.5. Les premiers paramètres obtenus sont toujours non satisfaisants, ce qui signifie que les RNN estiment mal lorsque peu de données lui sont fournies en entrée (ce problème est déjà mentionné dans la section 3.3.1). Cette mauvaise estimation peut être liée à du sous-apprentissage dû à la complexité du modèle, au faible nombre d'observations ou encore à l'inclusion d'aléa important lié aux recours dans les *incurred* du modèle pour les premières années d'accident. Pour rappel, les réserves de la table 2.4 associées au modèle de la section 2.2.3 estiment également de faibles provisions pour les premières années du *backtesting*.

Année	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Mack-Net	25,72%	19,17%	14,56%	17,32%	19,97%	17,45%	14,29%	16,81%	13,59%	10,45%	8,59%	10,18%	9,75%	9,59%	11,15%	7,99%

TABLE 3.5 : Paramètres USP finaux obtenus avec les modèles Mack-Net minimisant le CDR_{score} pour les années calendaires 2004 à 2019

Avec la métrique *Combined Score*. Enfin, il convient de modifier à nouveau la mesure d'entraînement en considérant la métrique *Combined Score*. La table 3.6 décrit les scores sur les données d'entrainement obtenus en choisissant $\alpha = 60\%$ pour chaque triangle lors du *backtesting*. Ce paramètre est à moduler par l'expert en fonction de la tâche à accomplir. Dans le cas des paramètres USP, on souhaite donner plus d'importance à la minimisation du CDR. Les performances sont décrites dans la table 3.7. La métrique *Combined Score* offre un meilleur score d'AvE_{score} par rapport à la table 3.4 mais en contrepartie un moins bon score de CDR_{score} (toutefois meilleur que celui de la table 3.2). D'après tous ces scores, dans une logique de pure minimisation de la MSEP du CDR, les paramètres USP les plus fiables sont donc ceux donnés en 3.5.

Année	$CDR_{\textbf{score}}$	$AvE_{\textbf{score}}$	***Combined Score***
2002	250,2	339,8	-0,37
2003	187,3	192,5	-0,34
2004	187,3	113,3	-0,40
2005	227,9	301,0	-0,34
2006	447,7	405,5	-0,41
2007	323,7	228,0	-0,36
2008	390,8	857,3	-0,28
2009	610,5	810,3	-0,26
2010	327,4	205,1	-0,29
2011	581,6	335,1	-0,21
2012	521,7	487,4	-0,26
2013	446,7	239,4	-0,24
2014	332,5	166,0	-0,30
2015	470,0	508,0	-0,26
2016	982,2	823,7	-0,16
2017	284,4	246,8	-0,29
2018	455,9	405,9	-0,20
2019	623,4	400,4	-0,20

TABLE 3.6 : *Combined Scores* obtenus lors de l'hyperparamétrage du modèle Mack-Net

CDR_{score}	AvE_{score}	MAE^{1yr}	$MAPE^{\text{1yr}}$	MAE^{ult}	$MAPE^{\text{ult}}$
801,8	457,7	124,7	1,32%	522,6	5,4%

TABLE 3.7 : Performances moyennes des modèles Mack-Net avec les hyperparamètres minimisant la métrique *Combined Scores* (milliers €)

.

Ainsi, en minimisant la métrique adéquate, le modèle Mack-Net, plus flexible du fait de sa structure en RNN, peut permettre de mieux calibrer le calcul des paramètres USP par rapport à la méthode de Merz & Wüthrich, en apprenant mieux lors de l'apprentissage et en sélectionnant dans la mémoire des RNN les informations pertinentes (malgré le bruit et les comportements anormaux inhérents aux données).

3.4.2 Paramètres Mack-Net obtenus sur le triangle

Tout l'intérêt des modèles Mack-Net et Mack Chain-Ladder repose sur leurs définitions des facteurs de lien et de volatilité. Les tables 3.8, 3.9 ainsi que la figure 3.17 permettent d'apprécier les différences des paramètres entre les deux modèles.

Développement	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Mack-Net	2.2644	1.1832	1.0832	1.0599	1.0371	1.0299	1.0256	1.0219	1.0175	1.0134	1.0108	1.0101	1.0094	1.0092	1.0092	1.0093	1.0093	1.0091	1.0090	1.0089	1.0088
Mack	2.3053	1.2031	1.0849	1.0616	1.0584	1.0281	1.0259	1.0502	1.0140	1.0269	1.0034	1.0296	1.0003	1.0129	1.0005	1.0003	1.0005	1.0000	1.0000	1.0000	1.0000

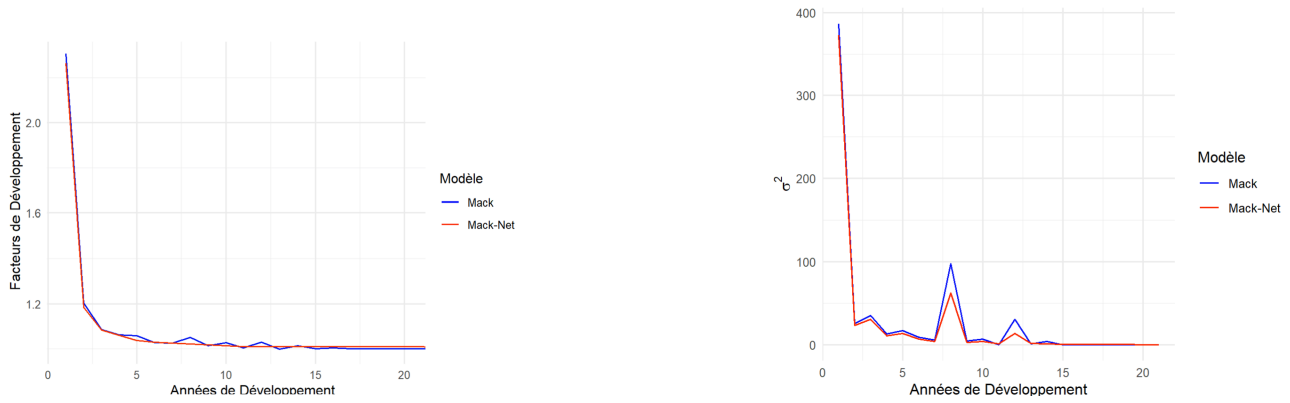
TABLE 3.8 : Facteurs de développement estimés $\hat{f}_j^p$ et $\hat{f}_j$ de Mack-Net et Mack du triangle de liquidation (cumulé) des paiements

Développement	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Mack-Net	369.0915	23.2958	30.4959	10.6735	14.1340	6.6777	3.8451	62.9208	2.7519	4.2350	0.2835	14.2589	0.6851	1.3018	0.1603	0.1133	0.1379	0.1193	0.0838	0.0575	0.0402
Mack	386.3954	25.3732	35.4894	13.1433	17.3027	9.3156	5.7507	97.9769	4.7386	7.0722	0.2638	30.7990	0.8910	3.7718	0.0082	0.1899	0.0062	0.0001	0.0000	0.0000	0.0000

TABLE 3.9 : Coefficients de volatilités estimés $\hat{\sigma}_j^{2,p}$ et $\hat{\sigma}_j^2$ de Mack-Net et Mack du triangle de liquidation (cumulé) des paiements

Remarque 4. *Avec les nouveaux coefficients Mack-Net, l'hypothèse 1.2.3 reste vérifiée sur l'ensemble du triangle de liquidation.*

Entre les deux modèles, les coefficients estimés semblent converger. D'après les figures 3.17a et 3.17b, le modèle Mack-Net semble lisser l'aléa statistique lié aux facteurs de développement et atténuer les valeurs $\hat{\sigma}_j^2$ trop volatiles dans le modèle de Mack. Cependant, le même problème que dans RAMOS-PÉREZ et al. (2022) (toutefois non-mentionné explicitement par les auteurs) apparaît. En effet, bien que les facteurs de lien soient décroissants, le dernier facteur $\hat{f}_{21}^p$ n'est pas strictement égal à 1, ce qui, théoriquement, exigerait l'inclusion d'un facteur de queue dans le triangle. Au vu de la profondeur de l'historique des données, il n'est pas raisonnable d'imaginer un écoulement des règlements non abouti après 22 années. Cette remarque est donc considérée comme une limite de plus à l'étude. De plus, ajouter une contrainte dans la fonction de coût du modèle pour faire converger strictement les coefficients à 1 pourrait faire apparaître un biais par rapport au modèle initial. Il est plutôt décidé d'affecter $\hat{f}_{21}^p$ à la valeur $\hat{f}_{21}$ (1, dans ce cas).



(a) Facteurs de développement $\hat{f}_j$ et $\hat{f}_j^p$ de Mack et Mack-Net

(b) Coefficients de volatilité $\hat{\sigma}_j^2$ et $\hat{\sigma}_j^{2,p}$ de Mack et Mack-Net

FIGURE 3.17 : Comparaison des paramètres du modèle de Mack Chain-Ladder et Mack-Net

Enfin, des simulations de bootstrap peuvent être effectuées pour obtenir une distribution des réserves du modèle Mack-Net, comme l'illustre la figure 3.18.

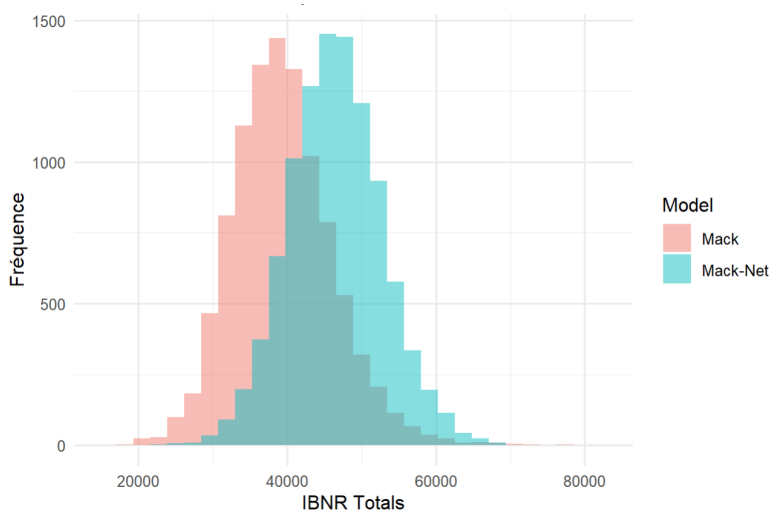
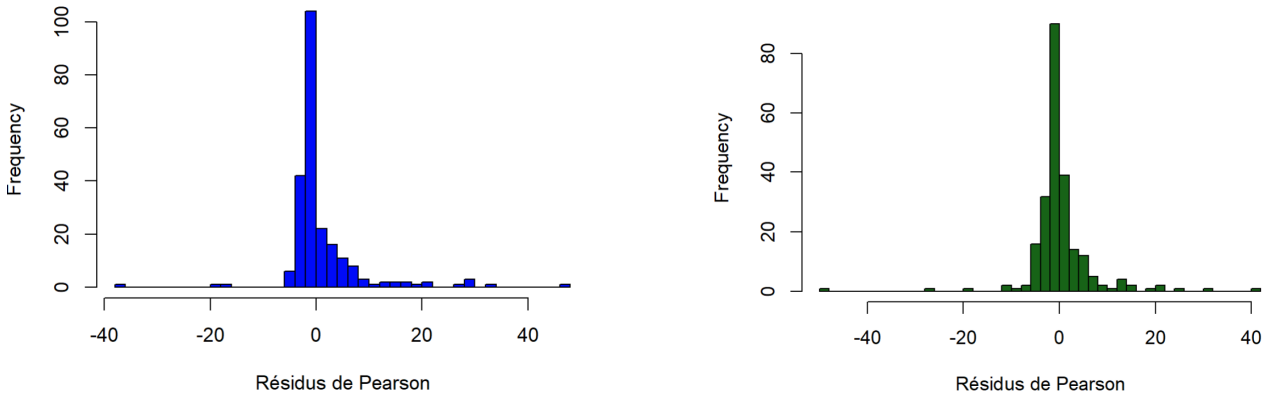


FIGURE 3.18 : Procédure bootstrap (10 000 simulations) du modèle Mack-Net

L'histogramme de la figure 3.19 des résidus de Pearson (de l'algorithme 1) permet d'observer leur distribution. De prime abord, les résidus de Mack semblent plus centrés que ceux de Mack-Net. Leur espérance est respectivement de 0,230 et 0,602, et leur variance de 49,04 et 51,57. Cela permet de valider le fait (comme en conclusion de la table 3.10) que la méthode de Mack réduit la distorsion et augmente la consistance des estimations à l'ultime. Cependant, même si les moyennes sont positives, la proportion de résidus négatifs dans le modèle Mack-Net est plus importante dans les résidus de Mack (65%, contre 60%) ce qui peut justifier une surestimation des provisions pour le modèle Mack-Net visible en figure 3.18. Enfin, quelques valeurs de résidus positifs élevés peuvent s'expliquer, dans le contexte de la RC automobile, par des sinistres considérés bénins *a priori* mais qui finalement s'avèrent être des sinistres graves (comme un sinistre corporel), d'où le fait que l'organisme d'assurance « sous-provisionne ».



(a) Résidus de Pearson de Mack-Net (b) Résidus de Pearson de Mack

FIGURE 3.19 : Comparaison des résidus de Pearson du modèle de Mack Chain-Ladder et Mack-Net

3.5 *Benchmark* du modèle Mack-Net

Afin de tirer parti de la potentielle robustesse modèle Mack-Net, il convient de comparer, à partir d'autres modèles traditionnels ou de *Deep Learning* existant sur le marché, et de métriques utilisées usuellement en provisionnement non-vie, les performances du modèle Mack-Net. Cette étape est une technique de vérification usuelle, avant de présenter les paramètres USP finaux obtenus avec le modèle Mack-Net. Elle permet d'apprécier les performances du modèle Mack-Net par rapport à des modèles de marché classiques (Mack, Munich Chain-Ladder) et également de *Deep Learning*. La section permet d'évaluer également la méthode la mieux adaptée aux données de la captive.

3.5.1 Le modèle *blended cross-classified neural network* (bCCNN)

Le modèle bCCNN (*blended cross-classified neural network*) propose une modélisation en réseaux de neurones embarquant le modèle linéaire généralisé Poisson décrit en section 2.2.2 (Gabrielli et al., 2019). Le modèle est construit grâce à la librairie `Keras` et est partiellement disponible dans l'article. Les preuves, elles mêmes non décrites par l'auteur, sont tout de même disponibles dans Kremer (1985) et Mack (1991b). Pour plus d'illustration, le mémoire de Tapsoba Wendinmanedge (2021) explore et applique ce modèle de manière détaillée. Les hypothèses du modèle sont identiques aux hypothèses 2.2.1.

En gardant les mêmes notations que dans la section 2.2.2, la fonction à prédire est décrite par

$$\mu^Y(\cdot;\cdot) : \{1,\dots,I\} \times \{0,\dots,J\} \to \mathbb{R},$$

où

$$(i,j) \mapsto \mu^Y(i,j;m) = \mathbb{E}[Y_{i,j}] = \mu^Y_{i,j}.$$

Couche *embedding*. La première étape du modèle consiste à intégrer les années d'accident i et de développement j dans une couche *embedding*. Cette intégration est une technique d'apprentissage dite par représentation, permettant de projeter les années d'accident et de développement sur un espace de dimension inférieure (1, ici). Le but d'un processus d'*embedding* est de capturer toutes les connexions, linéaires ou non, entre les variables explicatives et la variable cible. Généralement, l'*embedding* permet d'obtenir des résultats plus convaincants qu'un encodage binaire (*one-hot encoding*) du fait de la réduction de dimension et de la mesure directe de similarité entre les différentes années.

Les couches *embedding* sont définies analytiquement par les fonctions $a(i)$ et $b(j)$ telles que

$$a(i) : \{1,\dots,I\} \to \mathbb{R}, \quad i \mapsto a(i) = a^Y_i,$$

et

$$b(j) : \{0,\dots,J\} \to \mathbb{R}, \quad j \mapsto b(j) = b^Y_j.$$

Ces deux fonctions associent à chaque année d'accident et de développement le coefficient estimé par maximum de vraisemblance du modèle quasi-Poisson. Les composantes de l'*embedding* vont donc être des poids optimisés par le réseau de neurones au fur et à mesure de l'apprentissage. En effet, les deux premiers neurones de la première couche décrits sur la figure 3.20 sont donnés par

$$\mathbf{z}^{[0]}(i,j) = (a(i), b(j)) = \left(a^Y_i, b^Y_j\right).$$

Au total, le réseau de neurones du modèle est composé de K couches cachées et de la fonction d'activation $\phi = \tanh$. Cette fonction est notamment choisie car $\phi' = 1 - \phi^2$, ce qui aide pour le calcul

du gradient et $\phi \in [-1, 1]$, ce qui empêche l'explosion de gradient (problème inverse au problème évoqué en partie 3.2.1 où les valeurs du gradient deviennent excessivement grandes, ce qui a pour incidence de stopper l'apprentissage du réseau) et accélère ainsi la convergence du réseau.

Chaque couche cachée contient q_k neurones,

$$\mathbf{z}^{[k]} = \mathbf{z}^{[k]}(i,j) = \left(z_1^{[k]}(i,j), \ldots, z_{q_k}^{[k]}(i,j)\right) \in \mathbb{R}^{q_k},$$

$$\mathbf{z}_l^{[k]}(i,j) = \phi\left(\mathbf{c}_k + B'_k \mathbf{z}^{(k-1)}\right),$$

où $\mathbf{c}_k \in \mathbb{R}^{q_k}$ l'intercept et $\mathbf{B'}_k \mathbf{z}^{[k-1]} \in \mathbb{R}^{q_{k-1} \times q_k}$ la matrice de poids.

A la sortie du réseau, la fonction de sortie (avec la fonction d'activation exponentielle) est

$$(i,j) \mapsto \mu(i,j) = \mathbf{z}^{[K]}(i,j) = \exp\left\{a(i) + b(j) + \mathbf{c}_K + B'_K \mathbf{z}^{[K-1]}\right\}.$$

Enfin, la fonction de perte du modèle pour Y est issue de l'équation (2.3). En effet, maximiser la fonction de quasi-vraisemblance revient à minimiser, par descente de gradient, la fonction de perte de la déviance

$$\mathcal{L}\left((\mu_{i,j})_{i,j}, \mathcal{D}_I; \phi\right) = \frac{2}{\phi} \sum_{i+j \leq I} \mu_{i,j} - Y_{i,j} + Y_{i,j} \log\left(\frac{Y_{i,j}}{\mu_{i,j}}\right), \tag{3.2}$$

avec

$$\hat{\phi} = \frac{\mathcal{L}\left(\mu_{i,j}, D_I, 1\right)}{|D_I| - p}.$$

La variable p correspond au nombre de paramètres du modèle de la section 2.2.2, et D_I désigne les données d'entrainement. Enfin, la *skip connection* sur la figure 3.20 permet à la couche de sortie d'avoir un accès direct à des informations issues de la couche *embedding*. En effet, à l'initialisation des poids, la *skip connection* permet de considérer $a_i = \hat{a}_i^{\text{ODP}}$, $b_j = \hat{b}_j^{\text{ODP}}$, $\mathbf{c}_K = \hat{\mathbf{c}}^{\text{ODP}}$, $B'_K = 0$.

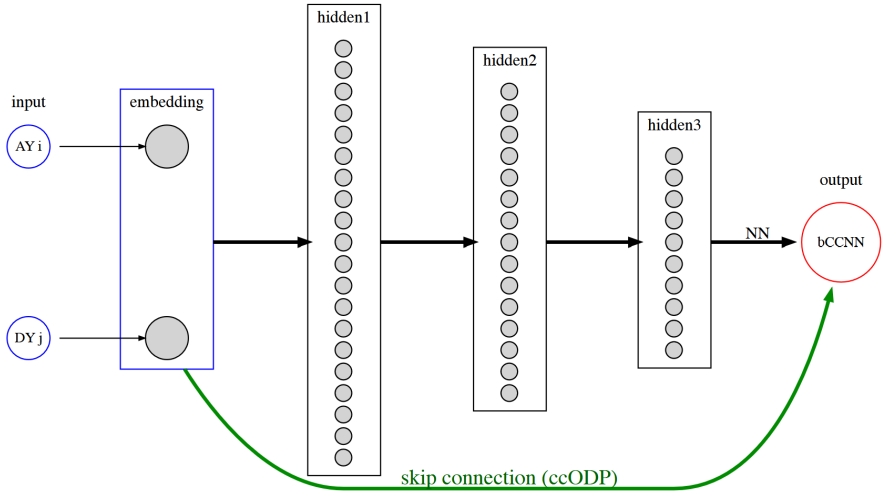


FIGURE 3.20 : Architecture du modèle de marché bCCNN (GABRIELLI et al., 2019)

3.5.2 Le modèle *DeepTriangle*

Le modèle *DeepTriangle* est présenté en annexe A.10. En effet, bien que ce modèle ait des similarités avec le modèle Mack-Net, une limite majeure quant à la pertinence de la comparaison fait obstacle. En effet, comme décrit dans l'annexe, le modèle *DeepTriangle* repose sur des données d'organismes d'assurance américains disponibles seulement pour une durée de dix années (de 1988 à 1997) ce qui empêche de pouvoir effectuer la comparaison sur un historique plus profond. Le modèle *DeepTriangle* disponible est de plus uniquement pré-entrainé sur ces données américaines. Une autre limite du modèle, du fait de l'absence des données ligne à ligne complètes, est l'impossibilité d'obtenir une statistique *as-if* des données de la captive « translatées »à la période 1988-1997, car les sinistres ouverts composant le triangle de liquidation sont à l'euro d'aujourd'hui. Les scores pour un triangle d'horizon dix ans entre le modèle Mack-Net et *DeepTriangle* sont toutefois fournis dans la table 3.11.

3.5.3 Résultats du *benchmark*

Comme déjà énoncé, en plus de comparer la méthode de Mack Chain-Ladder avec le modèle Mack-Net, il est intéressant, dans le cadre du *benchmark*, de confronter les scores du modèle bCCNN (basé sur du *Deep Learning*) présenté en section 3.5.1 et de modèles plus traditionnels utilisant, comme dans le modèle Mack-Net, le triangle des *incurred* : la méthode ECLRM et celle de Munich Chain-Ladder (Quarg et Mack, 2004), présentées en section 2.2.3. Dans le modèle Mack-Net, la prise en compte des paiements et des charges est moins contraignante que pour un modèle type Munich ou ECLRM, dû à la flexiblité des structures en réseaux de neurones. La combinaison du *Deep Learning* et la prise en compte des charges conduit finalement aux meilleures performances. En effet, la table 3.10 montre d'abord que le modèle Mack-Net est bien meilleur en termes de performances que le modèle bCCNN. Elle permet en outre d'observer de meilleurs scores à un an pour le modèle initial Mack-Net (AvE_{score}, $MAE^{1\text{yr}}$, et $MAPE^{1\text{yr}}$) par rapport à Chain-Ladder mais de moins bons résultats en vision à l'ultime. Ces résultats peuvent sembler logiques, puisque l'hyperparamétrage du modèle est basé sur la minimisation de la RMSE de la diagonale de validation. Ces résultats laissent donc penser que, même si le modèle Mack-Net est plus précis que Chain-Ladder sur la première diagonale prédite, les paramètres USP estimés, reposant sur l'erreur quadratique du CDR dans l'équation (1.9), pourraient être moins précis. Un changement de métrique de validation peut ainsi donc être envisagé, comme effectué dans la section 3.4. Enfin, de manière marginale, la table 3.11 montre que, sur un triangle de taille 10×10, le modèle Mack-Net fournit globalement de meilleures performances que le modèle *DeepTriangle.*

Métriques	**Mack-Net**	**Mack Chain-Ladder**	**ECLRM**	**Munich Chain-Ladder**	**bCCNN**
AvE_{score}	462,0	501,4	753,7	490,0	722,3
CDR_{score}	850,9	770,9	1381,1	1075,8	1296,2
$MAE^{1\text{yr}}$	143,7	163,5	304,9	129,4	342,0
$MAPE^{1\text{yr}}$	1,6%	2,1%	3,9%	1,6%	4,7%
MAE^{ult}	411,3	175,0	1070,9	181,0	1249,3
$MAPE^{\text{ult}}$	4,9%	2,3%	11,6%	1,5%	12,7%

Table 3.10 : Comparaison des performances moyennes du modèle Mack-Net (en milliers €)

Métriques	Mack-Net	*DeepTriangle*
CDR_{score}	697	1776
AvE_{score}	1028	1631
$MAPE^{\text{ult}}$	2,7%	1,7%
$MAPE^{\text{1yr}}$	2,6%	2,7%
MAE^{1yr}	243	257
MAE^{ult}	457	354

TABLE 3.11 : Comparaison entre Mack-Net et *DeepTriangle* sur une partie des données (triangle de taille 10x10)

Ainsi, ce chapitre permet d'explorer le modèle de provisionnement Mack-Net, modèle « branché »sur les hypothèses de Mack, basé sur des techniques de *Deep Learning* et pouvant clairement concurrencer le modèle de Mack Chain-Ladder. L'intérêt du modèle est qu'il offre une nouvelle approche pour le calcul des facteurs de développements et de volatilité et ainsi pour le calcul de paramètres USP. En minimisant les métriques adéquates sur l' ensemble d'entraînement, le paramètre USP finalement obtenu peut être fiabilisé grâce à des performances satisfaisantes. Quant à la métrique CDR_{score} utilisée comme métrique d'entrainement, E. Ramos-Pérez donne son avis : « *I have never seen this concept (CDR) used as an error measure for optimizing a reserving models based on machine learning methods. CDR can be very useful to compare the one year ahead performance of different models and it can be good as an error function but I have no references of the performance of this metric for hyperparameter optimization in neural networks. Conceptually it can make sense »(Source : échanges par mail avec E. Ramos-Pérez, 28 août 2024, 10h38)*. Enfin, d'après la table 3.12, bien que le temps d'une itération pour le modèle Mack-Net soit moindre qu'un modèle *DeepTriangle*, il reste quand même coûteux en temps par rapport aux méthodes déterministes comme le modèle de Mack qui fournit des résultats instantanément.

Mack-Net	bCCNN	*DeepTriangle*
1 minute	30 secondes	5 minutes

TABLE 3.12 : Temps d'itération des modèles avec CPU 12th Gen Intel(R) Core(TM) i5-1235U

3.6 Limites des travaux

Le présent mémoire possède évidemment des limites, notamment, en premier lieu le non-respect des contraintes d'un dossier USP (prérequis présentés en section 1.2). En matière de données, l'absence de données ligne à ligne empêche d'effectuer une analyse de marché complète entre les méthodes de provisionnement agrégé et individuel (en particulier, les méthodes d'apprentissage statistique étant généralement friandes en termes de volume de données). Cette absence de données ligne à ligne empêche également de segmenter les sinistres, d'inflater les triangles de liquidation (les sinistres ouverts étant évalués à l'euro d'aujourd'hui). L'absence de triangles de recours empêche enfin une estimation précise des provisions dossier/dossier, ce qui vient biaiser l'estimation et ajouter de la volatilité à certaines méthodes comme la méthode ECLRM, où les provisions dossier/dossier représentent la mesure de risque principale.

Par rapport aux méthodes utilisées (dont les limites sont indiquées en table 3.13), l'apprentissage profond nécessite souvent des données volumineuses et diversifiées pour entraîner efficacement les réseaux de neurones sous-jacents, ce qui peut être un défi pour les petites compagnies comme

des captives. Les risques consécutifs peuvent être du sous ou sur-apprentissage menant à de mauvaises performances. Le modèle Mack-Net, apparaît comme particulièrement complexe et difficile à interpréter, ce qui peut rendre son diagnostic difficile. La non-convergence vers 1 du dernier facteur de développement du modèle dans le chapitre 3 peut dans une certaine mesure être considérée comme une limite de plus. Les méthodes déterministes ou paramétriques comme Chain-Ladder ou encore le GLM quasi-Poisson présentent en ce sens, auprès du marché et du superviseur, l'avantage d'être facilement interprétables par rapport à des méthodes avec apprentissage profond.

En terme de résultats obtenus, bien que le modèle Mack-Net fournisse des résultats convaincants, les autres méthodes par apprentissage profond produisent des résultats plus médiocres. Le modèle bCCNN en témoigne puisqu'il conduit à la deuxième plus mauvaise performance dans la table 3.10 par rapport à la métrique CDR_{score}, et à la plus mauvaise en terme de MAE^{ult}.

Méthodes	**Mack-Net**	**Mack**	**MCL**	**bCCNN**	**ECLRM**	***DeepTriangle***
Données	Payés, charges et vecteur d'exposition	Payés	Payés et charges	Payés	Payés et provisions dossier/-dossier	Payés, provisions dossier/dossier, vecteur d'exposition et base NAIC
Performance CDR_{score}	1	2	3	4	5	6
Limites	Accès au vecteur des primes, dernier facteur différent de 1	Méthode standard	Méthode standard	Pas d'amélioration	Accès aux provisions dossier/-dossier	Longueur d'historique, pré-entraînement sur données externes, accès aux provisions dossier/dossier

TABLE 3.13 : Synthèse des données d'entraînement, ordre de performances et limites des méthodes du *benchmark*

Conclusion

Dans le cadre du calcul du SCR pour le risque de réserve d'une captive, le présent travail met en lumière les méthodes susceptibles de se substituer au paramètre imposé par la Directive Solvabilité II. Dans un contexte où de plus en plus de captives d'assurance ou de réassurance voient le jour sur le territoire national et où la Directive devient moins contraignante pour ces dernières avec notamment une refonte du principe de proportionnalité, le but du mémoire est d'explorer de nouvelles méthodes pour estimer le paramètre USP des réserves de la captive, notamment via les méthodes d'apprentissage profond. Il s'avère que le paramètre obtenu avec le modèle Mack-Net est plus précis que l'approche réglementaire, car le modèle permet de réduire l'erreur liée au calcul de la métrique CDR, fondamentale dans le calcul du paramètre USP. Cependant, la difficulté en matière de qualité des données et de vérification d'hypothèses rend l'approbation par le superviseur du paramètre USP dans la pratique plus ardue. Il existe aujourd'hui sur le marché une variété importante de méthodes pour quantifier la volatilité des réserves à un an, mais le mémoire a pour objectif d'étudier la contribution des méthodes de *Machine Learning* comme celles avec les réseaux de neurones récurrents. La question de l'interprétabilité de tels modèles demeure toutefois légitime.

Le mémoire contribue à promouvoir davantage les méthodes de *Deep Learning* sur des sujets réglementaires. A la lumière de la problématique, les résultats montrent que le modèle Mack-Net surperforme les méthodes standards de marché en fournissant un paramètre USP plus robuste, et donc, d'un point de vue métier, une estimation du SCR primes et réserves plus précise pour la captive. L'atout distinctif de cette méthode est qu'elle combine à la fois la simplicité d'utilisation, puisqu'elle ne nécessite pas les données individuelles, et la précision apportée par les réseaux de neurones. Le mémoire met ainsi en lumière une des rares méthodes de provisionnement agrégé par apprentissage profond, qui, via un entraînement selon la métrique construite souhaitée, permet de montrer que le *Machine Learning* améliore la prédictibilité des réserves.

Enfin, la validation d'un tel modèle de *Deep Learning* requiert parfois l'utilisation d'outils d'interprétabilité. En actuariat, certains modèles sont intrinsèquement interprétables mais certains modèles boîte noire (Rudin, 2019) nécessitent l'emploi de méthodes *post-hoc* pour signifier les interactions entre variables d'entrée, comme par exemple avec l'approche *SHapley Additive exPlanation* (SHAP). Au-delà de la prédiction de *Best estimates*, l'actuaire doit également pouvoir quantifier l'incertitude du modèle de *Deep Learning* (initialisation des poids, *dropout*,...) et essayer de la limiter au maximum. La technique d'aggrégation en réseaux de neurones, ou *nagging* (Richman et Wüthrich, 2020), correspondant à moyenner les prédictions de plusieurs réseaux indépendants, comme dans le modèle Mack-Net, est un parfait exemple. Ainsi, compte tenu de l'application croissante de l'apprentissage profond aux problèmes actuariels, il devient crucial pour l'actuaire de rester à jour dans ces domaines de pointe. Avec l'émergence de l'IA générative et l'opportunité d'exploitation pour l'actuaire, on pourrait imaginer dans les prochaines années des IA permettant d'interpréter les modèles de *Data Science* les moins interprétables à l'heure actuelle. Dans le contexte des captives, l'IA représente un défi, notamment pour les gestionnaires de captives de pouvoir leur transférer des données conformes afin d'optimiser les modèles de projection.

Bibliographie

ACPR (2023). Exigences en matière de qualité des données pour les organismes et groupes d'assurance soumis à la Directive Solvabilité 2. Rapp. tech. Banque de France. URL : https://acpr.banque-france.fr/sites/default/files/media/2023/12/01/20231201_notice_qdd_s2.pdf.

Actuelia (2015). La Qualité des Données sous Solvabilité II. Webinar. URL : https://www.actuelia.fr/qualite-des-donnees.

Actuelia (2024). Revoyure 2020 : Quid du principe de proportionnalité? Article en ligne. URL : https://www.actuelia.fr/post/revoyure-2020-quid-du-principe-de-proportionnalit%C3%A9.

Al-Mudafer, M., Avanzi, B., Taylor, G. et Wong, B. (2022). Stochastic loss reserving with mixture density neural networks. *Insurance: Mathematics and Economics* Volume 105.July.

Allaire, J. et Chollet, F. (2024). keras: R Interface to 'Keras'. R package version 2.15.0. URL : https://CRAN.R-project.org/package=keras.

Allianz (2023). The ART of Captives. Étude. URL : https://commercial.allianz.com/content/dam/onemarketing/commercial/commercial/solutions/commercial-art-captives.pdf.

Autorité de Contrôle Prudentiel et de résolution (2024). Revue de la Directive Solvabilité II : vers un régime proportionné. *Revue de l'ACPR* Volume 2024.N°61.

Badinter (1985). Loi n° 85-677 du 5 juillet 1985 tendant à l'amélioration de la situation des victimes d'accidents de la circulation et à l'accélération des procédures d'indemnisation. Loi ordinaire. URL : https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000693454.

Balona, C. et Richman, R. (2020). The Actuary and IBNR Techniques: A Machine Learning Approach. *Actuarial Society of South Africa* 2020.Octobre.

Baudry, M. et Robert, C. (2019). A machine learning approach for individual claims reserving in insurance. *Applied Stochastic Models in Business and Industry* Volume 35.N°1.

Bechouche, A. (2013). Utilisation des techniques avancées pour l'observation et la commande d'une machine asynchrone : application à une éolienne. Mém. de mast. UMMTO. URL : https://www.researchgate.net/figure/Fonctions-dactivation-dun-neurone-artificiel_fig2_322194356.

Boumezoued, A et Devineau, L. (2011). One-year reserve risk including a tail factor : closed formula and bootstrap approaches. *arXiv* Volume 2011.Numéro de juillet.

Budhiraja, A. (2016). Dropout in (Deep) Machine learning. Blog Medium. URL : https://medium.com/@amarbudhiraja/https-medium-com-amarbudhiraja-learning-less-to-learn-better-dropout-in-deep-machine-learning-74334da4bfc5.

Byrd, R., Lu, P., Nocedal, J. et Zhu, C. (1994). A limited memoryalgorithm for bound constrained optimization. Rapp. tech. Northwestern University, Department of Electrical Engineering et Computer Scienc. URL : https://users.iems.northwestern.edu/~nocedal/PDFfiles/limited.pdf.

Carnevale, G. et Clemente, G. (2020). A Bayesian Internal Model for Reserve Risk: An Extension of the Correlated Chain Ladder. *Risks* Volume 8.N°125.

Cavastracci, S et Tripodi, A. (2018). Overdispersed-Poisson Model in Claims Reserving: Closed Tool for One-Year Volatility in GLM Framework. *Risks* Volume 2018-6(4).N°139.

Cerchiara, R. et Magatti, V. (2013). Undertaking Specific Parameters or a Partial Internal Model under Solvency 2? Présentation lors du 30ème Congrès International des Actuaires. url : https://www.actuaries.org/dc2014/Handout/Paper1796/HANDOUT_ICA2014_Cerchiara_Magatti_def.pdf.

CNRS (2022). Introduction au Deep Learning - Données séquentielles et/ou temporelles. Réseaux de Neurones Récurrents (RNN). Formation Fidle portée par l'institut d'Intelligence Artificielle MIAI de Grenoble et le CNRS. url : https://cloud.univ-grenoble-alpes.fr/index.php/s/wxCztjYBbQ6zwd6?dir=undefined&path=%2FSaison%202022-2023&openfile=565851274.

Commission Européenne (14 mars 2014). Draft Delegated Acts Solvency II. Rapp. tech. Commission Européenne.

Commission Européenne (2010). QIS5 Technical Specifications. Rapp. tech. CEIOPS. url : https://www.knf.gov.pl/knf/pl/komponenty/img/technical_specifications_en_23614.pdf.

Compagnie Nationale des Services de Conseil en Risques & Assurances (2019). Société à Compartiments Protégés. Formation DDA. url : https://www.cnscra.fr/pcc.html.

Cramér, H (1946). Mathematical Methods Of Statistics. Princeton University Press.

Dahms, R. (2021). Stochastic Reserving. Présentation à l'ETH Zurich, Printemps 2021. url : https://ethz.ch/content/dam/ethz/special-interest/math/risklab-dam/documents/Lectures/Stochastic-Reserving-2021.pdf.

Dahms, R, Merz, M. et Wüthrich, M. (2009). Claims Development Result for Combined Claims Incurred and Claims Paid Data. *Bulletin Français d'Actuariat* Volume 9.N°18.

Delong, L et Szatkowski, M. (2021). One-year and ultimate reserve risk in Mack Chain Ladder model. *Risks* Volume 2021.N°9.

Denuit, M., Hainaut, D. et Trufin, J. (2019). Effective Statistical Learning Methods for Actuaries I: GLMs and Extensions. Springer International Publishing (Springer Actuarial).

Diers, D (2009). Stochastic re-reserving in multi-year internal models – An approach based on simulations. Colloque de l'ASTIN, Helsinki, Finland, June 1–4. url : https://www.actuaries.org.uk/system/files/field/document/S4_11_Diers.pdf.

Djuikem, C. (2024). Les Maths Derrière les Réseaux Neuronaux. https://www.clotilde-djuikem.com/.

EIOPA (2011). Report of the Joint Working Group on Non-Life and Health Non-SLT Calibration, 11/163. Rapp. tech. EIOPA. url : https://register.eiopa.europa.eu/Publications/Reports/EIOPA-11-163-A-Report_JWG_on_NL_and_Health_non-SLT_Calibration.pdf.

EIOPA (2015). Règlements Délégués (UE) 2015/35 de la Commission du 10 octobre 2014, Annexe XVII, Alinéas C et D. Rapp. tech. EIOPA. url : https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:32015R0035.

EIOPA (2024). Opinion on the supervision of captive (re)insurance undertakings : Cash pooling, Prudent Person Principle and Governance. Rapp. tech. EIOPA. url : https://www.eiopa.europa.eu/document/download/71e1ff0c-31d6-4226-b1de-80975299a86d_en?filename=EIOPA-BoS-24-176%20-%20Opinion%20on%20Supervision%20of%20Captives.pdf.

England, P. et Verrall, R. (2002). Stochastic claims reserving in general insurance. *Institute and Faculty of Actuaries*.

France Assureurs (2023). L'assurance automobile des particuliers en 2023. Étude. url : https://www.franceassureurs.fr/wp-content/uploads/lassurance-automobile-des-particuliers-en-2023.pdf.

Gabrielli, A. (2019). A neural network boosted double overdispersed Poisson claims reserving model. *ASTIN Bulletin* Volume 50.N°1.

Gabrielli, A, Richman, R. et Wüthrich, M. (2019). Neural network embedding of the over-dispersed Poisson reserving model. *Scandinavian Actuarial Journal* Volume 2020.1-29.

Gesmann, M., Murphy, D., Zhang, Y. W., Carrato, A., Wuthrich, M., Concina, F. et Dal Moro, E. (2023). ChainLadder. R package version 0.2.18. url : https://CRAN.R-project.org/package=ChainLadder.

Glorot, X. et Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. *Proceedings of Machine Learning Research* Volume 9.

Goma, T. (2020). Apport des réseaux neuronaux au provisionnement IARD : Application d'architectures récurrentes à la prédiction d'évolution temporelle de montants. Mémoire d'actuariat. ENSAE.

Gontier, D. (2021). Méthodes Numériques : Optimisation. Formation. url : https://www.ceremade.dauphine.fr/~gontier/Publications/methodesNumeriques.pdf.

Guy Carpenter (2021). EMEA Motor Study. Étude. url : https://www.guycarp.com/insights/2022/02/guy-carpenter-briefing-addresses-EMEA-motor-market.html.

Hoerl, A. et Kennard, R. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics* Volume 12.N°1.

Index Assurance (2023). Convention IRCA. Guide de l'assurance auto. url : https://www.index-assurance.fr/pratique/sinistre/convention-irca.

Institut des Actuaires (2021). Note technique - Groupe de travail Captives. Étude. url : https://www.institutdesactuaires.com/global/gene/link.php?doc_id=17059&fg=1.

JAGS (2023). Just Another Gibbs Sampler. Programme de simulation pour les modèles bayésiens. url : https://mcmc-jags.sourceforge.io/.

James, G., Witten, D. et Hastie T., T. R. (2013). An introduction to statistical learning. *Springer* Volume 112.

Kidger, P. et Lyons, T. (2020). Universal approximation with deep narrow networks. *Proceedings of Machine Learning Research* Volume TBD.

Kingma, D. et Ba, J. (2017). Adam: a method for stochastic optimization. *arXiv* Volume 2017.conference paper at ICLR 2015.

Klugman, S. A., Panjer, H. H. et Willmot, G. E. (2012). Loss Models: From Data to Decisions. 4ème édition. Society of Actuaries.

Kremer, E. (1982). IBNR Claims and the Two-Way Model of ANOVA. *Scandinavian Actuarial Journal* Volume 1982.N°1.

Kremer, E. (1985). Einführung in die Versicherungsmathematik. *ASTIN Bulletin* Volume 18.N°1.

Kuo, K. (2018). DeepTriangle: A Deep Learning Approach to Loss Reserving. *Risks* Volume 2019.N°97.

Lacoume, A. (2009). Mesure du risque de réserve sur un horizon de un an. Mémoire d'actuariat. Institut de Science Financière et d'Assurances.

L'Argus de l'assurance (2025). Solvabilité 2, une révision... et des questions. *L'Argus de l'assurance* Volume 2025.Numéro du 03 mars 2025.

Llosa, M., Massias, M. et Pillaudin, M. (2015). Journées d'étude IARD. Retour d'expérience projet USP. url : https://www.institutdesactuaires.com/global/gene/link.php?doc_id=3605&fg=1.

Mack, T. (1991a). A Simple Parametric Model for Rating Automobile Insurance or Estimating IBNR Claims Reserves. *ASTIN Bulletin* Volume 21.N°93.

Mack, T. (1991b). A simple parametric model for rating automobile insurance or estimating IBNR claims reserves. *ASTIN Bulletin* Volume 21.N°1.

Mack, T. (1993). Measuring the Variability of Chain Ladder Reserve Estimates. Faculty et Institute of Actuaries, Volume 2, Section D6.

Marsh Captive Solutions (2023). 2023 Captive Landscape. Étude. url : https://www.marsh.com/tw/en/services/captive-insurance/insights/captive-insurance-market-report-2023.html.

Merz, M et Wüthrich, M. (2008). Modelling the Claims Development Result for Solvency Purposes. Article du Forum Casualty Actuarial Society, Édition Automne 2008. url : https://www.casact.org/sites/default/files/database/forum_08fforum_21merz_wuetrich.pdf.

Meyers, G. (2015). Stochastic Loss Reserving Using Bayesian MCMC Models. *Casualty Actuarial Society Monograph Series* 2015.N°1.

Mulquiney, P. (2006). Artificial Neural Networks in Insurance Loss Reserving. *Atlantis Press* Volume JCIS-06.October 2006.

National Association of Insurance Commissioners (2015). Loss Reserving Data Pulled from NAIC Schedule P. Research Resources from the Casualty Actuarial Society.

Ndao, A. (2018). Calcul des USP en Santé Non SLT. Mémoire d'actuariat. Institut de Science Financière et d'Assurances.

Nouar, S. (2015). Méthode de provisionnement ligne-à-ligne en assurance non-vie. Mémoire d'actuariat. Institut de Science Financière et d'Assurances.

OpenAI (2024). ChatGPT. url : https://openai.com/.

Ortiz, L. (2019). Eléments d'Intelligence Artificielle faible en Provisionnement Non-Vie. Mémoire d'actuariat. ENSAE.

Parlement Européen et Conseil de l'Union Européenne (2009). Directive 2009/138/CE du 25 novembre 2009 sur l'accès aux activités de l'assurance et de la réassurance et leur exercice (solvabilité II). OJ L. 335/I.

Pillaudin, M (2013). Modélisation du Capital économique non-vie sous Solvabilité II. Lettre Actuariat Finance et Finance d'Optimind Winter, 2ème semestre 2013. url : https://www.optimind.com/medias/documents/265/201307_Lettre_Actuariat_et_finance_Second_semestre_2013_VF.pdf.

Pinkus, A. (1999). Approximation theory of the MLP model in neural networks. *Actanumerica* Volume 8.

Pittarello, G. (2023). Chain Ladder Plus: a versatile approach for claims reserving. Présentation à La Sapienza, Università di Roma. url : https://insurancedatascience.org/downloads/London2023/Gabriele_Pittarello.pdf.

Planchet, F. et Leroy, G. (2022). Barème de capitalisation 2022. Rapp. tech. La Gazette du Palais. url : https://lp.gazette-du-palais.fr/bareme-de-capitalisation.

Poon, J. (2019). Penalising Unexplainability in Neural Networks for Predicting Payments per Claim Incurred. *Risks* Volume 7.N°3.

Quarg, G. et Mack, T. (2004). Munich Chain Ladder: A Reserving Method that Reduces the Gap between IBNR Projections Based on Paid Losses and IBNR Projections Based on Incurred Losses. *Casualty Actuarial Society* Volume 2.N°2.

R core team (2023). R: A Language and Environment for Statistical Computing. Vienna, Austria : R Foundation for Statistical Computing. url : https://www.R-project.org/.

Ramos-Pérez (2024). MackNet: Mack-Net model: Blending Mack's model with Recurrent Neural Networks. R package version 1.0.0.

Ramos-Pérez, E, Alonso-González, P. et Núñez-Velázquez, J. (2022). Mack-Net model: Blending Mack's model with Recurrent Neural Networks. *arXiv* Volume 201.N°117146.

Reddi, S., Kale, S. et Kumar, S. (2019). On the Convergence of Adam and Beyond. *arXiv* Volume 2019.International Conference on Learning Representations.

Renshaw, E et Verrall, R. (1998). A stochastic model underlying the chain-ladder technique. *British Actuarial Journal* Volume 4.N°4.

Richman, R. et Wüthrich, M. V. (2020). Nagging Predictors. *Risks* 2020.Aout.

Rossouw, L. et Richman, R. (2019). Using Machine Learning to Model Claims Experience and Reporting Delays for Pricing and Reserving. *SSRN* Volume 2019.October.

Royer, C. (2022). Optimisation pour l'apprentissage automatique. Formation. Université Dauphine Tunis. url : https://www.lamsade.dauphine.fr/~croyer/ensdocs/TUN/PolyTUN.pdf.

Rudin, C. (2019). Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *arXiv* 2019.September.

Saxe, A., McClelland, J. et Ganguli, S. (2014). Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. *arXiv* Volume 2013.N°3.

Siegenthaler, F., Demarco, V. et Cerchiara, R. (2017). On the USP Calculation under Solvency II and its Approximation with Closed Form Formula. *Bulletin Français d'actuariat* Volume 17.N°33.

Souter, G. (2023). Captive Insurance Special Report. *Business Insurance* Volume 2023.Numéro de mars.

Suru, A. (2012). Assurance Dommages : les dessous d'un secteur qui vous protège. Collection Assurance-Audit-Actuariat. Economica.

Swiss Re (2022). Motor Bodily Injury Landscape. Étude. url : https://www.swissre.com/reinsurance/insights/motor-bodily-injury-landscape.html.

Tapsoba Wendinmanedge, A. (2021). Application de l'apprentissage profond au provisionnement agrégé non-vie. Mémoire d'actuariat. ENSAE.

Taylor, G. (2019). Granular and machine learning forms Risks. *Risks* Volume 2019.82.

Titon, E. et Talbi, T. (2024). Individual claim reserving: A complementary approach to aggregated methods. Milliman White Paper. url : https://fr.milliman.com/-/media/milliman/pdfs/2024-articles/2-29-24_individual-claims-reserving.ashx.

Vadjoux, T. (2021). Des assureurs flirtent avec la limite de solvabilité. *L'Agefi* Volume 2021.N°4.

Verrall, R. (2000). An Investigation into Stochastic Claims Reserving Models and the Chain-Ladder Technique. *Insurance: Mathematics and Economics* Volume 26.N°1.

Wikipédia (2023). Limited-memory BFGS. Page Wikipédia. url : https://en.wikipedia.org/wiki/Limited-memory_BFGS.

Wüthrich, M. V. et Merz, M. (2015). Stochastic Claims Reserving Manual: Advances in Dynamic Modeling. *Swiss Finance Institute* N°15-34.

Wüthrich, M. (2018). Neural networks applied to chain-ladder reserving. *European Actuarial Journal* Volume 23.N°2.

Zimbidis, A. A. (2021). Solvency II, Undertaking Specific Parameters (USPs) Validation, Generalization and Criticism. *Contemporary Mathematics* Volume 2.N°2.

Annexe A

Annexes

A.1 Utilisation de l'IA générative

L'IA générative comme ChatGPT (OpenAI, 2024) a pu être utilisée partiellement dans ce mémoire à des fins de reformulation.

A.2 Aboutissement à la formule de la méthode 1 des actes délégués

La méthode 1 (log-normale).

Démonstration. La démonstration est issue de Zimbidis (2021). La formule telle que définie dans les actes délégués suppose que

$$y_t \sim \mathcal{LN}(\mu, \nu).$$

La fonction densité pour y s'écrit donc

$$f(y) = \frac{1}{y\sqrt{2\pi\nu}} e^{-\frac{(\ln y - \mu)^2}{2\nu}}, \quad y > 0,\, y > 0.$$

Avec les hypothèses *4* et *5* de la section 1.2.2,

$$\begin{cases} \mathbb{E}(y_t) = \beta x_t, \\ \mathbb{V}(y_t) = e^{2\mu+\nu}[e^{\nu} - 1] = \sigma^2[(1-\delta)\bar{x_t} + \delta x_t^2]' \end{cases}$$

La résolution du système de deux équations à deux inconnues aboutit à

$$\begin{cases} \mu = \ln(\beta x_t) - \dfrac{1}{2\pi_t} \\ \nu = \dfrac{1}{\pi_t} \end{cases},$$

en définissant

$$\pi_t = \frac{1}{\ln\left[1 + \frac{\sigma^2}{\beta^2}(1-\delta)\frac{\bar{x}}{x_t} + \delta\right]}.$$

En posant

$V_t = \ln(y_t) - \ln(\beta x_t)$, on a alors la densité de V selon une loi normale centrée : $f_{V_t}(v) = \frac{1}{\sqrt{2\pi\nu}} e^{-\frac{1}{2\nu}v^2}$.

La vraisemblance vaut alors

$$L = \prod_{t=1}^{T} f(V_t) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\nu}} e^{-\frac{1}{2\nu}v_t^2}$$

S'en suit l'écriture de la log-vraisemblance

$$l = \ln(L) = \ln\left[\left(\frac{1}{\sqrt{2\pi}}\right)^T \left(\prod_{t=1}^{T}\pi_t\right)^{\frac{1}{2}} e^{-\frac{1}{2}\sum_{t=1}^{T}\pi_t v_t^2}\right].$$

Avec une dérivée seconde négative, la log-vraisemblance admet un maximum ou de manière équivalente un minimum de

$$l' = \sum_{t=1}^{T}\pi_t v_t^2 - \sum_{t=1}^{T}\ln\pi_t \quad . \tag{A.1}$$

Ainsi, en posant $\beta = \frac{\sigma}{e^\gamma}$ le ratio de *run-off*, on obtient dans l'équation (A.1) une fonction paramétrique en β

$$l'(\beta,\gamma,\delta) = \sum_{t=1}^{T}\pi_t(\gamma,\delta)\left(\ln\left(\frac{y_t}{x_t}\right) + \frac{1}{2\pi_t(\gamma,\delta)} - \ln\beta\right)^2 - \sum_{t=1}^{T}\ln\pi_t(\gamma,\delta).$$

Enfin, en supposant que $l'(\ln\beta,\gamma,\delta) \bmod l'(\beta,\gamma,\delta)$, on dérive

$$\frac{dl'(\ln\beta)}{d\ln\beta} = 0 \Rightarrow \sum_{t=1}^{T}\pi_t(\gamma,\delta)\ln\left(\frac{y_t}{x_t}\right) + \frac{T}{2} - \ln\beta\sum_{t=1}^{T}\pi_t(\gamma,\delta) = 0 \Rightarrow \ln\hat{\beta} = \frac{\frac{T}{2} + \sum_{t=1}^{T}\pi_t(\gamma,\delta)\ln\left(\frac{y_t}{x_t}\right)}{\sum_{t=1}^{T}\pi_t(\gamma,\delta)}.$$

On obtient donc bien un minimum de la fonction l', puis en remplaçant par la définition du ratio de *run-off*, on obtient finalement

$$\sum_{t=1}^{T}\pi_t(\gamma,\delta)\left(\ln\left(\frac{y_t}{x_t}\right) + \frac{1}{\pi_t(\gamma,\delta)} + \gamma - \ln\left(\hat{\sigma}(\gamma,\delta)\right)\right)^2 - \sum_{t=1}^{T}\ln\left(\pi_t(\gamma,\delta)\right).$$

L'équation (1.8) est donc vérifiée. □

A.3 Aboutissement à la formule de la méthode 2 des actes délégués

Pour passer des équations (1.3) à (1.4), une série de notations et d'approximations intermédiaires qu'il convient de rappeler est employée. L'intuition de la démonstration est issue de Merz et Wüthrich (2008) et de la version *draft* des actes délégués (Commission Européenne, 14 mars 2014).

Démonstration.

$$\hat{\Delta}_{i,J} = \frac{\hat{\sigma}^2_{I-i}}{\hat{f}^I_{I-i}{}^2 S^I_{I-i}} + \sum_{k=i}^{J-1} \frac{C^2_{I-j,j}\hat{\sigma}^2_j}{S^{I+1}_j \hat{f}^{I\,2}_j S^I_j},$$

$$\hat{\Phi}^I_{i,J} = \begin{cases} 0 & \text{si } i = 0, \\ \left[1 + \frac{\hat{\sigma}^{I\,2}_{I-i}}{(\hat{f}^I_{I-i})^2 C_{i,I-i}}\right] \left(\prod_{l=I-i+1}^{J-1} \left(1 + \frac{\hat{\sigma}^{I\,2}_l}{(\hat{f}^I_l)^2 (S^{I+1}_l)^2} C_{I-l,l}\right) - 1\right) & \\ \quad \approx \sum_{j=I-i+1}^{J-1} \left(\frac{C_{I-j,j}}{S^{I+1}_j}\right)^2 \frac{\hat{\sigma}^{I\,2}_j}{(\hat{f}^I_j)^2 C_{I-j,j}} & \text{si } i > 0. \end{cases}$$

$$\hat{\Psi}^I_{i,k} = \begin{cases} \hat{\Psi}^I_{i,1} = 0 & \text{si } i > 0 \\ \hat{\Psi}^I_{i,k} = \left(1 + \frac{\hat{\sigma}^I_{I-i}{}^2}{(\hat{f}I_{I-i})^2 S^{I+1}_{I-i}}\right) \left(1 + \frac{\hat{\sigma}^I_{I-i}{}^2}{(\hat{f}^I_{I-i})^2 C_{i,I-i}}\right)^{-1} \hat{\Phi}^I_{i,J} \approx \frac{\hat{\sigma}^2_{I-i}/(\hat{f}^I_{I-i})^2}{C_{i,I-i}} & \text{pour } k > i > 0, \end{cases}$$

$$\Lambda_{i,J} = \frac{C_{i,I-i}}{S^{I+1}_{I-i}} \frac{(\hat{\sigma}^I_{I-i})^2}{(\hat{f}^I_{I-i})^2} + \sum_{j=I-i+1}^{J-1} \left(\frac{C_{I-j,j}}{S^{I+1}_j}\right)^2 \frac{(\hat{\sigma}^I_j)^2}{(\hat{f}^I_j)^2 S^I_j},$$

$$\hat{\Gamma}^I_{i,J} = \hat{\Phi}^I_{i,J} + \hat{\Psi}^I_i = \left\{ \left[\left(1 + \frac{(\hat{\sigma}^I_{I-i})^2}{(\hat{f}^I_{I-i})^2}\right) \frac{1}{C_{i,I-i}}\right] \cdot \prod_{l=I-i+1}^{J-1} \left(1 + \frac{(\hat{\sigma}^I_l)^2}{(\hat{f}^I_l)^2 (S_l)^2} C_{I-l,l}\right) - 1 \right\} \approx \hat{\Phi}^I_{i,J} + \frac{(\hat{\sigma}^I_{I-i})^2}{(\hat{f}^I_{I-i})^2 S^{I+1}_{I-i}},$$

$$\hat{\Upsilon}^I_{i,k} = \widehat{\mathrm{Cov}}\left(\widehat{\mathrm{CDR}}_i(I+1), \widehat{\mathrm{CDR}}_k(I+1) \mid D_I\right)$$

$$= \left\{ \left(\left(1 + \frac{(\hat{\sigma}^I_{I-k})^2}{(\hat{f}^I_{I-k})^2}\right)^2 \frac{1}{S^{I+1}_{I-k}}\right) \cdot \prod_{l=I-k+1}^{J-1} \left(1 + \frac{(\hat{\sigma}^I_l)^2}{(\hat{f}^I_l)^2 (S^{I+1}_l)^2} C_{I-l,l}\right) - 1 \right\}$$

$$\approx \hat{\Phi}^I_{i,J} + \frac{(\hat{\sigma}^I_{I-i})^2}{(\hat{f}^I_{I-i})^2 S^{I+1}_{I-i}}.$$

En faisant l'hypothèse que

$$\mathbb{V}(\mathrm{CDR}_i(I+1)|D_I) = (\hat{C}^I_{i,J})^2 \hat{\Psi}^I_i,$$

et que

$$\widehat{\mathrm{MSEP}}_{\widehat{\mathrm{CDR}}_i(I+1)|D_I}\left(\widehat{\mathrm{CDR}}_i(I+1)\right) = (\hat{C}_{i,n})^2 \left(\hat{\Phi}^I_{i,J} + \hat{\Delta}^I_{i,J}\right),$$

ils estiment que l'équation (1.3) devient

$$\mathrm{MSEP}_{\mathrm{CDR}_i(I+1)|D_i}(0) = \underbrace{\mathbb{V}(\mathrm{CDR}_i(I+1) \mid D_I)}_{\text{erreur de processus}} + \underbrace{\widehat{\mathrm{MSEP}}_{\mathrm{CDR}_i(I+1)|D_I}\left(\hat{\mathrm{CDR}}_i(I+1)\right)}_{\text{erreur d'estimation}} = (\hat{C}_{i,J})^2 \left(\hat{\Gamma}^I_{i,J} + \hat{\Delta}^I_{i,J}\right).$$

Enfin, ils estiment que l'équation (1.4) devient

$$\widehat{\text{MSEP}}_{\sum_{i=1}^{I}\widehat{\text{CDR}}_i(I+1)|D_I}(0) = \sum_{i=1}^{I}(\hat{C}_{i,J})^2\left(\hat{\Phi}^I_{i,J}+\hat{\Delta}^I_{i,J}\right)+2\sum_{k>i>0}\hat{C}_{i,J}\hat{C}_{k,J}\left(\hat{\Upsilon}^I_{i,k}+\hat{\Gamma}^I_{i,J}\right).$$

En notant

$$\hat{Q}_j = \frac{\hat{\sigma}^2_j}{\hat{f}^2_j}, \quad S_j = S^I_j, \quad \text{et} \quad S'_j = S^{I+1}_j,$$

la formule de la MSEP ainsi obtenue est

$$\text{MSEP} = \sum_{i=1}^{I}\hat{C}_{i,J}\left(\frac{\hat{Q}_{I-i}}{C_{i,I-i}}+\frac{\hat{Q}_j}{S_i}\right)+\sum_{j=I-i+1}^{J-1}\frac{C_{I-j,j}}{S^{I+1}_j}\hat{Q}_j+2\sum_{i=1}^{I}\sum_{k=i+1}^{I}\hat{C}_{i,J}\hat{C}_{k,J}\left(\frac{\hat{Q}_{I-i}}{S_i}+\sum_{j=I-i+1}^{J-1}\frac{C_{I-j,j}}{S^{I+1}_j}\hat{Q}_j\right).$$

Par symétrie des coefficients croisés $\hat{C}_{i,J}\hat{C}_{k,J}$, on obtient ainsi la formule (1.9)

$$\text{MSEP} = \sum_{i=1}^{I}\hat{C}^2_{i,J}\cdot\frac{\hat{Q}_{I-i}}{C_{i,I-i}}+\sum_{i=1}^{I}\sum_{k=1}^{I}\hat{C}_{i,J}\cdot\hat{C}_{k,J}\cdot\left(\frac{\hat{Q}_{I-i}}{S_{I-i}}+\sum_{j=I-i+1}^{J-1}\frac{C_{I-j,j}}{S^{I+1}_j}\cdot\frac{\hat{Q}_j}{S_j}\right).$$

□

A.4 Preuves du modèle linéaire généralisé (GLM) Poisson sur-dispersé

A.4.1 Proposition 2.4

Démonstration.

$$\text{MSEP}(\tilde{Y}_{ij}) = \mathbb{V}(\hat{Y}_{ij})+\mathbb{E}\left[(\tilde{Y}_{ij}-\mathbb{E}[\hat{Y}_{ij}])^2\right] \approx \mathbb{V}(Y_{ij})+\mathbb{V}(\tilde{Y}_{ij}),$$

Or, l'approximation de Taylor donne

$$\mathbb{V}(\tilde{Y}_{ij}) = \mathbb{V}(h(\tilde{\eta}_{ij})) \approx [h'(\tilde{\eta}_{ij})]^2\mathbb{V}(\tilde{\eta}_{ij}),$$

et, dans sa forme compacte,

$$\widehat{\mathbb{V}}(\tilde{\eta}_{ij}) = x^\top_{ij}\mathbb{V}(\tilde{\beta}_{ij})x_{ij}.$$

Ainsi avec l'hypothèse *12*, on a

$$\text{MSEP}(\hat{Y}_{ij}) = \phi\hat{\mu}_{ij}+[h'(\hat{\eta}_{ij})]^2\mathbb{V}(\hat{\eta}_{ij}).$$

□

Démonstration. On a

$$\text{MSEP}(\tilde{R}) = \mathbb{E}[(R-\mathbb{E}(R)+\mathbb{E}(R)-\tilde{R})^2] \approx \mathbb{E}[(R-\mathbb{E}(R))^2]+\mathbb{E}[(\tilde{R}-\mathbb{E}(R))^2] \approx \mathbb{V}(R)+\mathbb{V}(\tilde{R}),$$

où $\hat{R}_i = \sum_{j=i+1}^{J}\hat{Y}_{ij}$, $\mathbb{V}(R)$ représente l'erreur de processus et $\mathbb{V}(\tilde{R})$ représente l'erreur d'estimation. D'où

$$\text{MSEP}(\hat{R}_i) = \mathbb{V}(R_i)+\mathbb{V}(\tilde{R}_i).$$

Par indépendance,

$$\widehat{\mathbb{V}}(R_i) = \sum_{j=i+1}^{J} \widehat{\mathbb{V}}(\hat{Y}_{ij}) = \phi \sum_{j=t-i+1}^{J} \hat{\mu}_{ij},$$

et

$$\widehat{\mathbb{V}}(\tilde{R}_i) = \sum_{j=t-i+1}^{J} \widehat{\mathbb{V}}(\tilde{Y}_{ij}) + \sum_{\substack{j_1,j_2=t-i+1 \\ j_1 \neq j_2}}^{J} \widehat{\mathrm{Cov}}(\tilde{Y}_{i,j_1}, \tilde{Y}_{i,j_2}).$$

Or, d'après la démonstration *supra*,

$$\widehat{\mathbb{V}}(\tilde{R}_i) = \sum_{j=t-i+1}^{J} \widehat{\mathbb{V}}(\tilde{Y}_{ij}) + \sum_{\substack{j_1,j_2=t-i+1 \\ j_1 \neq j_2}}^{J} \widehat{\mathrm{Cov}}(\tilde{Y}_{i,j_1}, \tilde{Y}_{i,j_2}),$$

et

$$\widehat{\mathrm{Cov}}(\tilde{\eta}_{i,j_1}, \tilde{\eta}_{i,j_2}) = x_{i,j_1}^{\top} \widehat{\mathbb{V}}(\tilde{\beta_{ij}}) x_{i,j_2}.$$

Avec

$$\mathbb{V}(\tilde{Y}_{ij}) = \mathbb{V}(h(\tilde{\eta}_{ij})) \approx [h'(\hat{\eta}_{ij})]^2 \mathbb{V}(\tilde{\eta}_{ij}),$$

on obtient que

$$\widehat{\mathbb{V}}(\tilde{R}_i) = \sum_{j=t-i+1}^{J} [h'(\hat{\eta}_{ij})]^2 \mathbb{V}(\tilde{\eta}_{ij}) + \sum_{\substack{j_1,j_2=t-i+1 \\ j_1 \neq j_2}}^{J} h'(\hat{\eta}_{i,j_1}) \cdot h'(\hat{\eta}_{i,j_2}) \cdot \widehat{\mathrm{Cov}}(\tilde{\eta}_{i,j_1}, \tilde{\eta}_{i,j_2}).$$

Avec la fonction de lien logarithme, la formule se réécrit

$$\widehat{\mathbb{V}}(R_i) = \sum_{j=t-i+1}^{J} \hat{\mu}_{ij}^2 \mathbb{V}(\tilde{\eta}_{ij}) + \sum_{\substack{j_1,j_2=t-i+1 \\ j_1 \neq j_2}}^{J} \hat{\mu}_{ij_1} \hat{\mu}_{ij_2} \widehat{\mathrm{Cov}}(\tilde{\eta}_{i,j_1}, \tilde{\eta}_{i,j_2}).$$

Pour rappel,

$$\widehat{MSEP}(\tilde{R}) = \widehat{\mathbb{V}}(R) + \widehat{\mathbb{V}}(\tilde{R}) \quad \text{avec} \quad \widehat{\mathbb{V}}(R) = \sum_{i+j>t} \widehat{\mathbb{V}}(Y_{ij}).$$

De manière similaire,

$$\widehat{\mathbb{V}}(\widetilde{R}) = \sum_{i+j>t} h'(\widehat{\eta}_{ij})^2 \widehat{\mathbb{V}}(\tilde{\eta}_{ij}) + \sum_{\substack{i_1+j_1>t \\ i_2+j_2>t \\ (i_1,j_1)\neq(i_2,j_2)}} h'(\widehat{\eta}_{i_1,j_1}) \cdot h'(\widehat{\eta}_{i_2,j_2}) \cdot \widehat{\mathrm{Cov}}(\tilde{\eta}_{i_1,j_1}, \tilde{\eta}_{i_2,j_2}),$$

et pour rappel, $\widehat{\mathbb{V}}(\tilde{\eta}_{ij}) = x_{ij}^{\top} \widehat{\mathbb{V}}(\tilde{\beta}_{ij}) x_{ij}$ donc par conséquent

$$\widehat{\mathrm{MSEP}}(\widetilde{R}_i) = \widehat{\phi} \sum_{j=t-i+1}^{J} \widehat{\mu}_{ij} + \sum_{j=t-i+1}^{J} \widehat{\mu}_{ij}^2 x_{ij}^{\top} \widehat{\mathbb{V}}(\tilde{\beta}) x_{ij} + \sum_{j_1,j_2=t-i+1}^{J} \widehat{\mu}_{ij_1} \widehat{\mu}_{ij_2} x_{i_1,j_1}^{\top} \widehat{\mathbb{V}}(\tilde{\beta}) x_{i_2,j_2}, \quad j_1 \neq j_2,$$

et

$$\widehat{\mathrm{MSEP}}(\widetilde{R}) = \widehat{\phi} \sum_{i+j>t} \widehat{\mu}_{ij} + \sum_{i+j>t} \widehat{\mu}_{ij}^2 x_{ij}^\top \widehat{\mathbb{V}}(\tilde{\beta}) x_{ij} + \sum_{\substack{i_1+j_1>t \\ i_2+j_2>t \\ (i_1,j_1)\neq(i_2,j_2)}} \widehat{\mu}_{ij_1} \widehat{\mu}_{ij_2} x_{i_1,j_1}^\top \widehat{\mathbb{V}}(\tilde{\beta}) x_{i_2,j_2}.$$

□

A.4.2 Proposition 2.5

Préliminaire. D'après Wüthrich et Merz (2015),

$$\frac{\widehat{f}_j^{(t+1)}}{\widehat{f}_j^{(t)}} = \alpha_j^{(t)} \frac{C_{t-j,j+1}}{C_{t-j,j}} \frac{\widehat{f}_j^{(t)}}{\widehat{f}_j^{(t)}} + \left(1 - \alpha_j^{(t)}\right) \quad \text{avec} \quad \alpha_j^{(t)} = \frac{C_{t-j,j}}{\sum_{i=1}^{t-j} C_{ij}}.$$

D'après Renshaw et Verrall (1998),

$$\widehat{f}_j^{(t)} = 1 + \frac{e^{\widehat{b}_{j+1}}}{\sum_{k=0}^{j} e^{\widehat{b}_k}} = \frac{\sum_{k=0}^{j+1} e^{\widehat{b}_k}}{\sum_{k=0}^{j} e^{\widehat{b}_k}} \quad j = 0, \ldots, J-1,$$

d'où

$$\prod_{j=t-i}^{J-1} \widehat{f}_j^{(t)} = \frac{\sum_{k=0}^{J} e^{\widehat{b}_k}}{\sum_{k=0}^{t-i} e^{\widehat{b}_k}},$$

et

$$\widehat{C}_{t-j,J}^{(t)} = C_{t-j,j} \frac{\sum_{k=0}^{J} e^{\widehat{b}_k}}{\sum_{k=0}^{j} e^{\widehat{b}_k}}.$$

Dans la même logique,

$$\begin{aligned}
C_{t-i,j+1} &= C_{t-j,j} + Y_{t-i,j+1}, \\
&= \widehat{C}_{t-j,J}^{(t)} \frac{\sum_{k=0}^{j} e^{\widehat{b}_k}}{\sum_{k=0}^{J} e^{\widehat{b}_k}} + Y_{t-j,j+1}, \\
&= \widehat{C}_{t-j,J}^{(t)} \frac{\sum_{k=0}^{j+1} e^{\widehat{b}_k}}{\sum_{k=0}^{J} e^{\widehat{b}_k}} + Y_{t-j,j+1} - \widehat{C}_{t-j,J}^{(t)} \frac{e^{\widehat{b}_{j+1}}}{\sum_{k=0}^{J} e^{\widehat{b}_k}}, \\
&= \widehat{C}_{t-j,J}^{(t)} \frac{\sum_{k=0}^{j+1} e^{\widehat{b}_k}}{\sum_{k=0}^{J} e^{\widehat{b}_k}} + Y_{t-j,j+1} - e^{\widehat{c}+\widehat{a}_{t-j}+\widehat{b}_{j+1}} \quad \text{car } \widehat{C}_{t-j,J}^{(t)} = \sum_{k=0}^{J} \widehat{Y}_{t-j,k} = e^{\widehat{c}+\widehat{a}_{t-j}} \sum_{k=0}^{J} e^{\widehat{b}_k}, \\
&= C_{t-j,J}^{(t)} \frac{\sum_{k=0}^{j+1} e^{\widehat{b}_k}}{\sum_{k=0}^{J} e^{\widehat{b}_k}} + \underbrace{Y_{t-j,j+1} - e^{c+a_{t-j}+b_{j+1}}}_{\xi_{t-j,j+1}} + \underbrace{e^{c+a_{t-j}+b_{j+1}} - e^{\widehat{c}+\widehat{a}_{t-j}+\widehat{b}_{j+1}}}_{\zeta_{t-j,j+1}}.
\end{aligned}$$

D'où

$$\begin{aligned}
\frac{C_{t-j,j+1}/C_{t-j,j}}{\widehat{f}_j^{(t)}} &= \underbrace{\frac{\widehat{C}_{t-j,J}^{(t)} \frac{\sum_{k=0}^{j+1} e^{\widehat{b}_k}}{\sum_{k=0}^{J} e^{\widehat{b}_k}}}{C_{t-j,j} \frac{\sum_{k=0}^{j+1} e^{\widehat{b}_k}}{\sum_{k=0}^{j} e^{\widehat{b}_k}}}}_{=1} + (\xi_{t-j,j+1} + \zeta_{t-j,j+1}) \frac{1}{C_{t-i,j}} \frac{\sum_{k=0}^{j} e^{\widehat{b}_k}}{\sum_{k=0}^{j+1} e^{\widehat{b}_k}}, \\
&= 1 + (\xi_{t-j,j+1} + \zeta_{t-j,j+1}) \frac{1}{\hat{C}_{t-j,j+1}^{(t)}}.
\end{aligned}$$

Démonstration. Pour rappel,

$$CDR_{i,t+1} = R_i^{(t)} - \left(Y_{i,t-i+1} + R_i^{(t+1)}\right) = C_{i,J}^{(t)} - C_{i,J}^{(t+1)},$$

$$= C_{i,t-i} \cdot \widehat{f}_{t-i}^{(t)} \cdot \widehat{f}_{t-i+1}^{(t)} \cdot \cdots \cdot \widehat{f}_{J-1}^{(t)} - C_{i,t-i+1} \cdot \widehat{f}_{t-i+1}^{(t+1)} \cdot \cdots \cdot \widehat{f}_{J-1}^{(t+1)},$$

$$= C_{i,t-i} \left(\prod_{j=t-i}^{J-1} \widehat{f}_j^{(t)} \right) \left(1 - \frac{C_{i,t-i+1}/C_{i,t-i}}{\widehat{f}_{t-i}^{(t)}} \prod_{j=t-i+1}^{J-1} \frac{\widehat{f}_j^{(t+1)}}{\widehat{f}_j^{(t)}} \right),$$

$$= \widehat{C}_{i,J}^{(t)} - \frac{\widehat{C}_{i,t-i+1}/C_{i,t-i}}{\prod_{j=t-i}^{J-1} \widehat{f}_j^{(t)}} \cdot \left(\alpha_j^{(t)} \frac{C_{t-j,j+1}/C_{t-j,j}}{\widehat{f}_j^{(t)}} + (1 - \alpha_j^{(t)}) \right), \text{ d'après Wüthrich et Merz (2015)}$$

$$= \underbrace{\widehat{C}_{i,J}^{(t)}}_{\text{ultime en } t} \underbrace{- \widehat{C}_{i,J}^{(t)} \frac{C_{i,t-i+1}/C_{i,t-i}}{\widehat{f}_{t-i}^{(t)}} \prod_{j=t-i+1}^{J-1} \left(\alpha_j^{(t)} \frac{C_{t-j,j+1}/C_{t-j,j}}{\widehat{f}_j^{(t)}} + \left(1 - \alpha_j^{(t)}\right) \right)}_{\text{ultime en } t+1}.$$

Enfin, on conclut avec Renshaw et Verrall (1998). □

A.4.3 Proposition 2.6

Démonstration. On introduit les poids suivants

$$q_{k+1}^{(t)} = \partial_{\log Y_{t-k,k+1}} \log \left(\sum_{j=t-J+1}^{J} \hat{C}_{i,J}^{(t+1)} \right) \Bigg|_0,$$

$$= \frac{\partial_{\log Y_{t-k,k+1}} \sum_{i=t-J+1}^{I} \widehat{C}_{i,J}^{(t+1)}}{\sum_{i=t-J+1}^{I} \widehat{C}_{i,J}^{(t+1)}} \Bigg|_0,$$

$$= \frac{\partial_{\log Y_{t-k,k+1}} \sum_{i=t-k}^{I} \widehat{C}_{i,J}^{(t+1)}}{\sum_{i=t-J+1}^{I} \widehat{C}_{i,J}^{(t+1)}} \Bigg|_0, \quad t - I \leq k \leq J - 1.$$

D'après la proposition de la section A.4.2,

$$\sum_{i=t-J+1}^{I} \widehat{C}_{i,J}^{(t+1)} = \sum_{i=t-J+1}^{I} \widehat{C}_{i,J}^{(t)} \left(1 + \frac{\zeta_{i,t-i+1} + \zeta_{i,t-i+1}}{\widehat{C}_{i,t-i+1}^{(t)}} \right) \prod_{j=t-i+1}^{J-1} \left(1 + \alpha_j^{(t)} \frac{\xi_{t-j,j+1} + \zeta_{t-j,j+1}}{\widehat{C}_{t-j,j+1}^{(t)}} \right),$$

d'où

$$\partial_{\log Y_{t-k,k+1}} \widehat{C}_{t-k,J}^{(t+1)} \Big|_0 = \partial_{\log Y_{t-k,k+1}} \widehat{C}_{t-k,J}^{(t)} \left(1 + \frac{\xi_{t-k,k+1} + \zeta_{t-k,k+1}}{\widehat{C}_{t-k,k+1}^{(t)}} \right) \prod_{j=k+1}^{J-1} \left(1 + \alpha_j^{(t)} \frac{\xi_{t-j,j+1} + \zeta_{t-j,j+1}}{\widehat{C}_{t-j,j+1}^{(t)}} \right) \Bigg|_0,$$

$$= \widehat{C}_{t-k,J}^{(t)} \left(\partial_{\log Y_{t-k,k+1}} \frac{\xi_{t-k,k+1}}{\widehat{C}_{t-k,k+1}^{(t)}} \right) \prod_{j=k+1}^{J-1} \left(1 + \alpha_j^{(t)} \frac{\xi_{t-j,j+1} + \zeta_{t-j,j+1}}{\widehat{C}_{t-j,j+1}^{(t)}} \right) \Bigg|_0,$$

$$= \widehat{C}_{t-k,J}^{(t)} \left(\partial_{\log Y_{t-k,k+1}} \frac{Y_{t-k,k+1}}{\widehat{C}_{t-k,k+1}^{(t)}} \right) \prod_{j=k+1}^{J-1} \left(1 + \alpha_j^{(t)} \frac{\xi_{t-j,j+1} + \zeta_{t-j,j+1}}{\widehat{C}_{t-j,j+1}^{(t)}} \right) \Bigg|_0,$$

$$= \widehat{C}^{(t)}_{t-k,J}\left(\frac{Y_{t-k,k+1}}{\widehat{C}^{(t)}_{t-k,k+1}}\prod_{j=k+1}^{J-1}\left(1+\alpha_j^{(t)}\frac{\xi_{t-j,j+1}+\zeta_{t-j,j+1}}{\widehat{C}^{(t)}_{t-j,j+1}}\right)\right)\Bigg|_0,$$

$$= \widehat{C}^{(t)}_{t-k,J}\frac{e^{c+a_{t-k}+b_{k+1}}}{\widehat{C}^{(t)}_{t-k,k+1}}\prod_{j=k+1}^{J-1}\left(1+\alpha_j^{(t)}\frac{\xi_{t-j,j+1}+\zeta_{t-j,j+1}}{\widehat{C}^{(t)}_{t-j,j+1}}\right),$$

pour $t-I\leq k\leq J-2$, et de même,

$$\partial_{\log Y_{t-J+1,J}}\widehat{C}^{(t+1)}_{t-J+1,J}\Big|_0 = \partial_{\log Y_{t-J+1,J}}\widehat{C}^{(t)}_{t-J+1,J}\left(1+\frac{\xi_{t-J+1,J}+\zeta_{t-J+1,J}}{\widehat{C}^{(t)}_{t-J+1,J}}\right)\Bigg|_0,$$

$$= \widehat{C}^{(t)}_{t-J+1,J}\left(\partial_{\log Y_{t-J+1,J}}\frac{\xi_{t-J+1,J}}{\widehat{C}^{(t)}_{t-J+1,J}}\right)\Bigg|_0,$$

$$= \partial_{\log Y_{t-J+1,J}}Y_{t-J+1,J}\Big|_0,$$

$$= e^{c+a_{t-J+1}+b_J}, \quad \text{pour } k=J-1.$$

D'où , en raisonnant de la même manière,

$$q^{(t)}_{k+1} = \frac{e^{c+a_{t-k}+b_{k+1}}}{\widehat{C}^{(t)}_{t-k,k+1}}\left(\widehat{C}^{(t)}_{t-k,J}+\alpha_k^{(t)}\sum_{i=t-k+1}^{I}\widehat{C}^{(t)}_{i,J}\right)\Bigg/\sum_{i=t-J+1}^{I}\widehat{C}^{(t)}_{i,J}, \quad t-I\leq k\leq J-1.$$

Comme $\partial_{\log x}\log\left(f(x)\right)=\frac{x}{f(x)}\frac{\partial}{\partial x}f(x)$, il est possible d'écrire

$$\partial_{Y_{t-k,k+1}}\sum_{i=t-J+1}^{I}\widehat{C}^{(t+1)}_{i,J}\Big|_0 = \frac{q^{(t)}_{k+1}}{e^{c+a_{t-k}+b_{k+1}}}\sum_{i=t-J+1}^{I}\widehat{C}^{(t)}_{i,J}.$$

De plus, par définition de ζ,

$$\partial_{\widehat{c}_{k+1}}\zeta_{t-k,k+1} = \partial_{\widehat{c}_{k+1}}\left(e^{c+a_{t-k}+b_{k+1}}-e^{c+\widehat{a}_{t-k}+\widehat{b}_{k+1}}\right) = -e^{c+\widehat{a}_{t-k}+\widehat{b}_{k+1}}.$$

Dans la même logique,

$$\partial_{\widehat{c}_{k+1}}\sum_{i=t-k}^{I}\widehat{C}^{(t+1)}_{i,J}\Big|_0 = -\frac{e^{c+a_{t-k}+b_{k+1}}}{\widehat{C}^{(t)}_{t-k,k+1}}\left(\widehat{C}^{(t)}_{t-k,J}+\alpha_j^{(t)}\sum_{i=t-k+1}^{I}\widehat{C}^{(t)}_{i,J}\right) = -q^{(t)}_{k+1}\sum_{i=t-J+1}^{I}\widehat{C}^{(t)}_{i,J},$$

et

$$\partial_{\widehat{a}_{t-k}}\sum_{i=t-k}^{I}\widehat{C}^{(t+1)}_{i,J}\Big|_0 = -q^{(t)}_{k+1}\sum_{i=t-J+1}^{I}\widehat{C}^{(t)}_{i,J},$$

$$\partial_{\widehat{b}_{k+1}}\sum_{i=t-k}^{I}\widehat{C}^{(t+1)}_{i,J}\Big|_0 = -q^{(t)}_{k+1}\sum_{i=t-J+1}^{I}\widehat{C}^{(t)}_{i,J}.$$

Par approximation de Taylor, on obtient

$$\frac{\sum_{i=t-J+1}^{I}\widehat{C}^{(t+1)}_{i,J}}{\sum_{i=t-J+1}^{I}\widehat{C}^{(t)}_{i,J}} = \frac{\sum_{i=t-J+1}^{I}\widehat{C}^{(t)}_{i,J}\left(1+\frac{\xi_{i,t-i+1}+\zeta_{i,t-i+1}}{\widehat{C}^{(t)}_{i,t-i+1}}\right)\prod_{j=t-i+1}^{J-1}\left(1+\alpha_j^{(t)}\frac{\xi_{t-j,j+1}+\zeta_{t-j,j+1}}{\widehat{C}^{(t)}_{t-j,j+1}}\right)}{\sum_{i=t-J+1}^{I}\widehat{C}^{(t)}_{i,J}},$$

$$\approx 1 + \sum_{k=t-I}^{J-1} \frac{q_{k+1}^{(t)}}{e^{c+a_{t-k}+b_{k+1}}} \xi_{t-j,j+1} - \sum_{k=t-I}^{J-1} q_{k+1}^{(t)} \left(c - \widehat{c} + a_{t-k} - \widehat{a}_{t-k} + b_{k+1} - \widehat{b}_{k+1} \right),$$

$$= 1 + \sum_{k=t-I}^{J-1} \frac{q_{k+1}^{(t)}}{\mu_{t-k,k+1}} \xi_{t-j,j+1} - \sum_{k=t-I}^{J-1} q_{k+1}^{(t)} \left(\widehat{\eta}_{t-k,k+1} - \widehat{\eta}_{t-k,k+1} \right),$$

$$= 1 + \sum_{k=t-I}^{J-1} \frac{q_{k+1}^{(t)}}{\mu_{t-k,k+1}} \xi_{t-j,j+1} - \sum_{k=t-I}^{J-1} q_{k+1}^{(t)} x_{t-k,k+1}^{\top} \left(\beta - \widehat{\beta} \right).$$

Avec le résultat de la proposition en section A.4.2, on obtient

$$\left(\sum_{i=t-J+1}^{I} \widehat{CDR}_{i,t+1} \right)^2 \approx \left(\sum_{i=t-J+1}^{I} \widehat{C}_{i,J}^{(t)} \right)^2 \left[\sum_{k=t-I}^{J-1} \frac{\left(q_{k+1}^{(t)} \right)^2}{\mu_{t-k,k+1}^2} \xi_{t-j,j+1}^2 \right.$$

$$\left. + \sum_{k_1=t-I}^{J-1} \sum_{k_2=t-I}^{J-1} q_{k_1+1}^{(t)} q_{k_2+1}^{(t)} x_{t-k,k+1}^{\top} \left(\beta - \widehat{\beta} \right) \left(\beta - \widehat{\beta} \right)^{\top} x_{t-k,k+1} \right],$$

$$= \left(\sum_{i=t-J+1}^{I} \widehat{C}_{i,J}^{(t)} \right)^2 \left[\sum_{k=t-I}^{J-1} \left(q_{k+1}^{(t)} \right)^2 \mu_{t-k,k+1}^2 \mathbb{V}(Y_{t-k,k+1}) + \sum_{k_1=t-I}^{J-1} \sum_{k_2=t-I}^{J-1} q_{k_1+1}^{(t)} q_{k_2+1}^{(t)} x_{t-k,k+1}^{\top} \mathbb{V}(\beta) x_{t-k,k+1} \right].$$

Ainsi, on obtient finalement

$$\widehat{\text{MSEP}} \left(\sum_{i=t-J+1}^{I} \widehat{CDR}_{i,t+1} \right) = \left(\sum_{i=t-J+1}^{I} \sum_{j=0}^{J} \widehat{\mu}_{ij} \right)^2 \times \left[\underbrace{\widehat{\phi} \sum_{k=t-I}^{J-1} \widehat{q}_{k+1}^2 \widehat{\mu}_{t-k,k+1}}_{\text{erreur de processus}} + \underbrace{\widehat{q}^{\top} X_{t+1}^{\top} \widehat{\mathbb{V}}(\widehat{\beta}) X_{t+1}^{\top} \widehat{q}}_{\text{erreur d'estimation}} \right].$$

□

A.5 Preuves de la méthode ECLRM

A.5.1 Proposition 2.7

Préliminaire 1.

Démonstration. D'après le modèle,

$$\left(\widehat{Y}_{i,j}^{Pa} \right)^{I+1} = R_{i,I-i+1} \prod_{k=I-i+1}^{j-2} \widehat{h}_k^{I+1} \widehat{f}_{j-1}^{I+1},$$

$$\widehat{C}_{i,J}^{I+1} = C_{i,I-i+1}^{In} + \sum_{j=I-i+2}^{J} \left(\widehat{Y}_{i,j}^{In} \right)^{I+1} = C_{i,I-i+1}^{Pa} + \sum_{j=I-i+2}^{J} \left(\widehat{Y}_{i,j}^{Pa} \right)^{I+1}.$$

On a

$$\mathbb{E}\left[\widehat{C}_{i,J}^{I+1} \mid D_I\right] = \mathbb{E}\left[C_{i,I-i+1}^{Pa} + R_{i,I-i+1} \sum_{j=I-i+2}^{J} \prod_{k=I-i+1}^{j-2} \widehat{h}_k^{I+1} \widehat{f}_{j-1}^{I+1} \mid D_I\right]$$

$$= C_{i,I-i}^{Pa} + R_{i,I-i} f_{I-i} + \sum_{j=I-i+2}^{J} \mathbb{E}\left[R_{i,I-i+1} \prod_{k=I-i+1}^{j-2} \widehat{h}_k^{I+1} \widehat{f}_{j-1}^{I+1} \mid D_I\right],$$

car $\mathbb{E}\left[C_{i,j+1} \mid B_j\right] = \left(C_{i,j}^{Pa} + R_{i,j} f_j, C_{i,j}^{In} + R_{i,j} g_j\right)'$.

Par définition, $\delta_j = \frac{R_{I-j,j}}{\sum_{i=0}^{I-j} R_{i,j}}$ d'où

$$\widehat{f}_j^{I+1} = \frac{\sum_{i=0}^{I-j} Y_{i,j+1}^{Pa}}{\sum_{i=0}^{I-j} R_{i,j}} = (1-\delta_j)\widehat{f}_j^I + \frac{Y_{I-j,j+1}^{Pa}}{\sum_{i=0}^{I-j} R_{i,j}}$$

$$\widehat{g}_j^{I+1} = \frac{\sum_{i=0}^{I-j} Y_{i,j+1}^{In}}{\sum_{i=0}^{I-j} R_{i,j}} = (1-\delta_j)\widehat{g}_j^I + \frac{Y_{I-j,j+1}^{In}}{\sum_{i=0}^{I-j} R_{i,j}}.$$

Les termes du produit dans $\mathbb{E}\left[\widehat{C}_{i,J}^{I+1} \mid D_I\right]$ sont indépendants car les années d'accident sont indépendantes, d'où le fait que

$$\mathbb{E}\left[R_{i,I-i+1} \prod_{k=I-i+1}^{j-2} \widehat{h}_k^{I+1} \widehat{f}_{j-1}^{I+1} \mid D_I\right] = \mathbb{E}\left[R_{i,I-i+1} \mid D_I\right] \prod_{k=I-i+1}^{j-2} \mathbb{E}\left[\widehat{h}_k^{I+1} \mid D_I\right] \mathbb{E}\left[\widehat{f}_{j-1}^{I+1} \mid D_I\right]$$

$$= R_{i,I-i} h_{I-i} \prod_{k=I-i+1}^{j-2} \mathbb{E}\left[\widehat{h}_k^{I+1} \mid D_I\right] \mathbb{E}\left[\widehat{f}_{j-1}^{I+1} \mid D_I\right].$$

Enfin, tout comme pour les coefficients g et h,

$$\mathbb{E}\left[\widehat{f}_j^{I+1} \mid D_I\right] = (1-\delta_j)\widehat{f}_j^I + \frac{1}{\sum_{i=0}^{I-j} R_{i,j}} \mathbb{E}\left[Y_{I-j,j+1}^{Pa} \mid D_I\right] = (1-\delta_j)\widehat{f}_j^I + \delta_j f_j.$$

Ainsi,

$$\mathbb{E}\left[\widehat{C}_{i,J}^{I+1} \mid D_I\right] = C_{i,I-i}^{Pa} + R_{i,I-i} f_{I-i} + R_{i,I-i} \sum_{j=I-i+2}^{J} h_{I-i} \prod_{k=I-i+1}^{j-2} \left((1-\delta_k)\widehat{h}_k^I + \delta_k h_k\right)$$

$$\times \left((1-\delta_{j-1})\widehat{f}_{j-1}^I + \delta_{j-1} f_{j-1}\right).$$

□

Préliminaire 2. D'après l'annexe 7 de Dahms et al. (2009), on admet que

$$\mathbb{V}\left(\widehat{CDR}_i(I+1) \mid D_I\right) = \mathbb{V}\left(\widehat{C}_{i,J}^{I+1} \mid D_I\right)$$

$$= \left(\widehat{Y}_{i,I-i+1}^{Pa}\right)^{I^2} \frac{\widehat{\alpha}_{I-i+1,I-i+1,I-i}}{R_{i,I-i}} + \sum_{j=I-i+2}^{I} \sum_{m=I-i+2}^{I} \left(\widehat{Y}_{i,j}^{Pa}\right)^I \left(\widehat{Y}_{i,m}^{Pa}\right)^I \left\{\frac{\widehat{\alpha}_{j,m,I-i}}{R_{i,I-i}} + \sum_{k=I-i+1}^{\min(j,m)-1} \delta_k^2 \frac{\widehat{\alpha}_{j,m,k}}{R_{I-k,k}}\right\}$$

$$+2\left(\widehat{Y}_{i,I-i+1}^{Pa}\right)^I \sum_{j=I-i+2}^{J} \frac{\widehat{\alpha}_{I-i+1,j,I-i}}{R_{i,I-i}} \widehat{Y}_{i,j}^{Pa}.$$

Préliminaire 3. D'après l'annexe 9 de Dahms et al. (2009), on admet que

$$R^2_{i,I-i}\widehat{\Delta}_i = R^2_{i,I-i}\left\{\left(\widehat{f}^I_{I-i}\right)^2\frac{\widehat{\alpha}_{I-i+1,I-i+1,I-i}}{\sum_{l=0}^{i-1}R_{l,I-i}} + 2\sum_{j=I-i+2}^{J}\widehat{f}^I_{I-i}\prod_{k=I-i}^{j-2}\widehat{h}^I_k\frac{\widehat{\alpha}_{I-i+1,j,I-i}}{\sum_{l=0}^{i-1}R_{l,I-i}}\right.$$
$$\left.+\sum_{j=I-i+2}^{J}\sum_{m=I-i+2}^{J}\prod_{k=I-i}^{j-2}\widehat{h}^I_k\widehat{f}^I_{j-1}\prod_{k=I-i}^{m-2}\widehat{h}^I_k\widehat{f}^I_{m-1}\left[\frac{\widehat{\alpha}_{j,m,I-i}}{\sum_{l=0}^{i-1}R_{l,I-i}}+\sum_{k=I-i+1}^{\min(j,m)-1}\delta_k^2\frac{\widehat{\alpha}_{j,m,k}}{\sum_{l=0}^{k-1}R_{l,k}}\right]\right\}.$$

Préliminaire 4. D'après l'annexe 8 de Dahms et al. (2009), on admet que

$$\widehat{\mathrm{Cov}}\left(\widehat{CDR}_i(I+1),\widehat{CDR}_n(I+1)\mid D_I\right)$$
$$=\sum_{j,m=I-i+2}^{I}\left(\widehat{Y}^{Pa}_{i,j}\right)^I\left(\widehat{Y}^{Pa}_{n,m}\right)^I\left\{\frac{\widehat{\alpha}_{j,m,I-i}}{R_{i,I-i}}+\sum_{k=I-i+1}^{\min(j,m)-1}\delta_k^2\frac{\widehat{\alpha}_{j,m,k}}{R_{I-k,k}}\right\}$$
$$+\sum_{j=I-i+2}^{J}\left(\widehat{Y}^{Pa}_{i,j}\right)^I\delta_{I-i}\frac{\widehat{\alpha}_{I-i+1,j,I-i}}{R_{i,I-i}}\left(\widehat{Y}^{Pa}_{n,j}\right)^I$$
$$+\left(\widehat{Y}^{Pa}_{i,I-i+1}\right)^I\sum_{j=I-i+2}^{J}\delta_{I-i}\frac{\widehat{\alpha}_{I-i+1,j,I-i}}{R_{i,I-i}}\left(\widehat{Y}^{Pa}_{n,j}\right)^I$$
$$=\sum_{j,m=I-i+1}^{I}\left(\widehat{Y}^{Pa}_{i,j}\widehat{Y}^{Pa}_{n,m}\right)^I\left\{\frac{\widehat{\alpha}_{j,m,I-i}}{R_{i,I-i}}+\sum_{k=I-i+1}^{\min(j,m)-1}\delta_k^2\frac{\widehat{\alpha}_{j,m,k}}{R_{I-k,k}}\right\}.$$

Préliminaire 5. D'après l'annexe 10 de Dahms et al. (2009), on admet que

$$R_{i,I-i}R_{n,I-n}\widehat{\Delta}_{i,n} = \left(\widehat{Y}^{Pa}_{i,I-i+1}\right)^I\sum_{j=I-i+1}^{J}\left(\widehat{Y}^{Pa}_{n,j}\right)\delta_{I-i}\frac{\widehat{\alpha}_{I-i+1,j,I-i}}{\sum_{l=0}^{i-1}R_{l,I-i}}$$
$$+\left(\widehat{Y}^{Pa}_{n,I-n+1}\right)^I\sum_{j=I-n+2}^{J}\left(\widehat{Y}^{Pa}_{i,j}\right)\delta_{I-n}\frac{\widehat{\alpha}_{I-n+1,j,I-n}}{\sum_{l=0}^{n-1}R_{l,I-n}}$$
$$+\sum_{j,m=I-i+1}^{J}\left(\widehat{Y}^{Pa}_{i,j}\right)^I\left(\widehat{Y}^{Pa}_{n,m}\right)^I\left[\delta_{I-i}\frac{\widehat{\alpha}_{j,m,I-i}}{\sum_{l=0}^{i-1}R_{l,I-i}}+\sum_{k=I-i+1}^{\min(j,m)-1}\delta_k^2\frac{\widehat{\alpha}_{j,m,k}}{\sum_{l=0}^{k-1}R_{l,k}}\right]$$
$$=\sum_{j,m=I-i+1}^{J}\left(\widehat{Y}^{Pa}_{i,j}\widehat{Y}^{Pa}_{n,m}\right)^I\left[\delta_{I-i}\frac{\widehat{\alpha}_{j,m,I-i}}{\sum_{l=0}^{i-1}R_{l,I-i}}+\sum_{k=I-i+1}^{\min(j,m)-1}\delta_k^2\frac{\widehat{\alpha}_{j,m,k}}{\sum_{l=0}^{k-1}R_{l,k}}\right].$$

avec $\Delta_{i,n}\overset{\text{def.}}{=}\Delta_i^{1/2}\Delta_n^{1/2}$.

Démonstration. On a

$$\mathrm{MSEP}_{\widehat{CDR}_i(I+1)|D_I}(0)=\mathbb{V}\left(\widehat{CDR}_i(I+1)\mid D_I\right)+\left[\mathbb{E}\left(\widehat{CDR}_i(I+1)\mid D_I\right)\right]^2.$$

D'après la décomposition erreur de processus et erreur d'estimation, l'erreur d'estimation s'écrit

$$\mathbb{E}\left[\widehat{CDR}_i(I+1) \mid D_I\right]^2 = \left(\widehat{C}_{i,J} - \mathbb{E}\left[\widehat{C}_{i,J}^{I+1} \mid D_I\right]\right)^2 \overset{\text{def.}}{=} R_{i,I-i}^2 \Delta_i.$$

De plus, on a d'une part

$$\widehat{C}_{i,J} = C_{i,I-i}^{Pa} + R_{i,I-i} \sum_{j=I-i+1}^{J} \prod_{k=I-i}^{j-2} \widehat{h}_k^I \widehat{f}_{j-1}^I.$$

D'autre part , pour le CDR observable, le préliminaire 1 implique que

$$\mathbb{E}\left[\widehat{CDR}_i(I+1) \mid D_I\right] = R_{i,I-i}\left(\widehat{f}_{I-i}^I - f_{I-i}\right) + R_{i,I-i} \sum_{j=I-i+2}^{J} \left[\prod_{k=I-i}^{j-2} \widehat{h}_k^I \widehat{f}_{j-1}^I - h_{I-i} \prod_{k=I-i+1}^{j-2} \left((1-\delta_k)\widehat{h}_k^I + \delta_k h_k\right)\right]$$

$$\times \left((1-\delta_{j-1})\widehat{f}_{j-1}^I + \delta_{j-1} f_{j-1}\right).$$

D'après le préliminaire 2, pour l'erreur de processus,

$$\mathbb{V}\left(\widehat{CDR}_i(I+1) \mid D_I\right) = \sum_{j,m=I-i+1}^{I} \left(\widehat{Y}_{i,j}^{Pa}\right)^I \left(\widehat{Y}_{i,m}^{Pa}\right)^I \left\{\frac{\widehat{\alpha}_{j,m,I-i}}{R_{i,I-i}} + \sum_{k=I-i+1}^{\min(j,m)-1} \delta_k^2 \frac{\widehat{\alpha}_{j,m,k}}{R_{I-k,k}}\right\}.$$

Pour l'erreur d'estimation, le préliminaire 3 implique que

$$R_{i,I-i}^2 \widehat{\Delta}_i = \sum_{j,m=I-i+1}^{J} \left(\widehat{Y}_{i,j}^{Pa}\right)^I \left(\widehat{Y}_{i,m}^{Pa}\right)^I \left[\frac{\widehat{\alpha}_{j,m,I-i}}{\sum_{n=0}^{i-1} R_{n,I-i}} + \sum_{k=I-i+1}^{\min(j,m)-1} \delta_k^2 \frac{\widehat{\alpha}_{j,m,k}}{\sum_{n=0}^{k-1} R_{n,k}}\right].$$

Ainsi, en sommant l'erreur de processus et d'estimation, on obtient finalement

$$\text{MSEP}_{\widehat{CDR}_i(I+1)|D_I}(0) = \mathbb{V}\left(\widehat{CDR}_i(I+1) \mid D_I\right) + R_{i,I-i}^2 \widehat{\Delta}_i,$$

$$= \sum_{j,m=I-i+1}^{J} \left(\widehat{Y}_{i,j}^{Pa}\right)^I \left(\widehat{Y}_{i,m}^{Pa}\right)^I \left[\frac{\delta_{I-i}^{-1} \widehat{\alpha}_{j,m,I-i}}{\sum_{n=0}^{i-1} R_{n,I-i}} + \sum_{k=I-i+1}^{\min(j,m)-1} \delta_k \frac{\widehat{\alpha}_{j,m,k}}{\sum_{n=0}^{k-1} R_{n,k}}\right].$$

Démonstration. De la même manière que *supra*,

$$\text{MSEP}_{\sum_{i=1}^I \widehat{CDR}_i(I+1)|D_I}(0) = \mathbb{V}\left(\sum_{i=1}^{I} \widehat{CDR}_i(I+1) \mid D_I\right) + \left[\mathbb{E}\left(\sum_{i=1}^{I} \widehat{CDR}_i(I+1) \mid D_I\right)\right]^2$$

$$= \sum_{i,\ell=1}^{I} \text{Cov}\left(\widehat{CDR}_i(I+1), \widehat{CDR}_\ell(I+1) \mid D_I\right) + \left[\mathbb{E}\left(\sum_{i=1}^{I} \widehat{CDR}_i(I+1) \mid D_I\right)\right]^2$$

$$= \sum_{i,\ell=1}^{I} \text{Cov}\left(\widehat{CDR}_i(I+1), \widehat{CDR}_\ell(I+1) \mid D_I\right) + \Delta,$$

car

$$\mathbb{E}\left[\sum_{i=1}^{I} \widehat{CDR}_i(I+1) \mid D_I\right]^2 = \left(\sum_{i=1}^{I} \widehat{C}_{i,J}^{(t)} - \sum_{i=1}^{I} \mathbb{E}\left[\widehat{C}_{i,J}^{(t+1)} \mid D_I\right]\right)^2 \overset{\text{def.}}{=} \Delta.$$

Le préliminaire 4 implique que

$$\widehat{\mathbb{V}}\left(\sum_{i=1}^{I}\widehat{CDR}_i(I+1)\mid D_I\right)=\sum_{i=1}^{I}\widehat{\mathbb{V}}\left(\widehat{CDR}_i(I+1)\mid D_I\right)+2\sum_{1\leq i<n\leq I}\sum_{j,m=I-i+1}^{I}\left(\widehat{Y}_{i,j}^{Pa}\right)^I\left(\widehat{Y}_{n,m}^{Pa}\right)^I$$

$$\left\{\frac{\delta_{I-i}\widehat{\alpha}_{j,m,I-i}}{R_{i,I-i}}+\sum_{k=I-i+1}^{\min(j,m)-1}\delta_k^2\frac{\widehat{\alpha}_{j,m,k}}{R_{I-k,k}}\right\}.$$

Enfin, le préliminaire 5 implique que

$$\widehat{\Delta}=\sum_{i=1}^{I}R_{i,I-i}^2\widehat{\Delta}_i+2\sum_{1\leq i<n\leq I}R_{i,I-i}R_{n,I-n}\widehat{\Delta}_{i,n}$$

$$=\sum_{i=1}^{I}R_{i,I-i}^2\widehat{\Delta}_i+2\sum_{1\leq i<n\leq I}\sum_{j,m=I-i+1}^{J}\left(\widehat{Y}_{i,j}^{Pa}\right)^\top\left(\widehat{Y}_{n,m}^{Pa}\right)^\top\left[\delta_{I-i}\frac{\widehat{\alpha}_{j,m,I-i}}{\sum_{l=0}^{i-1}R_{l,I-i}}+\sum_{k=I-i+1}^{\min(j,m)-1}\delta_k^2\frac{\widehat{\alpha}_{j,m,k}}{\sum_{l=0}^{I-k-1}R_{l,k}}\right],$$

d'où le résultat final.

□

A.6 Code associé à la méthode ECLRM

```
compute_msep <- function(I){
  J <- I
  XPa <- cum2incr(CPa)
  XIn <- cum2incr(CIn)
  R <- CIn - CPa

  f_I <- numeric(J)
  for (j in 0:(J - 1)) {
    num <- sum(XPa[1:(I - j), j+2])
    denom <- sum(R[1:(I - j), j+1])
    f_I[j + 1] <- num / denom
  }

  f_I_1 <- numeric(J)
  for (j in 0:(J - 1)) {
    num <- sum(XPa[1:(I - j + 1), j+2])
    denom <- sum(R[1:(I - j + 1), j+1])
    f_I_1[j + 1] <- num / denom
  }

  g_I <- numeric(J)
  for (j in 0:(J - 1)) {
    num <- sum(XIn[1:(I - j), j+2])
    denom <- sum(R[1:(I - j), j+1])
    g_I[j + 1] <- num / denom
  }

  g_I_1 <- numeric(J)
  for (j in 0:(J - 1)) {
    num <- sum(XIn[1:(I - j + 1), j+2])
    denom <- sum(R[1:(I - j + 1), j+1])
    g_I_1[j + 1] <- num / denom
  }

  h_I = rep(1, I) + g_I - f_I
  h_I_1 = rep(1, I) + g_I_1 - f_I_1
```

```
X_pa_hat_I <- matrix(NA, nrow = I+1, ncol = J+1)
for (i in 1:I) {
  for (j in ((I-i+1):J)) {
    if ((I - i) <= (j - 2)) {
      product_h <- prod(h_I[(I - i + 1):(j - 1)])
    }
    else {
      product_h <- 1
    }
    X_pa_hat_I[i+1, j+1] <- R[i+1, I - i + 1] * product_h * f_I[j]
  }
}

s_1_1 <- numeric(J)
for (j in 0:(J-2)) {
  sum_term <- 0
  for (i in 0:(I-j-1)) {
    sum_term <- sum_term + R[i+1, j+1] * ( (XPa[i+1, j+2] / R[i+1, j+1]) - f_I[j+1] )^2
  }
  s_1_1[j+1]  <- sum_term / (I - j - 1)
}
s_1_1[J] <- min(c(s_1_1[J-1]^2 / s_1_1[J-2], s_1_1[J-2], s_1_1[J-1]))

s_2_2 <- numeric(J)
for (j in 0:(J-2)) {
  sum_term <- 0
  for (i in 0:(I-j-1)) {
    sum_term <- sum_term + R[i+1, j+1] * ( (XIn[i+1, j+2] / R[i+1, j+1]) - g_I[j+1] )^2
  }
  s_2_2[j+1]  <- sum_term / (I - j - 1)
}
s_2_2[J] <- min(c(s_2_2[J-1]^2 / s_2_2[J-2], s_2_2[J-2], s_2_2[J-1]))

s_1_2 <- numeric(J)
for (j in 0:(J-1)) {
  sum_term <- 0
  for (i in 0:(I-j-1)) {
    sum_term<-sum_term+R[i+1, j+1]*((XPa[i+1, j+2]/R[i+1, j+1]) - f_I[j+1] )*((XIn[i+1, j+2] / R[i+1, j+1]) -g_I[j+1])
  }
  s_1_2[j+1]  <- sum_term / (I - j - 1)
}

delta <- numeric(J)
for (j in 0:(J-1)){
  delta[j+1] <- R[I-j+1, j+1] / sum(R[1:(I-j+1), j+1])
}

alpha <- function(j, m, k){
  if (m == j && j == k + 1) {
    return(s_1_1[k+1] / (f_I[k+1]^2))
  }
  else if ((m > j && j == k + 1) || (j > m && m == k + 1)) {
    return((s_1_2[k+1] - s_1_1[k+1]) / (f_I[k+1] * h_I[k+1]))
  }
  else if ((m >= j && j > k + 1) || (j >= m && m > k + 1)) {
    return((s_1_1[k+1] - 2 * s_1_2[k+1] + s_2_2[k+1]) / (h_I[k+1]^2))
  }
  else{
    return(NA)
  }
}

MSEP_0 = numeric(I)
for (i in 1:I){
  sum_total <- 0
  for (j in (I-i+1):J){
    for (m in (I-i+1):J){
      term_1 <- delta[I-i+1]^-1 * alpha(j, m, I-i) / sum(R[1:i, I-i+1])
      term_2 <- 0
      if (I-i+1 <= (min(j,m)-1)) {
```

```
          for (k in (I-i+1):(min(j,m)-1)){
            term_2 <- term_2 + delta[k+1] * alpha(j, m, k) / sum(R[1:(I-k), k+1])
          }
        }
        sum_total <- sum_total + X_pa_hat_I[i+1, j+1] * X_pa_hat_I[i+1, m+1] * (term_1 + term_2)
      }
    }
    MSEP_0[i] <- sum_total
  }

  cross_msep <- 0
  for (i in 1:(I-1)){
    for (n in (i+1):I){
      for (j in (I-i+1):J){
        for (m in (I-i+1):J){
          term_1 <- alpha(j, m, I-i) / sum(R[1:i, I-i+1])
          term_2 <- 0
          if (I-i+1 <= (min(j,m)-1)) {
            for (k in (I-i+1):(min(j,m)-1)){
              term_2 <- term_2 + delta[k+1] * alpha(j, m, k) / sum(R[1:(I-k), k+1])
            }
          }
          cross_msep <- cross_msep + 2 * X_pa_hat_I[i+1, j+1] * X_pa_hat_I[n+1, m+1] * (term_1 + term_2)
        }
      }
    }
  }

  msep <- sum(MSEP_0) + cross_msep
  return(msep)
}
```

A.7 Algorithmes BFGS, L-BFGS et L-BFGS-B

L'algorithme L-BFGS-B (Wikipédia, 2023) découle directement de l'algorithme L-BFGS (Wikipédia, 2023) qui dépend lui-même de l'algorithme BFGS (Gontier, 2021). Ces trois algorithmes appartiennent à la famille Quasi-Newton. Soit $F\colon \mathbb{R}^d \to \mathbb{R}$ de classe C^2, et soit x^* un minimum de F. Pour tout x_0, l'approximation quadratique de F est

$$F(x) \approx \tilde{F}_{x_0}\colon x \mapsto F(x_0) + \langle \nabla F(x_0), x - x_0\rangle + \frac{1}{2}(x - x_0)H_F(x_0)(x - x_0),$$

avec la matrice hessienne $H_F(x_0)$ définie positive (donc inversible).

La formule de récurrence minimisant la fonction $\tilde{F}$ est

$$x_{n+1} = x_n - [H_F(x_n)]^{-1}\nabla F(x_n).$$

L'algorithme BFGS repose sur l'approximation de la hessienne $H_F(x_{n+1})$ par une matrice B_{n+1} telle que B_{n+1}^{-1} se calcule aisément. On a

$$B_{n+1} = B_n - \frac{|B_n s_n\rangle \langle B_n s_n|}{\langle s_n, B_n s_n\rangle} + \frac{|y_n\rangle \langle y_n|}{\langle y_n, s_n\rangle},$$

avec $B_0 = Id$ et $s_n = x_{n+1} - x_n$, $y_n = \nabla F(x_{n+1}) - \nabla F(x_n)$, et $|\rangle\langle|$ le produit extérieur et $\langle,\rangle$ le produit scalaire.

En notant $W_n = B_n^{-1}$,

$$W_{n+1} = \left(I_d - \frac{|s_n\rangle \langle y_n|}{\langle s_n, y_n\rangle}\right) W_n \left(Id - \frac{|y_n\rangle \langle s_n|}{\langle s_n, y_n\rangle}\right) + \frac{|s_n\rangle \langle s_n|}{\langle s_n, y_n\rangle}.$$

Algorithme 5 Algorithme BFGS

Initialisation : Par convention, $x_{-1} = 0 \in \mathbb{R}^d$, $\nabla F(x_{-1}) = 0$
$x_0 \in \mathbb{R}^d$, $W_0 = I_n$
Calcul de x_{n+1} :
for $n \geq 0$ **do**
    Calculer $s_n = x_{n+1} - x_n$ et $y_n = \nabla F(x_{n+1}) - \nabla F(x_n)$
    Mettre à jour $W_{n+1} = \left(I_d - \frac{|s_n\rangle\langle y_n|}{\langle s_n, y_n\rangle}\right) W_n \left(Id - \frac{|y_n\rangle\langle s_n|}{\langle s_n, y_n\rangle}\right) + \frac{|s_n\rangle\langle s_n|}{\langle s_n, y_n\rangle}$
    Calculer la direction de descente $h_{n+1} = -W_{n+1}\nabla F(x_{n+1})$
    Choisir τ_{Wolfe}, le pas de temps optimal selon la règle de Wolfe, i.e tel que $\langle \nabla F(x_n + \tau h_n), h_n\rangle \geq c\langle \nabla F(x_n), h_n\rangle, 0 < c < 1$
    $x_{n+2} = x_{n+1} + \tau_{\text{Wolfe}} h_{n+1}$
end for

Ces notations permettent d'introduire l'algorithme 5.

Dans l'algorithme L-BFGS, $\rho_k = \frac{1}{y_k^\top s_k}$ est défini et la hessienne inverse vaut désormais

$$W_{k+1} = \left(I - \rho_k s_k y_k^\top\right) W_k \left(I - \rho_k y_k s_k^\top\right) + \rho_k s_k s_k^\top.$$

De plus, soit $q_i = \left(I - \rho_i y_i s_i^\top\right) q_{i+1}$, ou dit autrement, si $\alpha_i = \rho_i s_i^\top q_{i+1}$ alors $q_i = q_{i+1} - \alpha_i y_i$.

La direction de descente h s'écrit $h_{k-m} = W_k^0 q_{k-m}$, ou si $\beta_i = \rho_i y_i^\top z_i$, $h_{i+1} = h_i + (\alpha_i - \beta_i) s_i$.

L'algorithme 6 de descente de gradient, dans le cadre de l'algorithme L-BFGS, s'écrit de la manière suivante.

Algorithme 6 Calcul de la direction de descente L-BFGS

Initialisation des paramètres
$q = \nabla F(x_n)$
for $i = k, k-1, \ldots, k-m$ **do**
    $\alpha_i = \rho_i s_i^\top q$
    $q = q - \alpha_i y_i$
end for
$\gamma_k = \frac{s_{k-1}^\top y_{k-1}}{y_{k-1}^\top y_{k-1}}$
$W_k^0 = \gamma_k I$
$h = W_k^0 q$
for $i = k-m, k-m+1, \ldots, k-1$ **do**
    $\beta_i = \rho_i y_i^\top h$
    $h = h + s_i(\alpha_i - \beta_i)$
end for
$h = -h$
Retourner h comme direction de descente

Enfin, pour des problèmes sous contraintes bornées, l'algorithme L-BFGS-B est utilisé. Il revient à appliquer à chaque itération la méthode L-BFGS sur les variables libres de la fonction à minimiser.

A.8 Réserves estimées du chapitre 2

Année calendaire	Chain-Ladder	Bootstrap	GLM	ECLRM
2004	11704	11701	11704	6608
2005	13428	13515	13428	5702
2006	16300	16344	16300	6718
2007	21042	20972	21042	13293
2008	29534	26900	21584	11022
2009	25377	21969	22850	12529
2010	26104	28146	22471	17269
2011	32056	31719	28827	19424
2012	36732	34643	35063	22160
2013	30597	30511	29953	21269
2014	23234	23007	24881	18780
2015	26017	25567	27348	19811
2016	29366	28917	30486	11872
2017	30840	30434	31869	19053
2018	36658	35916	37530	27120
2019	40050	39345	40820	37791

TABLE A.1 : Réserves estimées selon les différents modèles du chapitre 2 (en milliers €)

A.9 Modèle interne partiel avec approche bayésienne

Un modèle interne partiel pour le risque de réserve, entièrement développé par CARNEVALE et CLEMENTE (2020) propose une extension du modèle *Correlated Chain-Ladder* (CCL) de MEYERS (2015). Elle se fonde sur une procédure MCMC (*Markov Chain Monte Carlo*) et suppose des corrélations entre les années d'accident. Au-delà de l'approche USP, cette méthode peut, dans une certaine mesure, présenter un intérêt, puisque les auteurs eux-même la présente comme une alternative comparable au *re-reserving* et à la formule de Merz & Wüthrich (plus proche d'un modèle interne partiel, cependant). Cette approche bayésienne utilise le programme de simulation JAGS (2023).

Le modèle CCL. Le modèle se base sur une distribution *a posteriori* (obtenue via le théorème de Bayes), combinant à la fois les observations et une distribution *a priori* (exogène) des paramètres. En considérant l'information contenue dans D_I, le modèle considère les paiements comme une fonction de paramètres $Y_{i,j} = f(\alpha_i, \beta_j)$ où α_i représente le logarithme de l'ultime pour l'année d'accident i, et β_j le niveau d'ultime payé jusqu'à l'année de développement j (qui dépend de la cadence de règlement du portefeuille).

Hypothèse A.9.1. *(a)* $C_{i,j} \sim \mathcal{LN}(\mu_{ij}, \sigma_j)$, *avec*

$$\mu_{0j} = \alpha_0 + \beta_j, \quad \mu_{ij} = \alpha_i + \beta_j + \rho \cdot (\log(C_{i-1,j}) - \mu_{i-1,j}) \quad \textit{pour } i > 0,$$

(b) Le paramètre ρ est un facteur de corrélation entre les années d'accident avec $\rho \sim \mathcal{U}(-1, 1)$.

(c)

$$\alpha_i \sim \mathcal{N}(\log(B_i) + \log(elr_i), \epsilon_i) \quad \textit{pour } i = 0, \dots, I,$$

où B_i représente la prime brute, $\log(elr_i)$ le logarithme du ratio S/P attendu (expected loss ratio) pour chaque année d'accident i, et ϵ le degré de précision d'inférence bayésienne choisi.

(d)

$$\log(elr_i) \sim \mathcal{U}(\gamma_i, \delta_i), \quad \beta_j \sim \mathcal{U}(-\eta, 0) \quad \textit{pour } j = 0, \dots, I-1,$$

(e) Le paramètre de variance logarithmique est tel que

$$\sigma_j^2 = \sum_{h=j}^{I} \tau_h,$$

où $\tau_h \sim \mathcal{U}(0,1)$.

En considérant K jeux de paramètres représentatifs des K scénarios de provisionnement possibles,

$$\Theta^{(k)} = \left\{ (\alpha_i)_{i=0}^{I}, (\beta_j)_{j=0}^{I-1}, (\sigma_j)_{j=0}^{I}, \rho \right\} \quad \text{avec } k = 1, \dots, K.$$

Le coût ultime des sinistres est alors simulé par

$$C_{i,I}^{(k)} \sim \mathcal{LN}(\mu_{i,I}^{(k)}, \sigma_I^{(k)}) \quad \text{pour } i = 1, \dots, I,$$

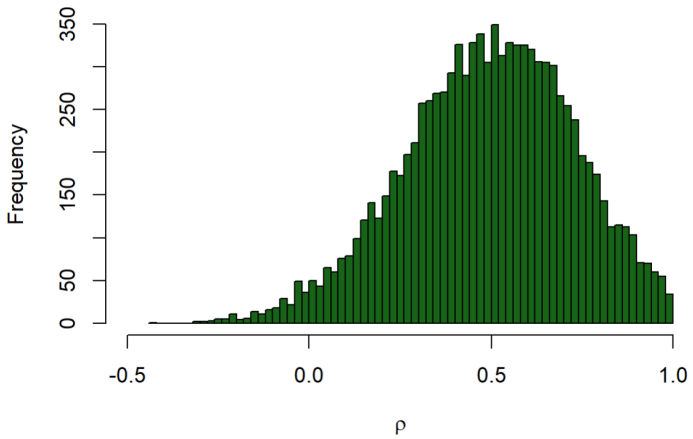
et les réserves sont évaluées comme

$$R^{(k)} = \sum_{i=1}^{I} C_{i,I}^{(k)} - \sum_{i=1}^{I} C_{i,I-i}.$$

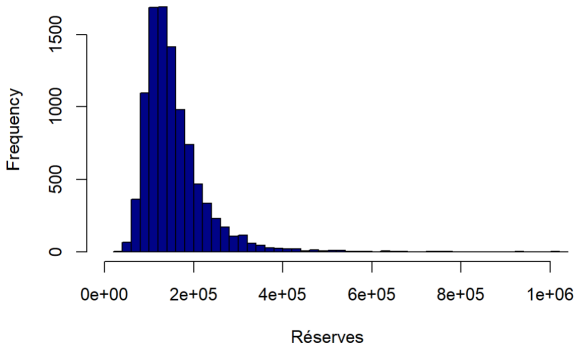
Vision à un an. À horizon un an, la charge en capital pour le risque de réserve est déterminée par le quantile à 99,5% (*Value-at-risk*) de la distribution du CDR

$$\text{SCR}_{0.995} = \text{VaR}_{0.995} \left(\sum_{i=1}^{I} Y_{i,I-i+1} + R^{(D_{I+1})} v_1 - R^{D_I} \right),$$

où v_1 est un facteur d'actualisation.



(a) Distribution a posteriori du paramètre ρ



(b) Distribution des ultimes des accidents de 2001 à 2010 (10 000 simulations) avec la méthode CCL (milliers €)

FIGURE A.1 : Modèle CCL

Via une procédure MCMC décrite dans l'algorithme 7, des simulations de pertes sachant les paramètres précédents sont effectuées pour obtenir la distribution des engagements à un an

$$Y^{1\text{yr}} = \sum_{i=1}^{I} Y_{i,I-i+1} + R^{(D_{I+1})}.$$

Algorithme 7 Algorithme MCMC pour la distribution du CDR

1: Simulation de K scénarios de paramètres $\Theta^{(k)}$ et obtention de la distribution a posteriori de $\mu_{i,j}$ et $\sigma^k = [\sigma_j^k]$, $i, j \in \{0, ..., I\}$.

2: Pour chaque k, simuler S réalisations des paiements à horizon un an $C_{i,I-i+1}^{(s,k)}$ avec

$$C_{i,j-1} \sim \log \mathcal{N}(\mu_{i,j-1}, \sigma_j) \quad (K \cdot S \text{ simulations }).$$

3: Calcul des $K \cdot S$ diagonales des paiements à un an

$$Y_{i,I-i+1}^{(k,s)} = C_{i,I-i+1}^{(k,s)} - C_{i,I-i}.$$

4: Probabilité a posteriori de chaque scénario, avec ϕ la densité log-normale et T le triangle original auquel la nouvelle diagonale simulée est ajoutée

$$L[T^{(s,k)} \mid \Theta^{(k)}] = \prod_{C_{i,j}^{(s,k)} \in T^{(s,k)}} \Phi_L(C_{i,j}^{(s,k)} \mid \mu_{i,j}^{(s,k)}, \sigma_{i,j}^{(s,k)}),$$

$$\mathbb{P}(\Theta^{(k)}|T^{(s,k)}) = \frac{L[T^{(s,k)}|\Theta^{(k)})] \cdot \mathbb{P}(\Theta^{(k)})}{\sum_{h=1}^{K} L[T^{(s,h)}|\Theta^{(h)})] \cdot \mathbb{P}(\Theta^{(h)})} = \frac{L[T^{(s,k)}|\Theta^{(k)})]}{\sum_{h=1}^{K} L[T^{(s,h)}|\Theta^{(h)})]}.$$

5: Itérations pour obtenir S distributions postérieures pour les paramètres.

6: Pour chaque *set* de paramètres, calcul

$$E[C_{i,j}^k] = \exp(\mu_{i,j}^k + \frac{\sigma_j^{2(k)}}{2}).$$

7: Calcul de

$$\hat{R}_s^{(D_{I+1})} = \sum_{k=1}^{K} \left(R_k^{(D_{I+1})} \cdot \mathbb{P}(\Theta^{(k)}|T^{(s,k)}) \right).$$

8: Distribution du CDR : itérations pour obtenir S *batch* de simulations en échantillonnant, pour chaque s une diagonale parmi les K possibilités

$$Y_s^{1yr} = \left(\sum_{i=1}^{t} \tilde{Y}_{i,t-i+1}^{(s)} \right) + \hat{R}_s^{(D_{I+1})}.$$

A.10 Le modèle *DeepTriangle* de Kévin Kuo

Le modèle *DeepTriangle* (Kuo, 2018) tient en partie inspiration du modèle Munich Chain-Ladder (Quarg et Mack, 2004). Il s'agit également d'un modèle de provisionnement basé sur des *RNN* (séquences à séquences) entrainés sur la base de données de la NAIC (2015). Cette base de données contient au total 200 triangles de liquidation issus de quatre branches d'activité (*truck liability, personal auto, workers compensation, other liability*). Quatre modèles sont entrainés, un pour toutes les entreprises de la même branche d'activité. Dans le présent travail, nous nous intéressons à la branche *truck liability*. Les deux variables à considérer pour estimer la séquence $(X_{i,j}, X_{i,j+1}, \dots, X_{i,I-i})$ sont

$$(X_{i,0}, X_{i,1}, \dots, X_{i,j-1}),$$

avec $X_{i,j} = \left(\frac{Y_{i,j}}{P_i}, \frac{R_{i,j}}{P_i}\right)$ où R désigne les provisions dossier/dossier (*claims outstanding*) et P désigne un vecteur d'exposition. Chaque organisme d'assurance détient un identifiant qui permet de la différentier des autres face à l'expérience.

Les données d'entrainement correspondent au triangle supérieur et celles de test correspondent aux paiements cumulés associés à la dernière année de développement.

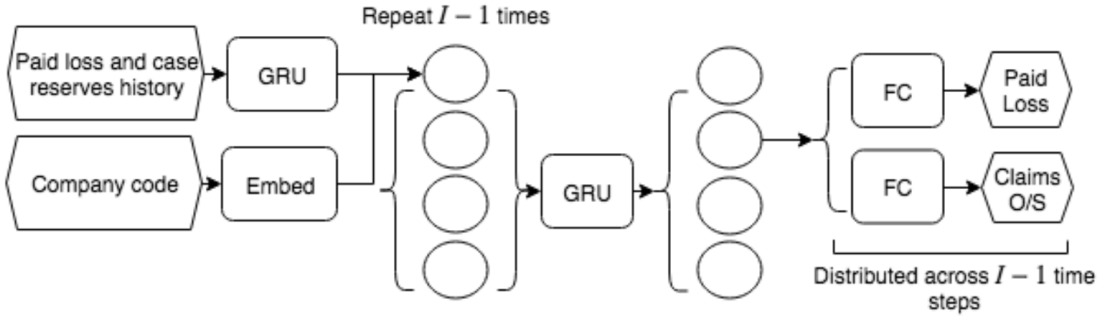


Figure A.2 : Architecture du modèle *DeepTriangle* (Kuo, 2018)

Comme expliqué dans les figures A.2 et A.3, le modèle utilise l'architecture des GRU, versions spécifiques des cellules LSTM pour analyser les séquences et dont les équations vérifient

$$\begin{aligned}
\hat{\mathbf{h}}^{<\mathbf{t}>} &= \tanh\left(W_h\left[\mathbf{r}^{<\mathbf{t-1}>} \circ \mathbf{h}^{<\mathbf{t-1}>}, \mathbf{x}^{<\mathbf{t}>}\right] + \mathbf{b_h}\right), \\
\Gamma_r^{<t>} &= \sigma\left(W_r\left[\mathbf{h}^{<\mathbf{t-1}>}, \mathbf{x}^{<\mathbf{t}>}\right] + \mathbf{b_r}\right), \\
\Gamma_u^{<t>} &= \sigma\left(W_u\left[\mathbf{h}^{<\mathbf{t-1}>}, \mathbf{x}^{<\mathbf{t}>}\right] + \mathbf{b_u}\right), \\
\mathbf{h}^{<\mathbf{t}>} &= \Gamma_u^{<t>}\hat{\mathbf{h}}^{<\mathbf{t}>} + (1 - \Gamma_u^{<t>})\mathbf{h}^{<\mathbf{t-1}>},
\end{aligned}$$

où $W_h, W_r, W_u, b_h, b_r, b_u$ sont les poids et biais de la cellule. Les GRU comportent un paramètre *dropout* pour éviter le surapprentissage. La couche *embedding* consiste en un l'apprentissage d'un*mapping* pour obtenir un vecteur associé « optimisé »(de taille $\mathbb{R}^{50-1}$) au sens où deux entreprises sont d'autant plus proches les unes des autres que la distance entre les deux vecteurs *embedding* est minimisé au sens de la distance euclidienne. Ces vecteurs représentent intuitivement les caractéristiques de chaque organisme, comme les volumes ou encore le mode de gestion des sinistres lors du provisionnement. Les résultats obtenus de la couche *embedding* et la première cellule GRU sont ensuite concaténés, réintégrés dans une seconde cellule puis intégrés dans deux couches entièrement connectées (FC).

Enfin, la fonction de coût est définie pour la séquence $(X_{i,j}, X_{i,j+1}, \dots, X_{i,I-i})$ par

$$\frac{1}{I-i-(j-1)} \sum_{k=j}^{I-i} \left(\frac{(Y_{i,k} - \hat{Y}_{i,k})^2 + (R_{i,k} - \hat{R}_{i,k})^2}{2} \right).$$

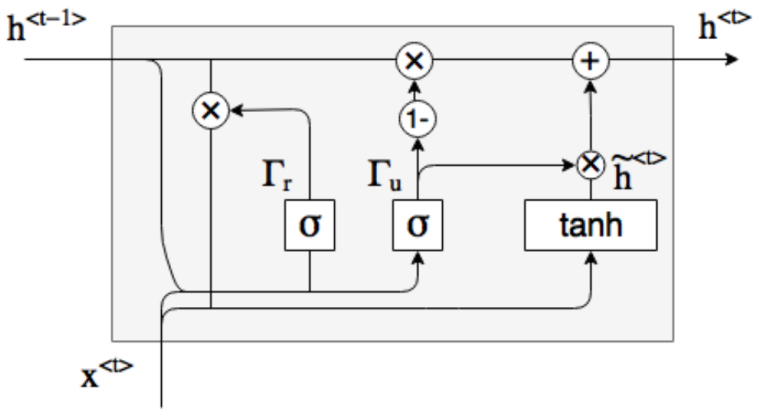


FIGURE A.3 : Architecture d'un GRU (KUO, 2018)

L'optmisation des paramètres s'effectue avec l'algorithme AMSGrad, variante de ADAM (REDDI et al., 2019).

Ce modèle présente une limite majeure. En effet, en pratique, un organisme d'assurance ne détient pas les données de sinistres de ses concurrents permettant d'entraîner un unique modèle comme le fait l'auteur. Dans le cas d'une captive d'un groupe industriel, on pourrait toutefois imaginer que le groupe dispose des triangles de liquidation de plusieurs de ses filiales.